Journal of King Saud University – Computer and Information Sciences 35 (2023) 145–174

HOSTED BY

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

Credit card fraud detection in the era of disruptive technologies: A systematic review

Asma Cherif [a,b], Arwa Badhib [a], Heyfa Ammar [b,c], Suhair Alshehri [a], Manal Kalkatawi [a], Abdessamad Imine [d]

[a] Department of Information Technology, Faculty of Computing & Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia
[b] Center of Excellence in Smart Environment Research, King Abdulaziz University, Jeddah, Saudi Arabia
[c] RISC-ENIT Research Lab, National Engineering School of Tunis, University of Almanar, Tunisia
[d] Lorraine University, CNRS, INRIA, Vandœuvre-lès-Nancy, France

ARTICLE INFO

Article history:
Received 11 July 2022
Revised 3 November 2022
Accepted 16 November 2022
Available online 5 December 2022

Keywords:
Credit card fraud detection
Machine learning
Deep learning
Class imbalance

ABSTRACT

Credit card fraud is becoming a serious and growing problem as a result of the emergence of innovative technologies and communication methods, such as contactless payment. In this article, we present an in-depth review of cutting-edge research on detecting and predicting fraudulent credit card transactions conducted from 2015 to 2021 inclusive. The selection of 40 relevant articles is reviewed and categorized according to the topics covered (class imbalance problem, feature engineering, etc.) and the machine learning technology used (modelling traditional and deep learning). Our study shows a limited investigation to date into deep learning, revealing that more research is required to address the challenges associated with detecting credit card fraud through the use of new technologies such as big data analytics, large-scale machine learning and cloud computing. Raising current research issues and highlighting future research directions, our study provides a useful source to guide academic and industrial researchers in evaluating financial fraud detection systems and designing robust solutions.

Contents

E-mail addresses: acherif@kau.edu.sa (A. Cherif), adhib@kau.edu.sa (A. Badhib), heyfa.amar@gmail.com (H. Ammar), sdalshehri@kau.edu.sa (S. Alshehri), mkalkatawi@kau.edu.sa (M. Kalkatawi), abdessamad.imine@loria.fr (A. Imine)
Peer review under responsibility of King Saud University.
This project was funded by the Institutional Fund Projects under grant No. (IFPRC-032-612-2020). The authors, therefore, gratefully acknowledge technical and financial support from the Ministry of Education and King Abdulaziz University, Jeddah, Saudi Arabia.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2022.11.008
1319-1578/

1. Introduction

1.1. Motivation

Already making room for digital commerce, cash as a method of payment has been further supplanted by the COVID-19 pandemic. Credit card payment is becoming the main representative of contemporary digital commerce within the global economy.

Different stakeholders, including issuers, banks, payment processors, and merchants, are constantly looking for ways to leverage multiple technological advancements to better align with their end-users' preferences. New technologies pave the way for innovative payment solutions. The abundance of Internet of Things (IoT) devices and improved connectivity, in-app payments and mobile device penetration are contributing to both the advancement and disruption of credit card payment systems. For example, many companies such as Amazon Go are currently experimenting with biometrics-based payments. Through tokenization (Liu et al., 2020), mobile IoT devices such as smartwatches are used to exchange information with nearby systems to make on-demand payments, giving rise to new communication models to perform transactions.

Credit cards are popular for online banking and are widely used in online transactions and e-commerce. However, the evolution and expansion of credit card use has led to the emergence of multiple forms of fraud. Fraudsters are using increasingly sophisticated approaches to conduct illegal transactions, resulting in significant losses for cardholders and banks. From theft, fishing for credit card information, and producing fake cards to mimic legitimate user behavior, today's fraudsters can conduct fraudulent transactions more easily than ever.

At the same time, the normalization of data and the increased use of neural networks is making the use of Artificial Intelligence (AI) and deep learning techniques essential by card issuers and banking services. Today, AI is paving the way in designing new methods to better handle next-generation credit card fraud detection by supporting increased approval rates, minimizing declined transactions, and enabling proactive monitoring of credit limits. However, there are many challenges when navigating intelligent banking transaction processing, including changing customer behavior that needs to be addressed so that legal operations are upheld. Due to these transitions and challenges, banks and payment processors are rapidly modernizing their payment technology, which can create security issues.

Consequently, it is essential to have robust and up-to-date credit card fraud detection systems in place. Credit card fraud detection helps identify suspicious transactions by classifying incoming transactions into two classes: legitimate and illegitimate transactions. Credit card fraud can take one of two forms: online and offline. In the case of online fraud, fraudsters carry out fraudulent transactions involving online purchases; while, in offline fraud, they make malicious transactions using a stolen credit card.

Many research works have addressed the problem of credit card fraud. Thus, it is essential to analyze their reported solutions to provide a roadmap for researchers in the field. Although there are previous surveys (see Section 2), many new methods have been proposed recently that require analysis. Moreover, existing surveys have mainly focused on detection models rather than exploiting new technologies and computational methods. Our study therefore offers a comprehensive investigation considering several aspects of credit card fraud detection with a focus on deep learning methods and disruptive technologies.

1.2. Contributions

Our research work presents a comprehensive survey of credit card detection with an emphasis on using recent advances and new technologies to cover not only machine learning algorithms but also the integration of recent advances, such as big data technologies, class imbalance issues and real-time aspects of the

detection. We also aim to review recent work published between 2020 and 2021 to provide relevant guidance for future researchers, as a huge body of new solutions has emerged over the past two years.

The contributions of this survey are threefold:

- Analysis of the most recent research conducted to solve the credit card fraud detection problem from three angles: (i) the machine learning method used and its effectiveness in solving the detection problem, (ii) the class imbalance problem and how it has been resolved in the literature, and (iii) the feature engineering issue.
- In-depth review of the state-of-the-art solutions in terms of implementation and testing. The main metrics used in the field are discussed and analyzed according to their relevance.
- Comprehensive presentation of research challenges along with possible research directions helping researchers and developers to provide more powerful and up-to-date solutions to navigate the credit card fraud detection problem.

1.3. Paper organization

The remainder of the paper is organized as follows. Section 2 explains the methodology when conducting this review. Section 3 presents the research background. Section 4 displays the proposed works based on traditional machine learning while Section 5 browses deep learning-based solutions. Section 6 emphasizes the class imbalance problem and summarizes state of the art techniques used for credit card fraud detection. Section 7 details the testing metrics with regard to their relevance to the credit card fraud problem. It also presents the different datasets included in the literature, and the best results achieved in the reviewed works. Section 8 highlights open issues and sets out some research directions. Finally Section 9 concludes the paper.

Table 1 shows the abbreviations used in this paper.

Table 1
List of Abbreviations.

Notation	Description
ADASYN	Adaptive Synthetic
AI	Artificial Intelligence
ANN	Artificial Neural Network
AUC	Area Under the Curve
AUC-PR	Area Under the Curve Precision–Recall
AUC-ROC	Area Under the Receiver Operating Characteristic
BBE	Balanced Bagging Ensemble
BMR	Bayes Minimum Risk
BMR-DT	BMR Decision Tree
BMR-LR	BMR Logistic Regression
BMR-RF	BMR Random Forest
BPNN	Back Propagation Neural Network
CNN	Convolution Neural Network
CNP	Card-Not-Present
DDM	Data-Driven Model
DL	Deep Learning
DT	Decision Tree
DWE	Dynamic Weighted Entropy
FDS	Fraud Detection System
FN	False Negative
FNN	Feed-Forward Neural Network
FP	False Positive
FPR	False Positive Rate
G-mean	Geometric Mean
GBT	Gradient Boosting
GNB	Gaussian Naïve Bayes
GRU	Gated Recurrent Unit
IoT	Internet of Things
KNN	K-Nearest Neighbor
LinR	Linear Regression
LOF	Local Outlier Factor
LogR	Logistic Regression
LSTM	Long Short-term Memory
MCC	Matthews Correlation Coefficient
ML	Machine Learning
MLP	Multi Layer Perceptron
MOPSO	Multiple Objectives Particle Swarm Optimization
NB	Naive Bayes
NN	Neural Network
PCA	Principal Component Analysis
PNN	Probabilistic Neural Network
RF	Random Forest
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
ROS	Random Oversampling
RUS	Random Undersampling
SEPA	Single European Payments Area
SMOTE	Synthetic Minority Oversampling
SVM	Support Vector Machine
SVM-RFE	Support Vector Machine-Recursive Feature Elimination
TN	True Negative
TP	True Positive
TNR	True Negative Rate
TPR	True Positive Rate
WELM	Weighted Extreme Learning Machine

2. Methodology

This research followed the strategy presented in Fig. 1. It consists of three main stages: the research design, the research methodology and the research output.

2.1. Related surveys

The researcher first searched for survey papers in the field and selected the most relevant ones. These papers were analysed according to the following criteria:

- The publication date;
- The coverage;
- The topics discussed.

Whilst many surveys have been proposed investigating smart credit card fraud detection systems, most of them have either a broad scope that covers many other areas in the commercial field, thus dispersing researchers, or have a very limited scope in a subfield of the area (short surveys, supervised approaches, or deep learning approaches). In addition, some papers only describe state-of-the-art solutions without providing a deep analysis of the pros and cons of the covered literature.

The work in Al-Hashedi and Magalingam (2021) covers research papers on financial fraud in general from 2009 to 2019 inclusive. It mainly discusses works based on data mining techniques and classifies the literature based on range of factors, including publication year, publisher, method used, and research area (credit fraud, cryptocurrency fraud, insurance, financial). It is a comprehensive review of research focusing on the detection of financial fraud, credit card fraud, insurance fraud and other types of fraud. The data mining techniques used to detect financial frauds are described. Datasets and validation metrics were also specified. Finally, the pros and cons of each data mining technique are specified in the paper. However, the review is restricted to 'classification' techniques only, and does not describe the complete detection chain, which is important in order to gain an idea of the features used for detection. Furthermore, the authors did not focus on credit card fraud and did not cover the class imbalance issue or feature engineering problems. There was also a lack of consideration for new trends such as big data.

In Lucas and Jurgovsky (2020), the authors considered the challenges of data-driven credit card fraud detection. Specifically, they focused on the data imbalance problem and how to address the evolving behaviour issue (dataset drift) in state-of-the-art solutions.

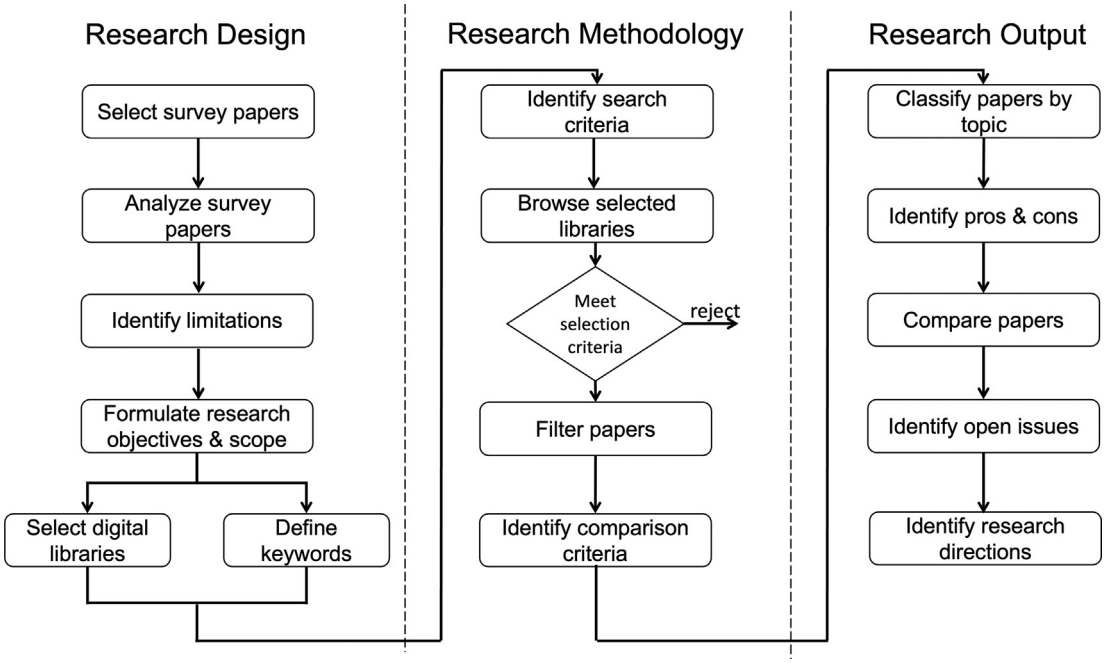

Fig. 1. Research strategy.

However, some papers were too old and no comparative analysis was provided. Popat and Chaudhary (2018) is a short survey in which only seven related works were discussed based on their machine learning techniques. However, the authors did not discuss the details of each reviewed work. In Kanika and Singla (2020), the authors analysed deep learning based fraud detection techniques for online transactions. The authors also provided information about the main datasets used and the results achieved. However, the scope of the study was limited to deep learning techniques.

In Mittal et al. (2020), the authors classified credit card fraud and summarized the main features described in the reviewed studies. They also discussed credit card fraud detection. A shortlist of research directions was suggested but lacks the detail needed to better guide researchers.

Table 2 shows a comparison of this study with the most recent surveys.

After identifying the limitations of existing surveys, the following research objectives were formulated:

- To review recent research papers;
- To survey papers using deep learning techniques;
- To analyse the usefulness of disruptive technologies and new methods for the credit card fraud detection;
- To study the security concerns related to credit card fraud detection.

2.2. Selection and screening of papers

The previous step led to the definition of research keywords and the selection of digital libraries for eligible studies, namely Elsevier, Springer, IEEE explore, ACM, etc.

The second phase identifies the search criteria as follows:

- Include works published from 2015;
- Include works related to the fields of data mining and AI;
- Include works investigating new technologies, such as IoT, big data, and cloud computing;
- Consider the security aspects of credit card fraud research.

Table 2
Comparison with recent survey papers.

Ref#	Publication year	Coverage	Reviewed papers	Topics							
				Class imbalance	Big data	Feature engineering	Credit card focus	Datasets	Pros and Cons	Open issues and future research directions	Security and privacy
(Al-Hashedi and Magalingam, 2021)	2019	2009–2019	25					✓	✓		
(Lucas and Jurgovsky, 2020)	2019	1998–2019	21	✓		✓	✓	✓			
(Popat and Chaudhary, 2018)	2018	1014–2017	7						✓		
(Kanika and Singla, 2020)	2020	1997–2017	24								
(Mittal et al., 2020)	2020	1994–2020	22	✓		✓	✓	✓			
Ours	2022	2015–2021	40	✓	✓	✓	✓	✓	✓	✓	✓

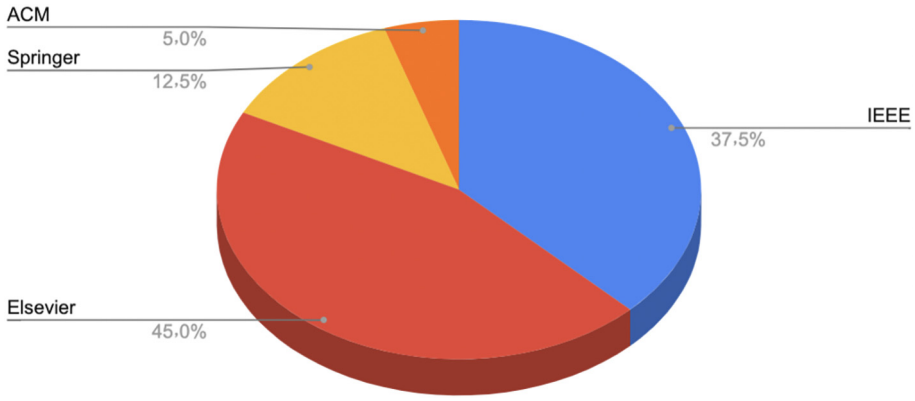

Fig. 2. Papers distribution according to the publisher.

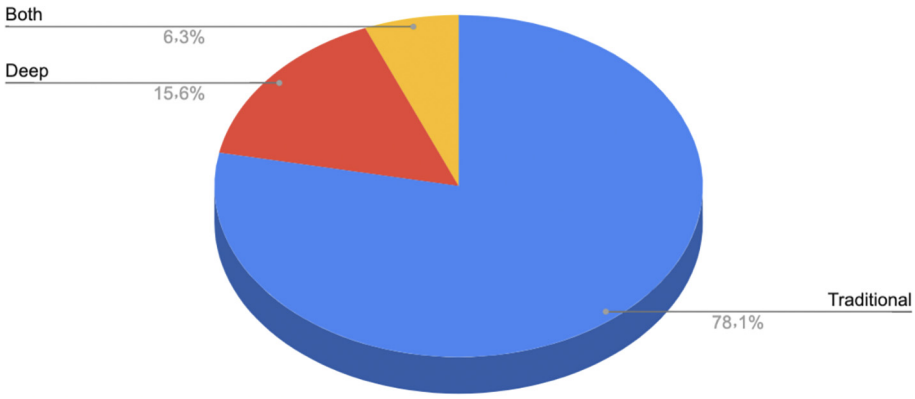

Fig. 3. Papers distribution according to the used technique.

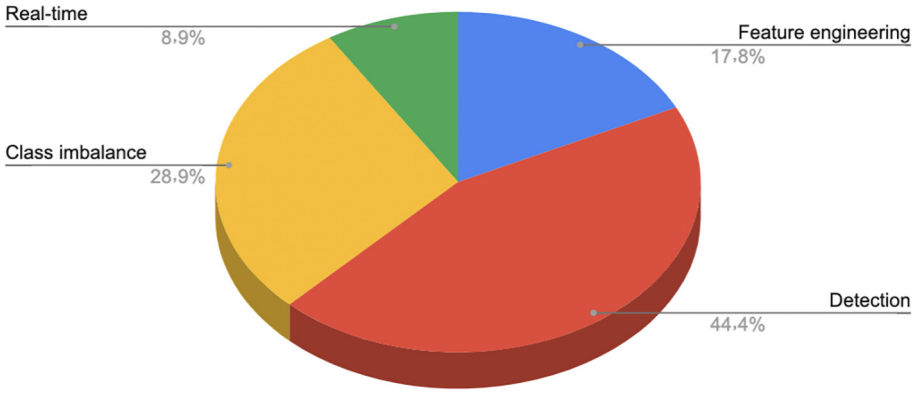

Fig. 4. Papers distribution according to the research problem.

Table 3
Taxonomy.

Main Topic	References
Detection	(Srivastava et al., 2016) (Kewei et al., 2021) (Sudha and Akila, 2021) (Fu et al., 2016) (Ingole et al., 2021) (Ali Yeşilkanat et al., 2020) (Mohammed et al., 2018) (Wang et al., 2018) (Padmanabhuni et al., 2019) (Roy et al., 2018) (Babu and Pratap, 2020) (Rtayli and Enneya, 2020) (RB and KR, 2021) (Forough and Momtazi, 2021) (Bagga et al., 2020) (Carcillo et al., 2021)
Class imbalance	(Benchaji et al., 2018) (Thennakoon et al., 2019) (Tran and Dang, 2021) (Li et al., 2021) (Kim et al., 2019) (Dornadula and Geetha, 2019) (Rtayli and Enneya, 2020) (Zhu et al., 2020) (Ingole et al., 2021) (Yang et al., 2019) (Baabdullah et al., 2020) (Akila and Srinivasulu Reddy, 2018) (Olowookere and Adewale, 2020)
Feature Engineering	(Thennakoon et al., 2019) (Lucas et al., 2019, 2020) (Correa Bahnsen et al., 2016) (Zhang et al., 2021) (Lucas et al., 2020) (Cochrane et al., 2021) (Han et al., 2021) (Jurgovsky et al., 2018) (Sudha and Akila, 2021)
Recommender systems	(Cui et al., 2021)
Optimization	(Han et al., 2021) (Soltani Halvaiee and Akbari, 2014) (Benchaji et al., 2021) (Zhu et al., 2020)
Real-time aspect	(Ali Yeşilkanat et al., 2020) (Soltani Halvaiee and Akbari, 2014) (Thennakoon et al., 2019) (Zhou et al., 2021)
Big data technologies	(Zhou et al., 2021) (Soltani Halvaiee and Akbari, 2014)
Security/privacy	(Yang et al., 2021) (Yang et al., 2019)

The titles and abstracts of all the retrieved papers that met these search criteria were screened and filtered for inclusion in the survey. Records marked as ineligible (irrelevant or beyond the scope of the study) were excluded. The final number of full-text relevant articles retrieved was 40. Fig. 2 shows the distribution of the retained papers according to the publisher.

2.3. Research output

Each selected model was reviewed descriptively to classify works into a comprehensive taxonomy, as illustrated below Fig. 3 summarizes the distribution of the selected papers according to the technique used.

As each research paper has a specific research problem according to the main challenges faced in the field, reviewed works were also classified according to the research problem in Fig. 4. Table 3 shows the distribution of the papers according to their topics. Fig. 5 displays the taxonomy of the state-of-the-art techniques suggested for the credit card fraud detection problem according to the main topic and subtopics investigated by the researchers.

3. Research background

A fraud is a planned deceit committed to acquire monetary gain. The increasing use of electronic payment modes such as credit and debit cards has led to a rise in credit card fraud. The growing popularity of mobile banking has led to an increase fraudulent pay-

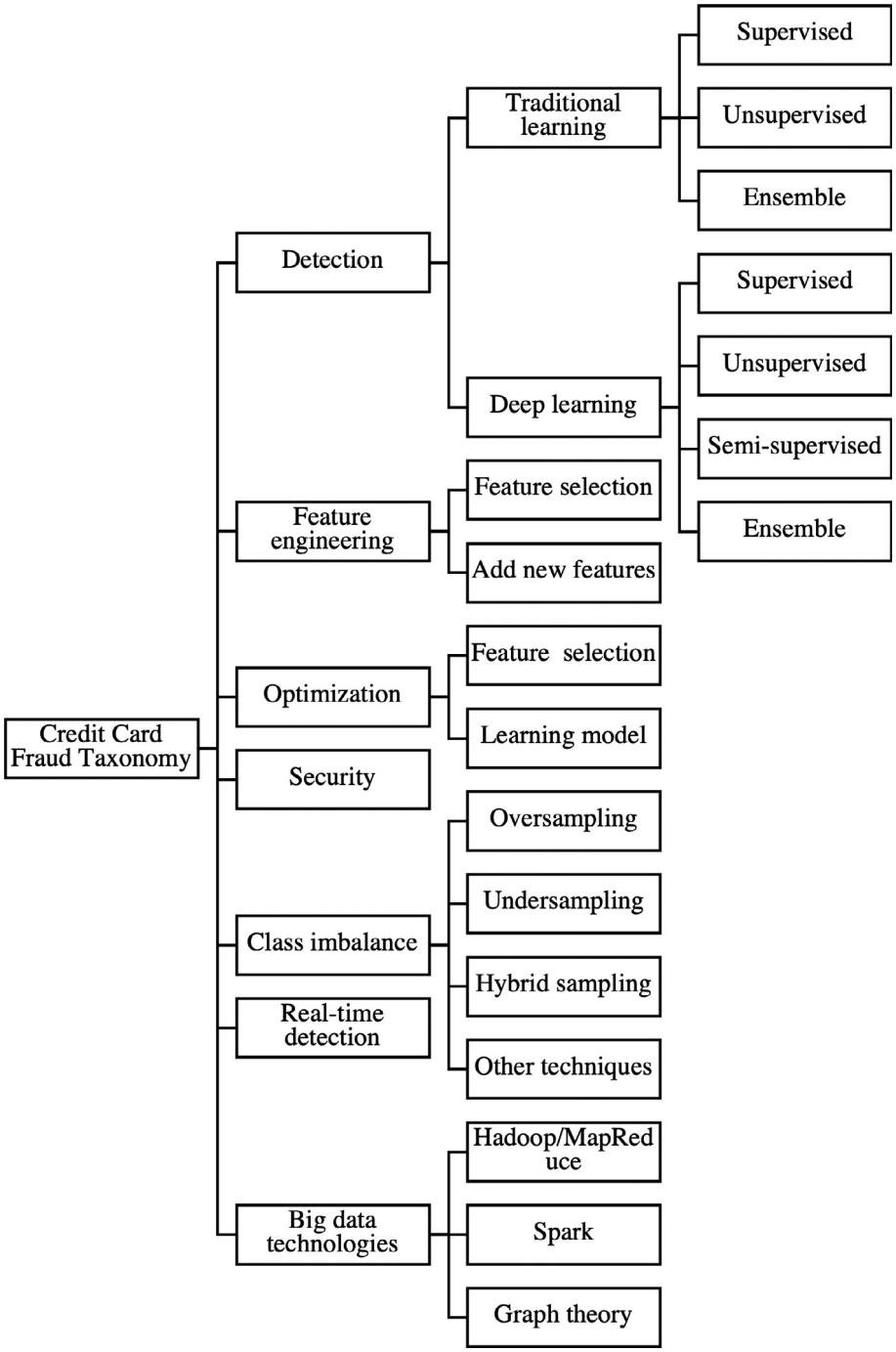

Fig. 5. Credit card fraud detection taxonomy.

ment transactions. As a result, financial losses are also increasing. Credit cards can be used either online or offline to purchase items. Online payments do not require the physical presence of the card and card data is consequently prone to attack. This type of fraud is also known as Card-Not-Present (CNP) fraud.

Furthermore, the use of contactless payments with chip cards and mobile device payments through near-field communication (NFC) is increasing, as this allows faster payment. These payment types have become particularly prevalent following the Covid-19 pandemic. They use short-range wireless communication technology that enables contactless payment (Vishwakarma et al., 2021). Unlike traditional payments, NFC payment depends on two other partners: the phone manufacturer and the mobile operator. The security policies of these partners can be manipulated by the telephony market rather than the payment market(Gerbaix, 2010), which leads to more security issues and exposes customers to more credit card fraud than when they pay using a physical card. Although small expenses are allowed on these types of payments, scammers can learn from user behaviour and make a large number of transactions before the customer alerts the bank. Therefore, it is important to provide robust detection solutions.

According to the seventh Single European Payments Area (SEPA) report published in 2021 (Sepa report, 2022) and analysis of data for the year 2019, the total value of fraudulent transactions amounted to 1.87 billion euros, of which 80% were CNP payments. In contrast, the share of fraud at ATMs and point-of-sale terminals decreased to 5% and 15% of the total fraud value, respectively (see Fig. 6). Compared to the card present frauds, CNP fraud has increased in recent years. CNP fraud is thus a very serious concern to the credit card industry.

3.1. Credit Card Fraud Detection System Design

A credit card Fraud Detection System (FDS) consists of a succession of detection modules performed to reject suspicious transactions (Kim et al., 2019; Dal Pozzolo et al., 2015; Dal Pozzolo et al., 2018). Few studies have investigated the design of a comprehensive framework for credit card fraud detection. The best-known model was suggested by Andrea et al. (Dal Pozzolo et al., 2018) (see Fig. 7. (a)). They defined five layers of control according to the operation of industrial partners: a Terminal, a Blocking Rules module, Scoring Rules, Data-Driven Model (DDM) and Investigators. The fraud control flows through five layers. First, the customers perform their transactions through a *Terminal*. These transactions are then forwarded to the *Blocking Rules Module* to perform an initial primary real-time, protection check according to logical rules. This step consists of checking the If-Then-Else rules for detecting defrauding patterns that have been already found by human investigators. These rules are kept confidential by the industry for security reasons (Dal Pozzolo et al., 2018)[1]. The next step is *Scoring Rules*, followed the *DDM*. This model includes *Investigators* which requires human intervention to double-check transactions.

The DDM uses historical transaction data that have been previously trained to build a classifier, or relies on another statistical model to detect fraudulent transactions. The scores of legal transactions are estimated. If the score overreaches a well-established threshold, the transaction is rejected and then forwarded to experts for further analysis. The DDM is usually trained from a labelled dataset and is fully automated. It is expected that the DDM will detect fraudulent transactions, going beyond the experience of investigators, which is missing in the rule-based module.

[1] For instance, a rule could be "IF the previous transaction in a different continent AND less than 1 h from the previous transaction THEN reject" (Dal Pozzolo et al., 2018).

If there are fraudulent transactions that are not detected by the classifier or the statistical model, the alerts are forwarded to investigators as soon as they are reported by the cardholder. The scoring module has to return accurate alerts to reduce both the number of false alarms and overlooked fraudulent transactions. This process is illustrated in Fig. 7. (a).

The previously-mentioned model was slightly modified by Kim et al. (2019) (see Fig. 7. (b)). The improvement mainly automated the scoring model and grouped it with the DDM model as a combined data-driven scoring model (DDSM). The new design also emphasizes the real-time aspect of the FDS by considering the whole process as real-time processing, unlike (Dal Pozzolo et al., 2018) who designed both scoring and DDM as near real-time processing.

In general, the fraud detection problem consists of a binary classification problem. Incoming transactions must be classified as either fraudulent or genuine. However, a FDS that quantifies the degree of fraud is more useful in practice. Since investigating suspicious transactions is time-consuming and the number of investigators is limited, a FDS needs to operate in a fully automated and distributed fashion. To enable this, the fraud detection system was redesigned, as shown in Fig. 8. It is suggested that the Blocking rules module be complemented by an intelligent model that estimates the risk of the transaction. If the risk is high, the transaction is rejected. The system moves forward only if the risk is low and all the predefined rules are passed. This ensures that even if the blocking rules are validated and the operation is risky, it will be rejected at an early stage. Then, once the risk is low to moderate, the transaction is passed to a Distributed DDSM, named 3DSM. This ensures that the learning module is distributed and relies on new technologies (edge/cloud computing). Human intervention should be minimized as much as possible to feed the learning system and update its behaviour. This design also emphasizes the necessity of using multiple data inputs for the learning module. This includes contextual information (for example location), and user behaviour information (for example spending or typing).

3.2. Credit Card FDS Challenges

Many challenges are faced when dealing with credit card fraud detection as depicted in Fig. 9. These issues are mainly classified into three categories:

- Data-related challenges: these challenges are related to the data required for building a robust credit card fraud detection system. They include:
  - Class imbalance;
  - Lack of real data;
  - Data drift/shift;
  - Data overlapping.
- Security-related challenges, which mainly involve privacy concerns.
- Deployment/Implementation challenges which are related to the efficiency of the final FDS and covers:
  - Distributed implementation.
  - Time complexity.

Data-related challenges. FDS relies on the DDSM module which mainly requires a high-quality dataset to build effective detection models. However, many issues are faced by researchers when dealing with credit card datasets.

- **Class imbalance**: An unbalanced or skewed dataset is a dataset in which the distribution of samples in the known categories is skewed (or biased). The distribution can range from a slight skew to a severe imbalance where the amount of raw data in

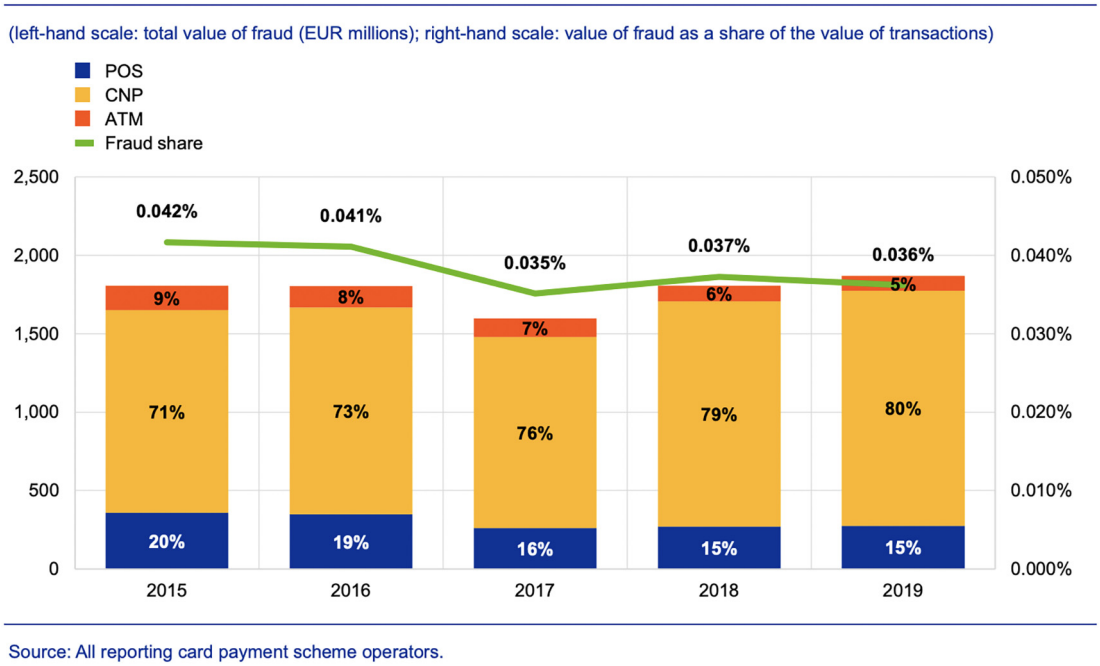

Fig. 6. Total value of card fraud using cards issued within SEPA.

the minority category is less than in the majority category. Unbalanced classifications present a challenge to predictive modelling. In fact, most machine learning algorithms used for classification assume an equal number of samples per class (Sun et al., 2009). Data imbalance results in predictive models with poor predictive performance, specifically for the minority class. The model will favor the majority class while the minority class is generally more relevant. Thus, the model is more exposed to classification mistakes for the minority class than the majority class. Traditional machine learning algorithms require a balanced dataset (Zheng, 2020). However, in practice, there are far fewer samples in the abnormal class compared to normal samples. In fraud detection specifically, the number of fraudulent transactions is very small compared to the number of legitimate transactions. An unbalanced dataset degrades the performance of most machine learning algorithms, such as Support Vector Machine (SVM) and random forests (Dablain et al., 2022). The techniques proposed to solve this problem go in two directions, either adapting the algorithm to make it powerful enough to handle unbalanced data sets, or solving the unbalanced data set using pre-process sampling methods (Lucas and Jurgovsky, 2020).

- **Lack of real data**: This challenge is also referred to as the unavailability of sufficient labelled data. Indeed, for many reasons, mainly related to privacy concerns, there is a lack of real data with which to build accurate models. In many cases, the data is not labelled, which requires additional effort to label data rows. Consequently, anomaly detection is one of the common methods used in detecting fraud. However, this is highly dependent on user behaviour and any variation can be detected as fraud. Anomaly systems rely on users' historical behaviour, which can be limited (Zheng et al., 2018). Some studies have investigated this limitation by utilizing information from other similar users, but this increases the problems of choosing similar individuals (Cui et al., 2021).
- **Data overlapping**: Overlap is faced when samples of different classes appear in the same area of the data space, making it difficult for the classifier to distinguish them (Denil and Trappenberg, 2010). In the context of credit card fraud detection, fraudulent and non-fraudulent transactions usually overlap because fraudsters mimic the behaviour of real cardholders in order to fool the FDS. Most research dealing with the data overlap problem considers it to be similar to the class imbalance problem (Li et al., 2021; Vuttipittayamongkol and Elyan, 2020). It was shown in Denil and Trappenberg (2010) that class imbalance and data overlap have interdependent effects on classifier performance. Moreover, the same study showed that, unlike the class imbalance problem, the data overlap causes a linear drop in the classifier performance.
  The most common solution reported in the literature to overcome the class imbalance and overlap problems consists of three main steps. The first step involves dividing the original dataset into an overlapping subset and a non-overlapping subset. Under-sampling is then applied to the overlapping subset to remove samples from the majority class. Finally, a classifier is used to detect the minority samples (Li et al., 2021).
- **Data drift**: Data drift is the main problem that inevitably degrades the accuracy of machine learning over time. It is the change in the input data form that results in the deterioration of its performance. Data drift might be caused by the change in the relationship between the characteristics, or the change in the covariate. This particularly occurs in the detection of credit card fraud due to changing customer behaviour. Indeed, fraudulent behaviours, methods and strategies develop and change over time. Additionally, user behaviour may also change over time, for example, if users spend more on holidays. All these changes cause variation in data distribution between testing and training data, which decreases the system's accuracy and performance (Becker and Becker, 2021). When building a credit card fraud detection model, it is crucial to select and use the proper features from raw data in particular. This could be achieved through the aggregation of raw data transactions to extract the spending behaviour patterns of the cardholder (Correa Bahnsen et al., 2016). This is usually done by aggregating the transactions to observe customers' spending patterns. This behaviour may evolve and such changes need to be considered in the model.

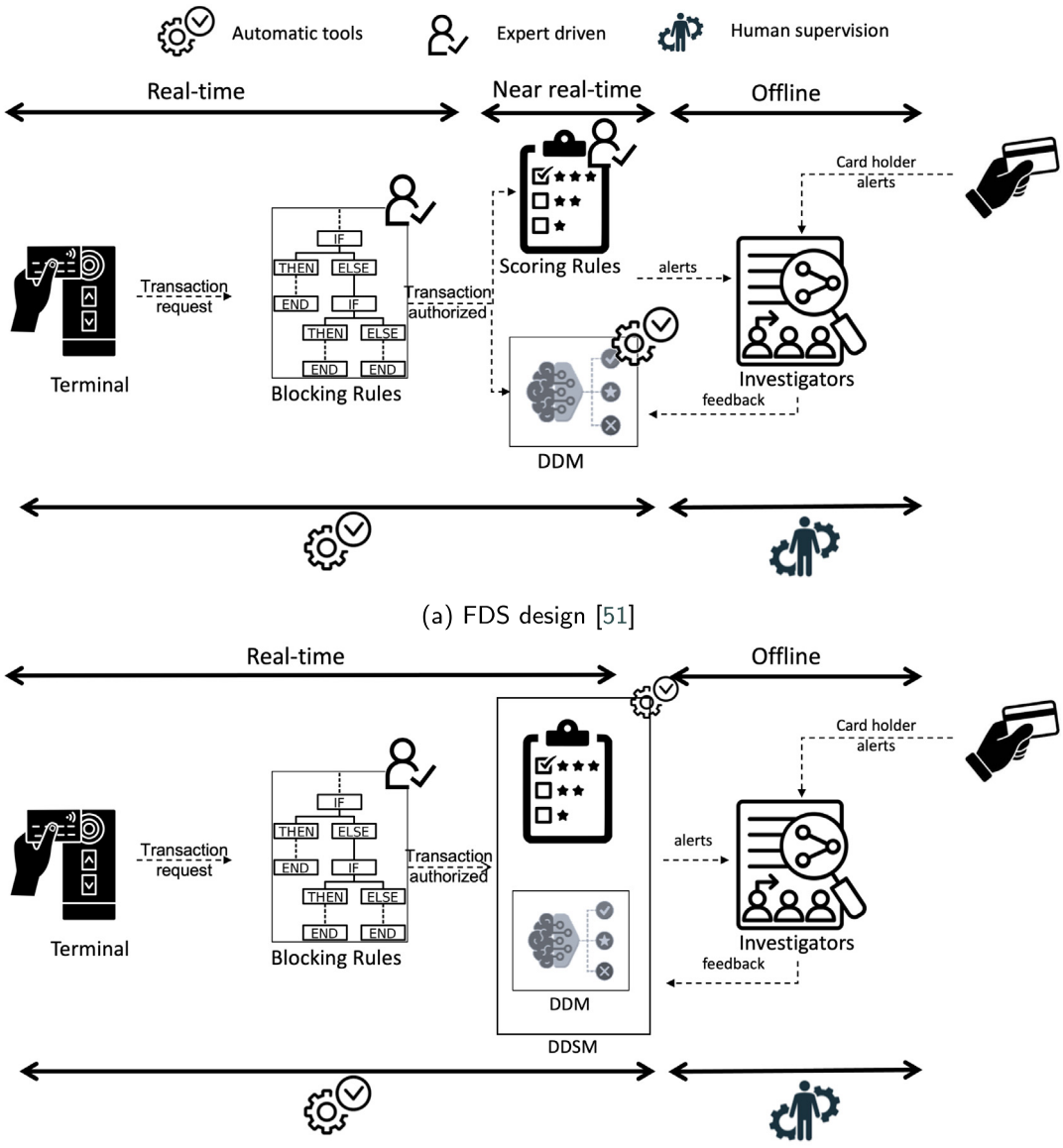

Fig. 7. Credit card system design evolution.

Security-related challenges. The security aspect of credit card transaction systems is a challenging problem. Detection models should preserve the privacy of the users and prevent attackers from learning their behaviour. Indeed, if an attacker acquires enough information about the learning model, he or she can imitate user behaviour to perform fake transactions that would not be detected as fraudulent ones. In Yang et al. (2019), the authors proposed a federated learning framework based on behavioral features. This model allows banks to train a fraud detection model while distributing training data over their local storage. Afterward, a common detection system is constructed by combining the updates that were computed locally by the local detection models. Therefore, involved banks can benefit from the shared model without disseminating their local dataset thus protecting sensitive cardholder data. Furthermore, it is important to devise protocols that prevent fraudsters from compromising the identity of users. For instance, Yang et al. (2021) discusses the interesting problem of the detection of identity fraud in cyber-physical systems (CPS) with clients having limited resources. The authentication between two clients is delegated to and proxied by a trusted authentication server. The authors propose a protocol to detect which authentication server is compromised in such a delegated authentication framework.

Implementation/deployment challenges.

- Distributed implementation: a robust FDS should be distributed to allow for a more robust solution that can be scaled up easily. Banking services have been characterized by a monolithic architecture for decades. Though this centralization is safe, it has several limitations, namely the high cost of scaling up to meet the increasing number of customers. With the emergence of cloud computing, banks have evolved to follow this computing paradigm. Elastic scalability provided by the cloud, in addition to the maturity level it has reached, has allowed many banks to shift to cloud deployment. A good FDS system should comply with the distribution principles to fit the infrastructure of the new generation of banking systems. However, distributing the

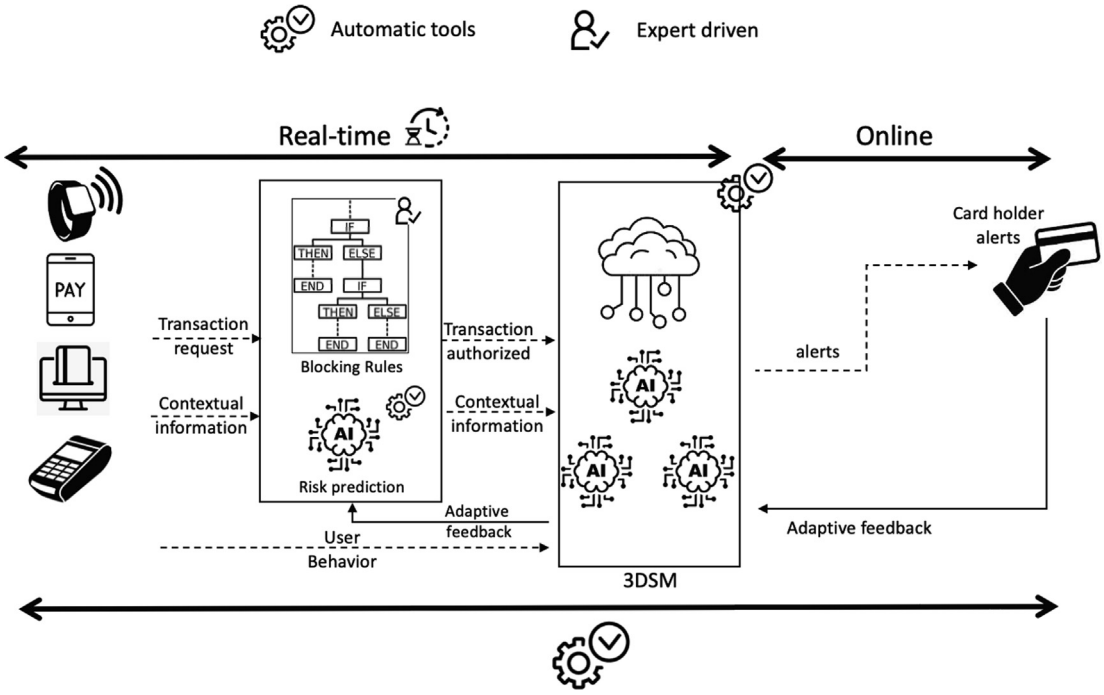

Fig. 8. Credit card system fully automated design.

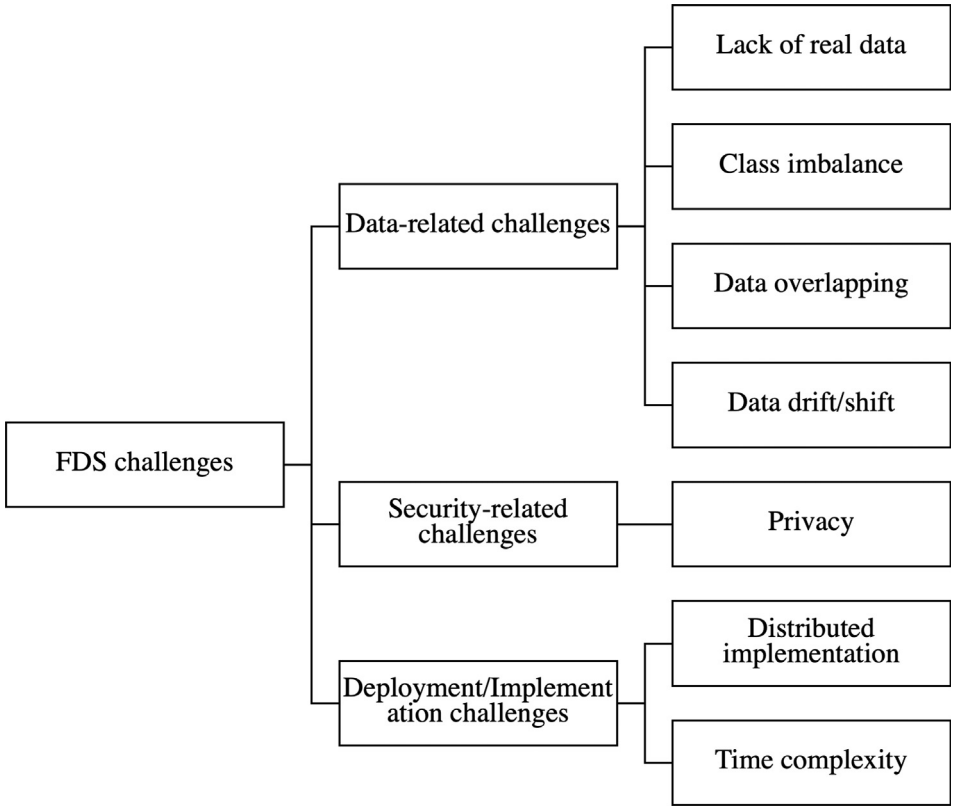

Fig. 9. Credit card fraud detection challenges.

components of FDS is challenging, and many aspects should be taken into account, such as latency, the adoption of new technologies and frameworks, and the integration of heterogeneous data.

- **Time complexity:** The real-time nature of online transactions requires a FDS to make decisions in milliseconds. The current systems usually suffer from a gap between the fraud detection time and the time that fraudulent users are excluded from the platforms. FDS should be able to provide decisions instantly based on streaming big data (Ali Yeşilkanat et al., 2020). A perfect FDS seeks to detect fraud even before the transaction is approved. Such a real-time requirement is hard to achieve as there is an inevitable delay due to information fetching (Dal Pozzolo et al., 2015; Mittal et al., 2020). Nevertheless, this might

also lead to usability issues and customer dissatisfaction if there are many false positives. Fraud detection might be applied during the transaction but should provide high responsiveness. New technologies that deal with streaming data and provide real-time analytics such as Apache Spark (Chambers and Zaharia, 2018) might provide better solutions in terms of responsiveness. Only a few studies have investigated the real-time aspect of fraud detection systems. These are explored in this survey to provide insights for future relevant research. In addition, the evaluation of prediction models has to take into account the responsiveness of the model. To this end, this metric is described in Section 7.

3.3. Machine learning techniques applied to FDS

As a branch of AI, Machine Learning (ML) systematically uses algorithms to synthesize the underlying connections between data and information (Awad and Khanna, 2015). The machine learning models suggested for credit card FDS are either traditional or deep. Moreover, the suggested solutions investigated both supervised and unsupervised techniques. While supervised techniques require labelled data to learn genuine and fraudulent transactions, unsupervised techniques do not require the use of labelled data, which avoids the issue of missing labelled data. Unlike supervised techniques, unsupervised models generally consider credit card fraud detection as an outlier detection problem (Chakraborty et al., 2022).

Domingos (2018) classifies machine learning algorithms into five approaches: symbolists, Bayesians, analogizers, evolutionaries, and connectionists, according to the preferred approach used to build machine learning models.

3.4. Symbolist approach

This approach focuses on logic theory and formal systems. It was a dominant paradigm from the 1950s to the 1980s. Decision-making is mainly directed by logical rules. This approach includes Decision Trees (DTs), Random Decision Forests (RF), production rule systems and Inductive logic programming. In the context of credit card fraud, this kind of machine learning has been largely studied through DT, as well as RF, which is a collective of DTs. To classify new objects based on their attributes, each tree is organized before the tree votes for a given class. The forest selects the class with the maximum votes (Breiman, 2001). Isolation forest is an anomaly detection algorithm developed by Fei Tony Liu (Liu et al., 2008; Liu et al., 2012). It determines the isolation of data, namely how distant a given data point is from the remaining data points. Isolation forest isolates anomalies using binary trees, thus producing a more rapid anomaly detector that finds anomalies directly without profiling all the regular instances. The time complexity of this algorithm is linear with an inferior memory need, which perfectly suits high-volume data (Chandola et al., 2009). Consequently, it allows the modelling of credit card fraud detection when dealing with a large volume of data. To the best of the author's knowledge, this algorithm has been investigated only in Ingole et al. (2021).

3.5. Bayesian approach

The Bayesian approach emphasizes the use of statistics to build models and uses probabilistic inference to predict outcomes. Bayesian ML algorithms include Hidden Markov chains, Naive Bayes (NB), Linear Regression (linR), Logistic Regression (logR), Gradient Boosting (GBT), Adaboost, etc. A Naive Bayesian model is simple and easy to construct and is well-suited for massive datasets. It is well established that NB outperforms several highly sophisticated classification models. It has been widely used in predicting credit card fraud, for example, (Mohammed et al., 2018; Thennakoon et al., 2019; Baabdullah et al., 2020). GBT and AdaBoosting Algorithm are also useful when dealing with massive data. Boosting is an ensemble learning technique that integrates the predictive power of many base estimators to improve the robustness of the prediction models. These algorithms have been investigated in several works addressing credit card fraud detection, such as (Padmanabhuni et al., 2019; Lucas et al., 2020; Ali Yeşilkanat et al., 2020).

3.6. Evolutionary approach

Evolutionary algorithms have their roots in biology and focus on steps and iterations rather than outcomes. When considering machine learning from the perspective of a biologist, it is the development of artificial intelligence optimization algorithms that is of most interest. Evolutionary machine learning models use genetic algorithms and evolutionary programming. These approaches are generally used to optimize machine learning models, either by optimizing the prediction of classifiers or through feature selection. In particular, the use of swarm intelligence in credit card fraud detection was investigated in terms of feature selection in Han et al. (2021).

3.7. Analogy-based approach

Analogies are based on the creation of feature classes. This is the approach psychologists prefer to use to solve machine learning problems. If data, old or new, is specified as one of the problem categories, the prediction of the result for that data will depend on its relationship to a particular category. The problems addressed through this approach are recommendation systems. Analogists use categories and types to identify groups of data points and predict future outcomes based on the results of other class members. The machine learning models under the analogies category are K-Nearest Neighbor (KNN) and SVM, which are unsupervised learning techniques, used to place members into their classes. The idea behind unsupervised learning is to construct a model that is able to correctly differentiate between fraudulent and genuine transactions based on the model's internal self-organization, which grasps patterns as probability densities or a combination of neural feature preferences. Unsupervised learning models the regularities in raw data, and could be used for anomaly detection by assuming that genuine transactions are more frequent than fraudulent ones in the test data. Though largely tested for credit card fraud detection, both KNN and SVM are computationally expensive and may show reduced performance in detecting credit card fraud for large datasets.

One of the unsupervised analogy-based solutions suggested in the literature to detect fraudulent transactions is the use of recommender systems. A recommender system, also known as a recommendation system or engine, is an information filtering system aiming at predicting the preference/rating that an end user may assign to an entity (Shapira, 2015). This type of system was originally designed to help recommend items to users and is thus useful in several contexts, such as social media platforms and open web content recommenders. Some studies have solved fraud detection by modelling the costumer as a user and his/her behaviour as an item before attaching the label (fraud/genuine) as the corresponding rating (Cui et al., 2021). By using collaborative filtering (Schafer et al., 2007), it was then possible to detect fraudulent transactions. Another unsupervised analogy-based technique is the Local Outlier Factor (LOF) (Breunig et al., 2000). LOF has been applied to the credit card fraud problem to detect fraudulent transactions as outliers (Ingole et al., 2021), identified based on local neighbourhoods.

It is a density-based technique that uses nearest neighbor search to identify anomalous points (Kotu and Deshpande, 2019). Finally, agglomerative clustering of the distance matrix has been used to solve the problem of datashift in the detection of credit card fraud (Lucas et al., 2019). It aims to classify a group of objects (such as notes or individuals) into subgroups called clusters with similar properties (Zepeda-Mendoza and Resendis-Antonio, 2013). This unsupervised approach has been used to solve the datashift problem in credit card fraud detection (Lucas et al., 2019).

3.8. Connectionist approach

Models in this category are mainly Artificial Neural Network (ANNs) (also known as NNs). An ANN is a model inspired by the biological neural networks of the human brain. This approach includes models such as ANN, reinforcement learning, and DL. In Connectionism, an input may have an output; but the path taken to get to that output is hidden. ANN comprises an interconnected set of artificial neurons. Generally, an ANN is adaptive, i.e. it changes its internal structure based on the external or internal information flowing through the network structure during the learning phase. Modern neural networks are nonlinear statistical data modelling tools. They are used to model complex associations between inputs and outputs or to discover hidden patterns in datasets (Daniel, 2013). ANNs have been used both as supervised and unsupervised techniques to solve the credit card fraud detection problem.

Feed-Forward Neural Network (FNN) is the first and most straightforward kind of ANN (Schmidhuber, 2015) in which the information flows in exclusively one path: from the input nodes, through the hidden nodes (if any), before reaching the output nodes. It is worth noting that a FNN structure does not contain cycles or loops. A Probabilistic Neural Network (PNN) is a special kind of FNN with a complex structure widely used in classification problems (Mohebali et al., 2020). Donald Specht proposed it (Specht, 1990; Specht, 1990) and it is based on Bayes theory and derived from Bayesian networks. It comprises an input layer, a pattern layer, a summation layer and an output layer.

DL is a subset of machine learning. It differs from classical machine learning by the data type it deals with and the learning methods it follows. According to Goodfellow et al. (2016), a DL network is a neural network with two or more hidden layers. Indeed, adding more hidden layers allows accuracy to be improved and refined. A network with only a single hidden layer is conventionally called “shallow”.

Convolution Neural Network (CNN) (aka ConvNet) is a class of deep networks that are most commonly applied to various computer vision tasks (Yamashita et al., 2018). CNN is conceived to automatically learn spatial feature hierarchies through back-propagation. It uses many building blocks including convolution layers, pooling layers, and fully connected layers. Though CNNs are mainly applied in image and video recognition as well as medical image analysis, has also been used to detect fraudulent transactions in Babu and Pratap (2020, 2021).

A Recurrent Neural Network (RNN) is a kind of deep ANN that recognizes the sequential characteristics of input data that uses patterns to predict the next likely scenario. It is generally used in speech recognition and natural language processing. Recurrent neural networks Unlike feedforward neural networks, RNNs use feedback loops to process a data sequence, thus allowing information to persist. This connects inputs and allows RNNs to deal with sequential and temporal data. This might be very useful to model spending behaviour and recognize unexpected behaviour as fraudulent transactions. The main problems with RNNs are gradient vanishing and exploding problems (Yue et al., 2018). The former occurs due to errors made during the training phase. If the gradients start to explode, the neural network becomes unstable and cannot learn from training data.

Learning to store information over extended time intervals by recurrent back-propagation is time-consuming (Hochreiter and Schmidhuber, 1997). This is solved by Long Short-term Memory (LSTM) which has a computational complexity per time step and weight equal to 0.1 (Hochreiter and Schmidhuber, 1997). Unlike FNN, LSTM has feedback connections. Thus, it processes not only single data points like images, but also sequential data such as speech or video. LSTM was applied to credit card fraud detection in Jurgovsky et al. (2018) to assemble a spending purchase profile of cardholders. This enhances the accuracy of fraud detection for new input transactions. It was found that an LSTM can model the hidden sequential patterns and the transaction history, thus improving fraud detection. However, LSTMs are weaker than standard feed-forward networks.

The Gated Recurrent Unit (GRU), one of the most common powerful RNN models, was proposed in Cho et al. (2014). It is considered to be an improvement and simplification of LSTM because it has fewer gates (half the number of gates in LSTM), by combining the input and forget gate into one gate and the hidden and cell state into a single state. GRU has a simplified structure as it uses fewer parameters than LSTM, which gives better performance. Experimental results show that GRU works better for intrusion detection systems (Xu et al., 2018). Combining LSTM and GRU allows the resulting network to benefit from the strengths of both units. Thus, it can learn long-term associations and short-term patterns. Indeed, the GRU is similar to a long short-term memory (LSTM) with a forget gate. It has fewer parameters than LSTM, as it does not have an output gate. Moreover, (Roy et al., 2018) showed that LSTM and GRU highly outperform the baseline artificial neural network and were able to properly model the order of transactions of an account. This represents valuable knowledge, allowing fraudulent and non-fraudulent transactions to be distinguished. transactions.

Autoencoder (Baldi, 2011) is an artificial neural network used for unsupervised learning. It compresses the input data into a lower dimension representation in a way that the output can be reconstructed from this compressed representation. Recently, it has been used for feature extraction, dimensional reduction and fraud detection because it can detect fraudulent transactions even if there are few fraudulent trained data (Chaquet-Ulldemolins et al., 2022). It also shows good performance in analyzing real-time, large-scale and complex data (Fan et al., 2018).

DeepWalk (Perozzi et al., 2014) is a scalable algorithm for learning node embedding that is used to represent vertices in a graph network to vectors based on social representation. It takes the network graph as an input and then uses a random walk to turn the graph into a sequence of random paths of connected nodes to be trained using the Word2Vec model. It is composed of two parts, a random walk generator and an update procedure, and is commonly used in social network detection. The random walk generates a random vertex based on the graph starting from the root until it reaches the maximum length (t). Additionally, the update procedure includes Skip-Gram to update the nodes embedding by maximizing the similarities. Another algorithm for embedding graph networks is Node2Vec. This algorithm is a semi-supervised mechanism to represent nodes in a network. It is similar to the DeepWalk algorithm; however, the generation of random walks is different (Grover and Leskovec, 2016).

4. Traditional machine learning based models

The following section describes studies addressing credit card fraud detection based on traditional techniques. These works are classified according to their approach into supervised and unsupervised techniques. A summary table is provided at the end of each subsection to highlight the main contribution of each category.

4.1. Supervised techniques

In Correa Bahnsen et al. (2016), the authors focused on feature engineering in the context of credit card fraud detection using aggregated features and periodic variables. They extend the transaction aggregation strategy and create a new feature set based on examining the recurring behaviour of the time of a transaction using the von Mises distribution (Mahmoudi and Duman, 2015), which is a continuous probability distribution close to the wrapped normal distribution around a circle. They compare state-of-the-art credit card fraud detection models (DT, logR, RF) against their previously suggested improvements based on the Bayes Minimum Risk (BMR) (BMR Decision Tree (BMR-DT), BMR Logistic Regression (BMR-LR), and BMR Random forest (BMR-RF)) (Bahnsen et al., 2014) and evaluate how the different sets of features affect the results, in addition to the cost-sensitive RF and cost-sensitive linR (Bahnsen et al., 2014). They demonstrated that periodic features increase savings (see Section 7) by 13% and increase performance by more than 200%.

Wang et al. (2018) presents a detection system based on Back Propagation Neural Network (BPNN). The model relies on swarm intelligence namely the whale population optimal solution (Mirjalili and Lewis, 2016), to find the initial weights and threshold of the neural network training model. The used neural network contains 2 layers of input and 2 layers of output with 20 hidden layers and it has been tested on 500 samples of a European dataset. The whale algorithm is used to find the initial optimal values of the model while the BPNN is used to correct and adjust these values in order to output the optimal model parameters. This technique solves some problems related to BPNN, such as network defects, slow convergence rate, and poor system stability. The main advantage of this study is that the neural network model is optimized using swarm intelligence to improve accuracy.

Mohammed et al. (2018) presents experiments comparing many ML techniques. Scalability was tested when working with highly imbalanced massive datasets. RF, Balanced Bagging Ensemble (BBE) with DT, and Gaussian Naïve Bayes (GNB) were used. The results show that many detection algorithms performed well with medium-sized datasets but struggled to maintain similar predictions when dealing with massive data. They deduced that a balanced bagging ensemble with the hybrid balancing technique has superior prediction and that RF is scalable and capable of detecting fraud with good accuracy. Another use of bagging was suggested in Akila and Srinivasulu Reddy (2018) in the form of a cost-sensitive Risk Induced Bayesian Inference Bagging model termed RIBIB.

Padmanabhuni et al. (2019) suggest ensemble learning composed of five individual machine learning classifiers: SVM, KNN, logR, NN, DT. They compared the final result with individual ML models in addition to other learning techniques such as Adaboost, RF, NN and Probabilistic Neural Network (PNN). The suggested model has higher accuracy than the other models considered in the study (82.47%) but has lower sensitivity than Adaboost, RF, NN, and PNN.

Thennakoon et al. (2019) proposes a real-time fraud detection system that uses Principal Component Analysis (PCA) for dimensionality reduction. The authors also used 10-fold cross-validation. To test their model, they used a private dataset containing fraudulent transaction logs combined with non-fraudulent transactions. The model was tested with four machine learning algorithms (SVM, NB, KNN, and logR). However, it is not clear how the real-time feature was achieved, and no testing was carried out to prove the transactions were checked in real time.

In Kim et al. (2019), the authors introduced a framework to compare the champion and challenger models, where the former is a hybrid ensemble model used for a credit card fraud detection system in a partner bank, and the latter is based on deep learning (see Section 5). The ensemble model is based on decision trees, logistic regressions, and shallow neural networks. The proposed model was evaluated in a real system. Additionally, the authors compared both models (champion and challenger) using offline testing and post-launch on a real dataset. Testing results showed that the best performance was achieved using the deep learning-based model (challenger).

In Rtayli and Enneya (2020), the authors proposed a credit card fraud detection based on a hybrid approach using the Support Vector Support Vector Machine-Recursive Feature Elimination (SVM-RFE) method to select the predictive features, and the Hyper-Parameters Optimization (HPO) method to estimate the best hyperparameter values for the RF. The approach was validated based on three datasets and the results showed high accuracy performance compared to existing models.

Ali Yeşilkanat et al. (2020) used a Gradient Boosting Tree (GBT) model to model a real-time credit card fraud detection system. The system works on streaming CNP transactions. For feature engineering, the authors combined many features, including numerical, hand-crafted numerical, categorical, and textual. The main contributions are the aggregation of categorical values and the use of the word embedding model. In addition, the authors used the sliding window approach in a given time frame to resolve the issue of data drift. The suggested system comprises two training modules: offline training and real-time detection. This scheme is used for real-time detection and represents different types of features using the character-level word method for each merchant. However, it does not take into account the time sequence of merchant and card interaction.

Olowookere and Adewale (2020) proposes a framework that uses both meta-learning ensemble learning and cost-sensitive learning to detect fraudulent transactions. In the first phase, the principle of the meta-learning ensemble is exploited to combine the predictions of three classifiers, namely the KNN, the DT and the multilayer perceptron. In the second phase, the Cost-Sensitive Logistic Regression algorithm is used. It includes cost-sensitive learning in the training phase of the meta-classifier. This is achieved by considering the misclassification cost of each transaction instance based on an instance-based cost matrix. Finally, to classify the new incoming transactions belonging to the test data, the base classifiers are first used to classify each new transaction. The resulting classifications are then sent to the cost-sensitive ensemble classifier to mark the final classification as either fraud or non-fraud. Although this framework may require heavy computations, the cost-sensitive ensemble classifier performs better than pure ensemble methods under high class-imbalance, according to the reported results.

In Sudha and Akila (2021), the authors proposed a system that uses users' operational and transaction features to detect suspicious operations. The system uses RF and M-class Support Vector Machine (M-class SVM) in order to classify the collected data as legitimate or suspect. It consists of two phases, the first of which uses RF to classify users' operation features when they make payments (for example time, display balance, change password) into legitimate and fraudulent. Second, it uses M-class SVM to classify transaction features, such as account number, amount, and type, to detect fraud actions. The model is easy to implement, but the performance of the system can degrade if there are many trees in the RF that slow the algorithm and make it unsuitable for real-time detection (Speiser et al., 2019).

Tran and Dang (2021) solved the data imbalance problem (see Section 6) and and compared four machine learning algorithms (RF, logR, KNN, DT) to evaluate the performance of the obtained balanced dataset. The results showed better performance with a balanced dataset than with a skewed imbalance dataset. Similarly, (RB and KR, 2021) investigated the use of multiple algorithms, such

as SVM, KNN and ANN, to detect fraudulent credit card transactions. A comparison of the algorithms led to the conclusion that the ANN algorithm outperforms the other machine learning algorithms based on accuracy. The suggested model is however very simple, with no details in terms of the data preprocessing, normalization and sampling. The testing was only based on accuracy, precision and recall, and it is important to consider additional metrics for verification.

In Sudha and Akila (2021), the behaviour of users of a banking website was used to classify transactions as either genuine or fraudulent. The operational and transactional features are combined into a single feature in a model with three phases. First, the user behaviours are collected and analyzed using the Web Markov Skeleton Process (based on a Markov chain). The probability matrix of the transitions from and to the pages of the website is estimated, as well as the time spent on each page. Secondly, the operational features could be operation type, operation mode, operation device, operation time, source IP address, or location information. These features are then passed to the RF algorithm to classify the operation as either honest or suspect. The transaction features include transaction type, transaction device, transaction amount, account number, and balance after a transaction. They are sent to the M-class SVM classifier to classify the transaction as honest or suspect. Lastly, the majority voting ensemble classification is used to accurately predict fraudulent transactions. However, the execution time is not specified. Moreover, the model relies on the analysis of individual user behaviour, which may be computationally expensive. In addition, the authors did not provide a performance comparison with state-of-the-art techniques. Another question that arises is whether analyzing user behaviour while visiting the bank's website contributes to improved fraud detection.

In Lucas et al. (2020), the authors used the features of credit card fraud detection as a sequence of transactions rather than individually. For feature generation, Hidden Markov Models (HMM) were used in the construction of historical features and the construction of multiple perspectives used HMM-based features. These features were applied to Random Forest and the results increased the precision-recall AUC to 18.1%. To validate the proposed solution, the authors also applied the new feature engineering method to Adaboost and logistic regression classifiers and showed the stability of this increase. The main benefit of this work is the improvement of the feature engineering phase.

Baabdullah et al. (2020) proposed a credit card fraud detection approach to solve problems related to imbalanced data using different techniques. This model uses two levels for data resampling (oversampling and undersampling). Additionally, to improve classification performance, several supervised algorithms, namely logR, NB, DT, RF, and KNN, are applied over the resampled datasets. Finally, to evaluate the performance, the performance evaluation metrics for the five classification models were compared with those from different datasets.

Bagga et al. (2020) evaluates and compares the performance of nine techniques (logR, NB, RF, KNN, Multi Layer Perceptron (MLP), Adaboost, Quadrant Discriminant Analysis, etc.) with pipelining and ensemble learning on the credit card fraud detection problem. The pipelining was used for feature selection. Ensemble learning was used through the bagging classifier. To overcome the imbalanced data the ADASYN method was applied. However, the model has an accuracy of 100%, which may be due to an overfitting problem. More testing is required with different datasets.

In Cochrane et al. (2021), the authors presented a pattern analysis for attributes related to fraud with a view to applying indicators to increase the accuracy of fraud detection systems. It was found that some attributes are strongly correlated with fraud activities, showing that a transaction is fraudulent. After building the correlation chart, this information can be applied to the training model. Two public datasets that simulate transactions were used to train and test the model. Furthermore, three machine learning algorithms were used on both datasets (DT, linR, logR) to build the predictive and prescriptive analysis. The results from the predictive models were then combined to determine whether or not the transaction is fraudulent. However, further experiments with private financial data are needed in order to improve the accuracy of fraud detection and to increase the speed of the system.

Ingole et al. (2021) proposed a credit card fraud detection system that focuses on orchestrating various services using the Oracle SOA suite. They tested different machine learning models (supervised, unsupervised, traditional, and deep). The supervised traditional classifiers tested were SVM, isolation forest, random forest regressor, local outlier factor (LOF), and NNP. The Oracle SOA suite model has been deployed on the Google Cloud Platform (GCP) to provide an online solution. The authors dropped the time attribute in the preprocessing phase, a choice justified by the low impact on the predicted result. However, the time attribute might be of considerable importance in detecting fraud as it is part of user behaviour. Furthermore, although the model is implemented in a cloud-based environment, it does not provide a real-time solution.

The authors in Han et al. (2021) proposed a system called INUM, which is an information utilization method. INUM can be used to assist any multimodal multiobjective evolutionary algorithm MMEA in solving multimodal multiobjective problems MMOP. INUM works by extracting a decision vector from the optimal one, in order to produce elite solutions. This method enhances performance in objective and decision spaces. It has been used in credit card fraud detection by optimizing feature selection combined with extreme learning, namely WELM (Weighted Extreme Learning) which is a particular type of single-layer feed-forward NN. The authors used MRPS (MO_Ring_PSO_SCD) (Yue et al., 2018) which is an improved particle swarm optimization algorithm. The main benefit of this solution is to take decision space and objective space into consideration. Furthermore, it can be used to enhance the performance of any multi-modal multi-objective evolutionary algorithm. However, the model achieves near values of the G-mean and mean run time of the other multi-objective evolutionary algorithm. Moreover, the growth rate (large population and more computational budgets), gives low results while dealing with some imbalanced multi-modal multi-objective problems.

Table 4 summarizes the previously discussed works and highlights the key concepts and ML algorithms used.

4.2. Unsupervised techniques

In Soltani Halvaiee and Akbari (2014), an AIS-based Fraud Detection Model (AFDM) was proposed that uses an artificial immune system algorithm. The authors also parallelized computation and used Hadoop to distribute the calculation, which increased accuracy by 25%. This paper proposes using cloud computing by deploying a fraud detection system on a cloud-based file system, namely Hadoop, thus achieving data parallelization. However, parallelization focuses more on the computation cost rather than the accuracy of detecting fraudulent transactions.

Srivastava et al. (2016) presents a fraud detection system based on ANN. The model is trained on past transactions based on the transaction parameters and user profile before outputting a probability value. According to this value, the output is classified into one of four classes (non-fraudulent, doubtful, suspicious, and fraudulent). Whilst the suggested model is simple and easy to understand and implement, no details about the dataset, normalization phase, and testing were provided. The model was trained only once, and it is not clear how it can take into account the data drift issue. In addition, the comparative study of the literature was not justified.

Table 4
Summary of supervised traditional machine learning-based models for credit card fraud detection.

Ref#	Key aspects of the suggested solution					ML algorithms												
	Optimization	Feature engineering	Real-time aspect	Behavior	Scalability	NB	KNN	LogR	RF	DT	SVM	NN	PNN	BPNN	adaboost	LinR	GBT	MLP
(Correa Bahnsen et al., 2016)		aggregated features and periodic variables		periodic behavior				✓	✓	✓								
(Wang et al., 2018)	Whale													✓				
(Mohammed et al., 2018)					yes	✓			✓	✓								
(Akila and Srinivasulu Reddy, 2018)						✓												
(Padmanabhuni et al., 2019)							✓	✓	✓	✓	✓	✓	✓		✓			
(Thennakoon et al., 2019)		PCA	yes			✓	✓	✓			✓							
(Kim et al., 2019)		step-wise feature selection						✓		✓		✓						
(Rtayli and Enneya, 2020)		SVM-RFE							✓		✓							
(Ali Yeşilkanat et al., 2020)		Combination of numerical, hand-crafted numerical, categorical and textual attributes	sliding window														✓	
(Olowookere and Adewale, 2020)		meta-learning ensemble techniques and cost-sensitive learning					✓			✓								✓
(Sudha and Akila, 2021)		operational and transaction features							✓		✓							
(Tran and Dang, 2021)		class imbalance					✓	✓	✓	✓								
(RB and KR, 2021)							✓				✓	✓						
(Sudha and Akila, 2021)									✓									
(Lucas et al., 2020)		HMM						✓	✓						✓			
(Baabdullah et al., 2020)						✓	✓	✓	✓	✓								
(Bagga et al., 2020)		pipelining				✓	✓	✓	✓						✓			✓
(Cochrane et al., 2021)		correlation						✓		✓						✓		
(Ingole et al., 2021)		dropping the time attribute							✓		✓	✓						
(Han et al., 2021)	MRPS	optimized feature selection										✓						

In Cui et al. (2021), the authors presented an anomaly detection mechanism to detect online banking fraud. The system is designed to solve problems related to previous fraud detection systems namely the limited amount of historical data for users, the skewed nature of transaction data and the absence of a uniform way of treating users' attribute values. The authors presented a ranking metric embedding system called (ReMEMBeR), which uses multi-contextual behaviour to reduce the false positive and error rate. They designed a pseudo recommender system, which deals with individuals as pseudo-users and their behaviours as pseudo items to solve historical user data problems. The system utilizes collaborative filtering to benefit from other similar users' behaviours. The ranking method (legitimate/fraudulent) is based on mapping the pseudo-user to the pseudo item in terms of whether he/she likes or dislikes this. The performance of the system was tested under four concepts: real-world transaction, skewed data, model combination (with the machine learning algorithms SVM, RF, NN2L, logR) and multiple contextual behaviours.

Lucas et al. (2019) presents a method for credit card fraud detection based on quantifying the covariate shift (i.e. differences in behaviours). In this method, each day's transactions are classified against other days to test efficiency. If the classification is efficient, then the days are similar; otherwise, it means that there is a covariate shift between these days. To characterize data, they are presented in a distance matrix to clarify the covariate shift. The RF algorithm is then used to cluster the distance matrix. After clustering, the authors observed four similar clusters that differ from others, namely: working days, school holidays, Saturdays and Sundays. The data shift was integrated as a new feature in the credit card fraud detection system. The main advantage is that the proposed model solves the problem of the variety in users' purchase behaviours and fraudulent mechanisms, since it changes over time. However, the dataset used is relatively old (from 2015), whereas fraudulent strategies have evolved over the past seven years. Furthermore, the results show only a slight difference when compared with and without the application of the covariate shift as a feature.

Ingole et al. (2021) used unsupervised models, amongst other techniques, to solve the credit card detection problem as an outlier detection problem. They used Isolation Forest, which is similar to the random forest model, and built upon DT, but allowing for outlier detection. They also used the Local Outlier Factor, which is based on calculating an anomaly score that measures how isolated a sample is from its neighboring sample.

Table 5 summarizes the supervised techniques suggested for credit card fraud detection and highlights the key concepts, the real-time aspect if any and the consideration of behavior features algorithms used.

5. Deep learning models

Few deep learning-based solutions for fraud detection have been suggested recently, and most of these studies investigated supervised deep models.

The authors of Fu et al. (2016) proposed a supervised model based on CNN to capture the intrinsic patterns of fraud behaviour using real-world massive transactions of a major commercial bank. The model has two phases: the training phase (offline) and the prediction phase (online). In the training phase, the features are extracted and transformed from the raw data within a fixed time period. Several features are extracted based on previous users' transactions, but these traditional features by themselves cannot be useful for representing sophisticated consumer patterns. As a result, the authors introduce a new feature called trading entropy, which is designed to represent the relationship between the user's transaction and total transaction amount during a specific time interval. Though the authors claim to have achieved superior per-

Table 5
Summary of unsupervised traditional machine learning techniques used for credit card fraud detection.

Ref#	ML algorithm	Key concept	Real-time aspect	Behavior
(Soltani Halvaiee and Akbari, 2014)	AIS	Cloud computing & Parallelization	Hadoop MapReduce	
(Srivastava et al., 2016)	NN	User profile		past behavior
(Cui et al., 2021)	recommender system & SVM/LR/RF/NN2L	ranking metric embedding system		collaborative filtering & multi-contextual behavior
(Lucas et al., 2019)	Agglomerative clustering + RF	quantifying the covariate shift as a new feature		clustering according to the days
(Ingole et al., 2021)	Isolation Forest (DT) & LOF	Anomaly detection		

Table 6
Summary of deep learning models for credit card fraud detection.

Ref#	Key concepts	Type	DL algorithm									
			CNN	LSTM	GRU	FNN	ANN	RNN	Auto-encoder	Deep Walk	Node2 Vec	DBN
(Fu et al., 2016)	features matrix	supervised		✓								
(Jurgovsky et al., 2018)	aggregate historical behaviour	supervised		✓								
(Roy et al., 2018)	new features	supervised		✓	✓		✓	✓				
(Kim et al., 2019)						✓						
(Babu and Pratap, 2020)		supervised	✓									
(Forough and Momtazi, 2021)		supervised		✓	✓	✓						
(Kewei et al., 2021)		supervised					✓					
(Ingole et al., 2021)	SOA architecture	supervised unsupervised	✓						✓			
(Zhou et al., 2021)		semi-supervised								✓	✓	
(Zhang et al., 2021)	Hierarchical feature selection	unsupervised	✓					✓				✓

formance compared with some state-of-the-art methods, the F1-score is very low and the detection accuracy needs to be increased Choi and Lee (2018). In addition, more metrics are required to better evaluate the model.

Roy et al. (2018) assessed the efficacy of four deep learning algorithms: RNN, GRU, LSTM, and ANN. Feature engineering was performed by adding new features to the ones provided in the dataset. It is worth noting that the dataset contains 80 million anonymous transactions collected during 8 months and provided by a financial institution. The sensitivity metric was used to discover the hyper-parameters that have the highest impact on the model performance. The authors concluded that the model performance is affected by the network size: the larger the network, the better the performance. Moreover, the performance analysis showed that LSTM and GRU outperform the baseline ANN model. However, according to Kim et al. (2019), the evaluation mechanism suggested in Roy et al. (2018) is not applicable to real-world fraud detection systems due to their complexity and several constraints.

Kim et al. (2019) suggested a champion-challenger framework in which the champion model is an ensemble learning model while the challenger is a deep-learning-based model. The authors performed parameter tuning and tested several models. They selected the feed-forward architecture and ignored recurrent neural networks because the average number of transactions per card is around 10 and 62% of cards have fewer than five transactions. The testing results show that the challenger model outperforms traditional ensemble learning.

Babu and Pratap (2020) suggested using CNN to solve the fraud detection problem. The authors used a public dataset that already relies on PCA to hide sensitive data. Though the model achieved a high level of accuracy, it did not solve the skewness problem. This model is simple and fast in terms of computation time as it does not use any derivative variables or high-dimensional inputs. The model was tested with and without the max pooling layer, concluding that the accuracy decreases when the max pooling layer is applied. Meanwhile, only the accuracy was considered to assess the model while the dataset is highly imbalanced which may bias the results. Thus, more appropriate metrics should be applied to check the performance of the solution. Moreover, no comparison with similar works was provided.

In Forough and Momtazi (2021), the authors proposed an ensemble model for credit card fraud detection using two sequential classification models, LSTM and GRU networks. The outputs of these models are aggregated and fed to a multi-layer FNN as a voting method to predict fraud transactions. The authors claimed that this approach is the first to use an ensemble of sequential models in fraud detection. The model was validated based on two real datasets and compared against two baseline deep sequential models, solo LSTM and solo GRU, and a third ensemble model from the literature. The results showed that the proposed ensemble method outperforms the previous three models. The authors did not discuss how to handle imbalanced data by using the Area Under the Curve of Precision–Recall (AUC-PR) in their comparison. This was selected for its applicability to highly class-imbalanced settings (see Section 7).

Another scheme using LSTM to perform sequence classification is proposed in Jurgovsky et al. (2018). It utilizes the LSTM network to aggregate the historical behaviour of users in order to improve the accuracy of fraud detection. This model was tested with a real dataset and compared with RF. The results showed that LSTM improves fraud detection and shows good performance in face-to-face transactions. However, it does not improve online e-commerce transactions (Lucas et al., 2020). Additionally, this solution requires manual feature construction before classification, which may not be suitable for large-scale and complex real-world detection systems (Cheng et al., 2020).

In Kewei et al. (2021), the authors used a deep learning architecture that comprises three hidden layers consisting of 512, 256, and 32 neurons. They combined several techniques, including feature engineering, memory compression, mixed precision, and ensemble loss to boost the performance of the suggested model. They used aggregated features and emphasized the hour feature as it impacts user behaviour. Experiments show that the suggested model outperforms traditional machine-learning-based methods that were tested, such as Naive Bayes and SVM.

Ingole et al. (2021) used CNN to classify fraudulent transactions as part of a service-oriented system. Though CNN is usually used on image datasets, the authors also showed that it gives good results on FDS. Ingole et al. (2021) also used autoencoders as an unsupervised deep learning classifier in their model which comprises several other classifiers. Autoencoders were used for anomaly detection where it is trained with genuine transactions. When a fraudulent transaction is fed into the model, it fails to regenerate at the output layer, and the transaction is thus considered fraud. The final solution is implemented using REST-API. However, REST services have security concerns compared to SOAP services.

In Zhou et al. (2021), the authors proposed an Internet financial fraud detection for big data using Node2Vec, which is a semi-supervised graph algorithm to represent nodes and edges. Node2-Vec represents the features in a low-dimensional vector using random walk. The large dataset was processed on Apache Spark GraphX clusters and Hadoop. The proposed approach consists of four main modules:

1. Reprocessing to clear data, remove redundancy and remove empty fields;
2. Sampling normal data features by dividing the data into parts and analyzing each part to collect the normal features;
3. Graph embedding using the Node2Vec algorithm to represent the topological features in a low-dimensional vector in the network graph (implemented in Apache Spark GraphX);
4. Prediction that performs the classification process with a deep neural network to output the final results.

To evaluate the experimental results, three machine learning algorithms were compared: SVM, DeepWalk, Node2Vec. The main advantage of the model is the lower-dimensional representation it provides using graphs and the distribution aspect. In addition, the graphs algorithm can detect high fraud risk features. Finally, the model only detects between 60% and 69% of fraudulent samples. However, Node2Vec is a static graph; once a new node is added, it needs to be repeated on the whole graph to embed the new node (Kazemi and Abhari, 2020). Moreover, node topology affects the node embedding and not node attributes or properties (Kazemi and Abhari, 2020).

A feature engineering framework based on homogeneity-oriented behaviour analysis (HOBA) is proposed in Zhang et al. (2021). The analysis of credit card-based transactions commonly uses the Recency-Frequency-Monetary (RFM) principle. In Zhang et al. (2021), the authors suggest adding the geographical location of the transaction to the RFM, giving rise to the RFM-Location (RFML) principle. Furthermore, as the transactions are heterogeneous (for example purchase transactions and cash withdrawals), the four above customer behaviours are analyzed jointly with a given characteristic (for example the monetary value of a given cash withdrawal). Two strategies are used to fulfill HOBA: a transaction aggregation strategy (aggregation characteristic, aggregation period, transaction behaviour measure based on RFML, aggregation statistics) and a rule-based strategy to generate some categorical feature variables. The combination of the four aggregation elements gives rise to 160 possible feature values. With this number of features, the model would be too complex to update

for a real-time fraud detection system. Thus, the authors suggest using hierarchical feature selection. They construct a Deep Belief Network (DBN) by stacking many restricted Boltzmann machines (RBM) before training the first layer of the RBM. The second layer, Bernoulli-Bernoulli RBM, is then constructed by treating the hidden layer of the first RBM as the visible input layer. In the same way, more RBMs can be used layer-by-layer to construct a DBN, where the hidden layer of a lower RBM is used as the visible layer of the RBM of the next layer. The main benefit of this layer-by-layer greedy training approach is that it requires no class labelling. A final layer, which represents the class labels of the training data, is added to the DBN to deal with classification problems. CNN and RNNs are also used in this study.

Table 6 summarizes the deep learning techniques used for credit card fraud detection and highlights the key concepts, the type (supervised/unsupervised) and the algorithms used.

6. Class imbalance solutions

Data imbalance occurs when the dataset has an unequal distribution of classes. The class imbalance can be an inherent property or caused by difficulty in collecting data due to the high cost, privacy concerns, and effort required (Elrahman and Abraham, 2014). Credit card transactions commonly comprise an imbalanced dataset, including very few fraudulent transactions compared to genuine ones.

To overcome this issue, a range of methods are suggested, namely, under-sampling (Yen and Lee, 2006) and oversampling (Chawla et al., 2002). However, these solutions are challenging because the credit card datasets are highly imbalanced, and instances of the dataset individually carry relevant information (for example transactions owned by the same cardholder) (Ali Yeşilkanat et al., 2020).

Another solution to the class imbalance problem is the use of cost-sensitive learning. It is a sub-field of machine learning that focuses on using models on data that have uneven penalties or costs when making predictions (Balaji et al., 2011). The costs represent the penalty associated with an incorrect prediction.

Many studies have addressed the class imbalance problem as one of the challenging issues faced in designing credit card fraud detection models.

6.1. Oversampling Techniques

In Benchaji et al. (2018), the authors propose using a genetic algorithm along with K-mean clustering to generate new input data from the minority class (fraudulent instances). This will produce a more balanced dataset and thus enhance the accuracy of the system. An autoencoder is also used for feature selection to generate discriminative features for fraudulent instances. However, the authors discussed the theory of each algorithm used more than the suggested solution itself and did not provide evidence for the proposed model.

In Dornadula and Geetha (2019), the authors developed profiles for cardholders and used Synthetic Minority Oversampling (SMOTE) and One-class SVM to handle the imbalanced data. They measured the performance using Matthews correlation coefficient (MCC) which is used to deal with imbalanced datasets (see 7). Similarly, Rtayli and Enneya (2020) used oversampling to solve the class imbalance issue and proposed a hybrid system that uses SMOTE.

In Tran and Dang (2021), the authors tried to increase the effectiveness of machine learning models to detect credit card fraud by utilizing two resampling techniques for imbalanced datasets: SMOTE and daptive Synthetic (ADASYN), which is an improvement of SMOTE. The authors employed four machine learning algorithms (RF, logR, KNN, and DT) to compare and evaluate the performance of the balanced dataset produced from both SMOTE and ADASYN. However, the resampling method used might have led to overfitted data. Similarly, Bagga et al. (2020) used ADASYN to solve the class imbalance. However, they achieved an accuracy of 100%, which may indicate overfitting caused by the used oversampling technique.

In Yang et al. (2019), SMOTE was also used as an oversampling technique to balance the dataset. The performance testing of the model on a large-scale real dataset showed that the federated learning achieves an average of test Area Under the Curve (AUC) to 95.5%, which is about 10% higher than traditional systems. However, this scheme needs to take reliable measurements into consideration in order to preserve privacy (Li et al., 2020).

Ingole et al. (2021) oversampled their data set with varying sample sizes before reaching 20,000 samples to achieve the best performance. They used the sklearn.util.resample utility which is based on bootstrapping. The suggested solution is simple but might lead to overfitting issues. It is also tedious to find the oversampling rate for larger datasets as it requires testing several sampling sizes.

6.2. Undersampling Techniques

As discussed in Section 5, (Roy et al., 2018) compared several deep learning models. In addition, they solved the class imbalance and scalability problems through random undersampling. They tested several ratios and found that 10:1 non-fraudulent: fraudulent was the most suitable for credit card detection. However, this result is highly dependent on the dataset used.

Li et al. (2021) proposed a method to handle the class imbalance with overlap based on a divide-and-conquer approach. They trained an anomaly detection model on the minority class in order to exclude a few outliers of the minority class and many samples from the majority class. The remaining samples, therefore, constitute an overlapping subset with a low imbalance ratio. Afterwards, they dealt with the resulting overlapping subset by using a nonlinear classifier to distinguish samples. They also proposed a new assessment criterion called Dynamic Weighted Entropy (DWE) to evaluate the quality of the set, which takes into account the trade-off between the number of excluded outliers of the minority class and the ratio of class imbalance of the overlapping subset.

6.3. Hybrid Techniques

Mohammed et al. (2018) tested Random Oversampling (ROS) with many classifiers for massive datasets. They also tested various types of SMOTE, namely the original version, the borderline1 and borderline2, SVM-SMOTE, SMOTEENN, and SMOTETomek using RUS. BBE was tested, which is balanced internally with Random Undersampling (RUS) along with SMOTE, to build a hybrid balancing model. The results showed that the hybrid techniques are scalable and achieved the best performance.

Thennakoon et al. (2019) suggests a fraud detection system that uses oversampling and undersampling techniques to solve the data imbalance problem. It uses SMOTE to over-sample fraudulent instances, as well as condensed nearest neighbor CNN and RUS to undersample the genuine records.

Baabdullah et al. (2020) performed a comparative experimental study to detect credit card fraud, and to tackle the imbalance classification problem by applying different machine learning algorithms to handle imbalanced datasets along with over- and under-sampling methods. The study allows experimental insights to be obtained into the use of ML techniques to detect card fraud, especially with imbalanced datasets. The authors used SMOTE as

an oversampling technique and NearMiss as an undersampling technique.

6.4. Other Balancing Techniques

Alongside over-sampling, under-sampling, and hybrid techniques that combine both over- and under-sampling, many researchers have suggested other solutions to handle the class imbalance issue, such as cost-sensitive learning and extreme learning. Cost-sensitive learning assigns different costs to the types of misclassification errors. Afterwards, specialized methods are used to take these costs into account. Cost-sensitive learning aims to minimize the cost of a model on a training dataset rather than minimizing the prediction error, as in regular learning (Sammut and Webb, 2010).

Moreover, parallel one-class extreme learning (P-ELM) has been suggested in Li et al. (2018) to solve the class imbalance problem. It combines one-class classifiers and Extreme Learning Machine (ELM). The suggested solution is applicable to both multi-class and binary classification problems. The one-class ELM relies on the fact that all trained objects should be in a feature space that contains just the target class, thus being useful for outlier detection when only the target class data is available (Leng et al., 2015). P-ELM addresses the class imbalance problem using the Bayesian theory. The training data set is divided into k subsets based on the class features, where k is the number of classes. The resulting datasets are then fed in parallel to separate Gaussian kernel-based one-class ELM, then the conditional probability is used to determine the output class. Another ELM-based solution for the imbalance issue has been proposed in Xiao et al. (2017). It uses class-specific cost regulation extreme learning machine (CCR-ELM) and deals with binary and multi-class classification problems. In this approach, the regularization parameters need to be tuned to get the best results, which is computationally expensive. Although a kernel extension is used in CCR-ELM to improve the performance, the calculation of regularization parameters does not take into account the class distribution and overlapping samples (i.e., close negative and positive data points) (Raghuwanshi and Shukla, 2018). As such, CCR-ELM has been further improved in Raghuwanshi and Shukla (2018) wherein a class-specific solution for ELM imbalance problems (CS-ELM) is proposed. It has better performance than CCR-ELM and uses a class skewness parameter to deal with overlapping samples effectively.

In the credit card fraud detection field, both cost-sensitive and extreme learning have been suggested. For instance, Kim et al. (2019) uses cost-sensitive learning in addition to random sampling to solve the class imbalance problem. Olowookere and Adewale (2020) contributed to the field of class imbalance in credit card fraud detection field handling by using both ensemble learning and cost-sensitive learning. The main idea is to incorporate cost-sensitive learning in the meat classifier in ensemble learning without enforcing cost-sensitive learning on each individual classifier. The authors showed that their solution outperforms pure ensemble methods under high class imbalance. Akila and Srinivasulu Reddy (2018) solves the class imbalance problem by adding constraints to their bagging ensemble model to reduce the imbalance levels in the bags thus achieving improved predictions. The authors propose an Overlapped-Majority Bagging (OMB) model that creates bags by selecting 60% of training samples for each new bag while ensuring that all samples from the minority class instances are selected, while the rest are randomly selected from the majority class. The value of 60% was selected based on multiple experiments and is recommended as a minimum threshold to avoid internal imbalance. Another paper that addressed the class imbalance is Fu et al. (2016), with cost-based sampling applied to the CNN classifier.

According to Zong et al. (2013), WELM is well-suited for imbalanced datasets. By assigning different weights for each example, the weighted ELM can be generalized to cost-sensitive learning (Zong et al., 2013). Some research has focused on using WELM in credit card detection, mainly to deal with the imbalance problem. Zhu et al. (2020) applied multiple optimization techniques to optimize a WELM and compared their performance to the credit card imbalanced classification problem. The experimental results show that WELM with a dandelion algorithm (i.e. a swarm intelligence algorithm inspired by dandelion behaviour) (Li et al., 2017) with a probability-based mutation (Zhu et al., 2019) can perform better than WELM with improved particle swarm optimization, as well as the bat algorithm, genetic algorithm, dandelion algorithm and self-learning dandelion algorithm.

Han et al. (2021) tackled class imbalance by providing an optimization model for feature selection combined with WELM. The model in Han et al. (2021) was compared with other optimization algorithms, namely Multiple Objectives Particle Swarm Optimization (MOPSO), and showed good performance results.

Table 7 summarizes the reviewed techniques that have been suggested to solve the class imbalance issue in the field.

Table 7
Summary of the sampling techniques used in the literature.

Ref#	Over-sampling	Under-sampling	Cost-sensitive learning
(Benchaji et al., 2018)	genetic + K-means		
(Roy et al., 2018)		random	
(Mohammed et al., 2018)	ROS, SMOTE (many flavors)	RUS	
(Thennakoon et al., 2019)	SMOTE	CNN & RUS	
(Dornadula and Geetha, 2019)	SMOTE + One-class SVM		
(Kim et al., 2019)		random	entropy
(Rtayli and Enneya, 2020)	SMOTE		
(Tran and Dang, 2021)	SMOTE & ADASYN		
(Bagga et al., 2020)	ADASYN		
(Yang et al., 2019)	SMOTE		
(Baabdullah et al., 2020)	SMOTE	NearMiss	
(Olowookere and Adewale, 2020)			ensemble
(Ingole et al., 2021)	bootstrap		
(Li et al., 2021)		optimization + outlier detection	
(Zhu et al., 2020)			WELM + dandelion
(Han et al., 2021)			WELM + MRPS
(Akila and Srinivasulu Reddy, 2018)			Bayesian inference + bagging
(Fu et al., 2016)			cost-based sampling

Table 8
Public credit card fraud detection datasets.

Name	Publisher	Type	Description				Used in
			Overview	Main features	Imbalance ratio	Size	
PaySim data	Kaggle	synthetic	Generated by using PaySim which simulates mobile money transactions based on a sample of real transactions extracted from financial logs, for one month, of a mobile money service implemented in an African country.	11 features including the type, amount, and customer information.	fraudulent transaction rate < 0.2: 1142 fraudulent and 1047433 legitimate.	1048575	(Rtayli and Enneya, 2020; Cochrane et al., 2021)
CERT Insider Threat Dataset	kilthub	synthetic	A public dataset for insider threat detection evaluation. The dataset also has the ground truth that indicates the malicious activities committed by insiders.	email.csv (email operations like send or receive), logon.csv (logon and logoff operations), http.csv (web browsing: visit, download, or upload operations), file.csv (open, write, copy or delete involving a removable media device), and decive.csv (usage of a thumb drive: connect or disconnect).	5 insiders (in version r6.)	4000	(Zheng, 2020)
Credit Card Fraud Detection	Kaggle	Real	Provides transactions occurring within two days. In order to protect customer privacy, it only contains numerical input variables, which is the result of the PCA transformation.	28 features obtained with PCA, in addition to 'time', 'number', and 'transaction amount'. Time is the interval time (in seconds) between the current transaction and the previous one; Amount is the transaction amount.	0.172%: 492 fraudulent.	284,807	(Wang et al., 2018; Babu and Pratap, 2020; RB and KR, 2021; Rtayli and Enneya, 2020; El hlouli et al., 2020; Olowookere and Adewale, 2020; Ingole et al., 2021; Yang et al., 2019; Mohammed et al., 2018; Baabdullah et al., 2020; Han et al., 2021; Li et al., 2021; Dornadula and Geetha, 2019; Tran and Dang, 2021; Forough and Momtazi, 2021)
ccFraud	Revolution analytics	Not mentioned		7 attributes: Gender, State, Cardholder, Balance, NumTrans, NumIntTrans, CreditLine, and fraudRisk, which represent the target variable.	Fraud cases rate = 5.96%	10 million	(Rtayli and Enneya, 2020; Mohammed et al., 2018)
IEEE-CIS	Kaggle	Real	IEEE Computational Intelligence Society (IEEE-CIS) dataset for fraud detection. This dataset includes around 1 million records and is proposed by Vesta company for competition. It is public in Kaggle competition section.	Two types of features: features related to the transaction (e.g., product, card, address) and features related to the identity (e.g., device information and type)	3.5%	1 million	(Kewei et al., 2021)
Abstract data set for Credit card fraud detection	Kaggle	Real	Uploaded by a data engineer from Larsen & Toubro Infotech, based out of Mumbai India.	11	14.57%:448 fraudulent	3075	(Cochrane et al., 2021)

Table 9
Dataset types used in the literature.

Dataset type		References
Public		
	Real	(Wang et al., 2018; Babu and Pratap, 2020; Rtayli and Enneya, 2020; El hlouli et al., 2020; Olowookere and Adewale, 2020; Ingole et al., 2021; Yang et al., 2019; Rai and Dwivedi, 2020; Mohammed et al., 2018; Itoo et al., 2021; Baabdullah et al., 2020; Han et al., 2021; Li et al., 2021; RB and KR, 2021; Dornadula and Geetha, 2019; Tran and Dang, 2021; Kewei et al., 2021; Cochrane et al., 2021; Bagga et al., 2020; Forough and Momtazi, 2021; Zheng, 2020)
	Synthetic	(Cochrane et al., 2021; Rtayli and Enneya, 2020; Zheng, 2020)
	Not specified	(Padmanabhuni et al., 2019; Zhu et al., 2020)
Private		
	Real	(Thennakoon et al., 2019; Sudha and Akila, 2021; Zhou et al., 2021; Lucas et al., 2019; Sudha and Akila, 2021; Li et al., 2021; Soltani Halvaiee and Akbari, 2014; Kim et al., 2019; Akila and Srinivasulu Reddy, 2018; Forough and Momtazi, 2021; Correa Bahnsen et al., 2016; Zhang et al., 2021; Lucas et al., 2020; Fu et al., 2016; Ali Yeşilkanat et al., 2020)
	Synthetic	(Cui et al., 2021)
	Not specified	(Carcillo et al., 2021) (Jurgovsky et al., 2018)

7. Fraud detection datasets and testing parameters

This section discusses the available datasets used for credit card fraud detection and aims to provide a deep understanding of the evaluation metrics that researchers relied on to test their models. The suitability of these metrics is discussed and the best performance achieved in the reviewed literature is presented.

7.1. Dataset

Many datasets have been used in the literature to test and validate credit card fraud detection models. Generally, researchers rely on private datasets. However, there are publicly available datasets that could be also used. It is worth noting that the availability of a well-established dataset is important to allow researchers provide a clear comparative study of their results and thus show the superiority of a given solution over another. Since the availability of such a dataset is still an issue in the credit card fraud detection field, as most banks refuse to share their data because of privacy concerns and the sensitivity of customers' data, finding datasets for financial research is difficult. As a result, some studies have examined the creation of synthetic datasets (Dandekar et al., 2018).

Synthetic data is artificially created rather than collected from real-world cases and can thus be used to train and validate machine learning models. Although it is an old technique (Robin, 1993; Drechsler, 2011) and addresses the problems of data unavailability and low label quality, synthetic data has not been widely used for credit card fraud detection. This method has several advantages, including reducing constraints when using sensitive data, tailoring data needs to certain conditions that cannot be obtained with real datasets, controlling the data distribution utilized in testing, and simplifying the performance comparison (Ikram Ul Haq et al., 2016).

The main public repositories that provide credit card fraud datasets are the Kaggle datasets[2], UCI repository[3], GitHub[4] and the Kilthub repository[5].

Kaggle datasets offers a public platform for publishing datasets and have been a source for several researchers to test their algorithms. Kaggle contains several public and synthetic datasets for credit cards and has been used extensively in the field of credit card fraud detection.

The main repositories and public datasets for credit card fraud detection are shown in Table 8. These include PaySim, Credit card fraud detection, and IEEE-CIS. In addition to Kaggle, the UCI repository is a collection of datasets, database and domain theories for machine learning. It is an essential source for analyzing machine learning algorithms. Furthermore, it contains 588 datasets, including some credit card datasets, namely Statlog (German Credit Data) Data Set, and the Credit Approval Data Set (used in Padmanabhuni et al. (2019), Han et al. (2021), Zhu et al. (2020)).

Additionally, GitHub contains several synthetic datasets for credit card fraud detection. One of these datasets was produced by Brandon Harris and is used in the Sparcov program[6]. It is a synthetic dataset that contains labelled transactions generated using Faker to produce information about merchants and customers. Another synthetic dataset was adapted from a previous one and is used for real-time credit card fraud analysis[7]. It was generated between June and October 2021 using Faker to generate the transactions and customer profiles.

Finally, the Kilthub repository is an institutional source managed by the library of Carnegie Mellon University. University researchers use this repository to share their materials, including datasets, such as the CERT insider threat dataset used in Zheng (2020).

Table 9 classifies the main datasets used in state-of-the-art solutions and refers to the papers they were used in.

7.2. Fraud detection testing metrics

Choosing the right metric is crucial to assessing machine learning models. Different metrics have been suggested to evaluate ML models in various applications. This section provides the metrics that were used in the surveyed papers and emphasizes the most useful ones that better suit the specificity of the credit card fraud detection problem. Indeed, examining a single metric does not usually provide a full picture of the problem. Meanwhile, the imbalance aspect of the datasets used makes relying on the basic ML metric useless, as illustrated below. In addition, most studies have used misclassification measures to evaluate the various solutions, and thus do not take into account the financial costs of the fraudulent transaction (Correa Bahnsen et al., 2016).

[2] https://www.kaggle.com/datasets.
[3] https://archive.ics.uci.edu/ml/contact.html.
[4] https://github.com/.
[5] https://kilthub.cmu.edu/articles/dataset/Insider_Threat_Test_Dataset/12841247/1.
[6] https://github.com/namebrandon/Sparkov_Data_Generation.
[7] https://github.com/atoti/notebooks/tree/master/notebooks/credit-card-fraud-detection.

The following section first presents the confusion matrix used to define testing metrics (Glossary of Terms, 1998) before summarizing the most commonly used ones to demonstrate their applicability to the credit card fraud detection problem.

The confusion matrix, or error matrix, is a key concept in classification performance. It is a tabular visualization of the model predictions against the ground-truth labels. Each row of the confusion matrix describes the instances in a predicted class, while each column defines the instances in an actual class. In a two-class problem such as the fraud detection problem, the aim is to discriminate between fraudulent and non-fraudulent transactions. In this way, the fraud event can be assigned in the first row as *positive* and the non-fraud event in the second row as *negative*. *'True'* is then assigned to a correct prediction, and *'False'* to a false prediction. Thus, in the first row, the True Positive (TP) predictions and the False Postive (FP) predictions are given. Similarly, the False Negative (FN) predictions and the True Negative (TN) predictions are specified in the second row.

		Actual value		
		Positive	Negative	Total
Prediction outcome	Positive	TP	FP	$TP+FP$
	Negative	FN	TN	$FN+TN$
	Total	$TP+FN$	$FP+TN$	N

A distinction is made between classification metrics, statistical metrics, and cost metrics.

7.2.1. Classification metrics

- Accuracy: Accuracy is the most straightforward metric. It is defined as the number of correct predictions divided by the total number of instances, multiplied by 100:

$$Accuracy = 100 \times (TP + TN)/N$$

  In many cases, accuracy is not a good indicator of the performance of a ML model, especially when the class distribution is imbalanced, which is the case for credit card datasets. In this case, even if all samples are predicted as the most frequent class, the model will have a high accuracy rate, which is meaningless since the model is not learning anything. It is instead predicting everything as the majority class. As such, the precision should be considered.
- Precision: This is one of the most commonly used metrics. It is used to check the positive predictions of the system by finding the ratio of the true positive over the total positive predictions.

$$Precision = \frac{TP}{(TP + FP)}$$

- Recall (aka Sensitivity or True Positive Rate (TPR)): This is another important metric. It refers to the model's ability to correctly detect transactions as fraudulent conditioned on them truly being fraudulent. It is used to measure the percentage of actual positives that are correctly identified (Wang and Zheng, 2013).

$$Recall = TPR = \frac{TP}{TP + FN}$$

- Specificity (aka True Negative Rate (TNR)): This refers to the model's ability to correctly identify non-fraudulent transactions. As such, it is used to measure the percentage of actual negatives which are correctly identified and is defined by:

$$Specificity = \frac{TN}{TN + FP}$$

- False Positive Rate (FPR): This represents the rate of fraudulent transactions that are classified or considered as positive. In order to achieve the system's best performance, this rate should be low (Kumari and Mishra, 2019).

$$FPR = \frac{FP}{FP + TN}$$

- F1-measure: In many applications, both recall and precision are important. Therefore, the F1-measure (or score) combines these two into a single metric, which is the harmonic mean of precision and recall defined as:

$$F1 - measure = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

  Nevertheless, the F-measure performs well when the data is balanced. The Adjusted F1-Measure (AGF) is an improvement of the F-Measure that better suits imbalanced data (Akosa, 2017).
- F2-measure: This measure considers the recall as twice important as precision. It is defined by:

$$F2 - measure = 5 \times \frac{Precision \times Recall}{4 \times Precision + Recall}$$

- Alert rate: This is the ratio of alerted transactions among all the transactions. The alert rate is defined by Kim et al. (2019):

$$A = \frac{TP + FP}{N}$$

7.2.2. Visual Representations

- Receiver Operating Characteristic (ROC) curve: It is a plot that shows the performance of a binary classifier as a function of its cut-off threshold. It mainly displays the TPR against the FPR for different threshold values (Hanley and McNeil, 1982). The ROC curve is a famous curve to examine the overall model performance for choosing a good cut-off threshold.
- Area Under the Curve (AUC): AUC stands for area under the curve. It is a plot that visualizes the tradeoff between TPR and FPR for every threshold on one chart. The higher TPR and the lower FPR is for each threshold the better. Thus, models that have curves that are more top-left-side are better.
- Area Under the Receiver Operating Characteristic (AUC-ROC) (aka AUROC): The area under the receiver operating characteristic (AUROC) is one of the most important metrics used for checking the performance of any classification model. AUROC is a performance metric for discrimination; it shows whether the model is able to correctly rank examples and discriminate between cases. However, this metric should not be used when data is heavily imbalanced. Indeed, the false positive rate for highly imbalanced datasets is pulled down due to a large number of true negatives (Saito and Rehmsmeier, 2015).
- Area Under the Curve of Precision–Recall (AUC-PR) (aka as AUCPRC): It relies on the Precision-Recall curve, which merges precision and recall in one visualization. For every threshold, precision and recall are calculated and plotted. The higher on the y-axis the curve, the better the model's performance. Knowing at which recall the precision starts to fall fast helps in choosing the threshold and delivering a better model. Like the AUC-ROC score, it is possible to compute the area under the precision-recall curve to obtain one metric describing the model performance. AUC-PR can be thought of as the average of

precision scores calculated for each recall threshold. The AUC-PR is relevant when there is a need to choose the threshold that fits a given business problem. It is also useful when the data is heavily imbalanced (Saito and Rehmsmeier, 2015) as AUC-PR concentrates primarily on the positive class.

7.2.3. Statistical metrics

- Matthews correlation coefficient (MCC): This is used to measure the quality of both binary and multiclass problems. It considers true and false positives and negatives. It is commonly viewed as a balanced measure that can be utilized even if the classes are highly imbalanced. The MCC is by nature a correlation coefficient value ranging between −1 and +1. A coefficient of +1 indicates a perfect prediction, 0 indicates an average random prediction, and −1 is an inverse prediction.
- Geometric Mean (G-mean) This metric measures the balance between the classification performance on the majority and minority classes. A low G-Mean is interpreted as a poor performance in the classification of the positive samples even if the negative samples are correctly classified. It is important to use G-mean to avoid over-fitting the negative class and under-fitting the positive class (Akosa, 2017).

$$G-mean=\sqrt{sensitivity\times specificity}$$

- K-S statistics: The K-S test is a non-parametric and distribution-free test. Indeed, this statistical test makes no assumption about the data distribution. The K-S test can be used to compare a sample with a reference probability distribution, or to compare two samples. Given a distribution P, the K-S test is used to evaluate:
  - Null Hypothesis: The samples come from P
  - Alternative Hypothesis: The samples do not come from P

7.2.4. Cost metrics

- Cost-reduction rate: A fraud scoring model estimates scores for genuine transactions. If a score exceeds a predefined cutoff value δ, the transaction is rejected and sent to investigators for further analysis. To reduce costs caused by missed alerts, a scoring model aims to reduce the false alarm rate by returning precise alerts. A false positive costs the same as the transaction analysis and incurs the cost of contacting the cardholder C_c. A missed fraud FN costs as much as the amount $Amount_i$ of the transaction t_i. The total cost C of a model can be measured for each transaction t_i as follows:

$$C_\delta=\sum_{i=1}^{N}c_i\times C_c+y_i\times(1-c_i)\times Amount_i \tag{1}$$

where $c_i=1$ if the system arises an alert for the transaction t_i, and 0 otherwise. The actual status of the transaction (whether fraudulent or not) is reflected by the variable y_i, where $y_i=1$ if the transaction is fraudulent, and 0 otherwise. To compare the performances of a new fraud detection model with a second one, one concern is to measure how much cost is reduced by the new model through the cost reduction rate Cr_δ. This is calculated as follows:

$$Cr_\delta=\frac{C_\delta^{new}}{C_\delta^{old}} \tag{2}$$

where C_δ^{new} and C_δ^{old} are the costs of respectively the new and the old models. The cost metric is a valuable metric that has not been widely used in the credit card fraud detection field.

- Mean run time: The run time is the duration of the execution of an algorithm. When an algorithm is executed several times on the same data input, the mean over all the obtained run times is computed, giving rise to the mean run time. This metric is important as it compares the reliability of the detection results with the time duration taken to achieve those results.
- Cost savings: The cost savings $\mathscr{S}$ of using an algorithm is defined by the cost of the algorithm versus the cost of not using any algorithm at all. It is expressed by Correa Bahnsen et al. (2016):

$$\mathscr{S}=\frac{C-C_\delta}{C} \tag{3}$$

In the case of credit card fraud, the cost C of not using an algorithm is $C=\sum_{i=1}Ny_i\times Amount_i$ In this case, the cost savings are defined by:

$$S=\frac{\sum_{i=1}^{N}y_ic_iAmount_i-c_iC_c}{\sum_{i=1}^{N}y_iAmount_i} \tag{4}$$

- Classification metrics-based cost function (CCF) (Gadi et al., 2008): Some research works such as (Soltani Halvaiee and Akbari, 2014, Akila and Srinivasulu Reddy (2018)) use a cost metric based on the classification metrics as follows:

$$CCF=\alpha\times FN+\beta\times FP+\gamma\times TP$$

For instance, (Soltani Halvaiee and Akbari, 2014) uses 100, 10, and 1 for α,β, and γ respectively while (Akila and Srinivasulu Reddy, 2018) uses the amount of denomination for the transaction as a value of α and $\beta=\gamma$ refers to the administrative cost.

7.2.5. Discussion

Several metrics have been suggested to assess the performance of credit card fraud detection models. Table 10 summarizes the results achieved by the reviewed models. It provides an overview of the testing results obtained for the metrics described for each paper, thus providing a unified general view of the performance of the state-of-the-art techniques. Indeed, different solutions may refer to different terminology for the testing metrics.

As shown in Table 10, for some of these metrics the accuracy was widely used, but not enough to allow the detection system to be evaluated properly. It can be observed from the table that most studies rely on classification metrics, while a few use statistical and cost metrics. It is worth noting that it is important to take into account cost metrics when testing the efficiency of the model, in addition to considering metrics that better fit unbalanced datasets like the MCC. Additional research effort should be directed towards unifying the testing metrics used in credit card fraud detection to permit a better comparison of the state-of-the-art solutions. In addition, it is important to assess the processing time of the suggested model, as many studies claim to be real-time but do not consider the mean run-time metric.

8. Open research problems and research directions

In this section, we give an exhaustive overview on recent advances that has impact on credit card fraud detection. We highlight the main challenges and the future research directions that should be investigated.

Table 10
Metrics used by the surveyed papers.

Ref#	accuracy	precision	recall	F1-score	F2-score	FPR	TNR	ROC	MCC	G-mean	AUC	K-S statistics	cost-reduction rate	AUC-ROC	AUC-PR	Mean run time	Alert rate	CCF
(Wang et al., 2018)	96.40%	98.64%	97.83%	98.04%														
(Thennakoon et al., 2019)	91%								74%									
(Padmanabhuni et al., 2019)	82.47%	98.54%	90.61%	89.72%			76.12%											
(Babu and Pratap, 2020)	99.62%																	
(Kewei et al., 2021)	95.8%													91%				
(Cui et al., 2021)	99.59%	86.42%	81.69%	82.48%			99.86%			90.29%								
(Sudha and Akila, 2021)	98%	86%	94%	90%														
(Zhou et al., 2021)		97%	96%	73%	71%													
(Lucas et al., 2019)														97.3%	30.3%			
(Cochrane et al., 2021)	87%																	
(Tran and Dang, 2021)	99.99%	99.98%	100%	99.99%			99.98%		99.98%	99.99%	100%							
(Han et al., 2021)										88.92%						25982.09		
(Sudha and Akila, 2021)	98.5%	92.5%	81.5%	87%														
(Olowookere and Adewale, 2020)														100%				
(Li et al., 2021)				79%											77%	5.5		
(Soltani Halvaiee and Akbari, 2014)		52.6%	51.8%			1.7%												70
(Kim et al., 2019)		–	87.5%									80.4%	23.7%	94.8%			0.96%	
(Akila and Srinivasulu Reddy, 2018)	96.46%		100%			0.8%	99%		90%		99%							15279
(RB and KR, 2021)	99.92%	97.43%	89.76%															
(Dornadula and Geetha, 2019)	99.98%	99.96%							99.96%									
(Bagga et al., 2020)	99.99%	100%	100%	100%					85%									
(Rtayli and Enneya, 2020)	100%	97%	100%	99%										100%	97%			
(Forough and Momtazi, 2021)		97.91%	77.95%	82.29%										86.68%	67.07%			
(Correa Bahnsen et al., 2016)																		
(Zhang et al., 2021)	98.25%	62.60%	75%	57.7%							97.6%							
(Zhu et al., 2020)	98.18%	94.67%	98.21%	93.96%						97.88%	97.89%							
(Jurgovsky et al., 2018)															40.4%			
(Lucas et al., 2020)															31.7%			
(Fu et al., 2016)			33%															
(Ingole et al., 2021)	99.51%	99.99%	100%	99.8%			99.60%											
(Mohammed et al., 2018)		94.12%	93.88%	90.72%		0%	100%							98.11%				
(Yang et al., 2019)	99%		88%	95.34%							96.90%							
(Ali Yeşilkanat et al., 2020)			98.9%			0.3					96.8%							
(Baabdullah et al., 2020)	99%	94%	98%											95%	92%			
(Zheng, 2020)	71.80%	81.98%	55.69%	65.37%										94.96%				

8.1. Big data technologies

A huge amount of data is continually being generated and collected. With the increasing use of social media, more and more textual and image data is being collected. Meanwhile, the increase in use of IoT, and the emergence of smart technologies and smart cities has led to a large amount of sensed data that needs to be processed to afford decision-makers clearer insights.

Big Data and Data Science describe the application of software and technology integrated with cutting-edge algorithms and techniques to acquire better understandings, produce informed findings, and forecast risks and gains (Doko and Miskovski, 2019). Big data is characterized by greater variety, rising volumes, and velocity, known as the three vs. More Vs of relevance have been proposed by researchers, such as veracity, which refers to data quality and value, i.e., transforming the data tsunami into business (Ishwarappa, 2015). Consequently, big data provides larger and more complex data sets, mainly derived from new data and heterogeneous sources. These data sets are characterized by their large volume, with the result that traditional data processing software is unable to handle them. Meanwhile, these massive volumes of data are useful for supporting business intelligence. In particular, big Data can support the processing of large-scale banking data in real-time, thereby enhancing fraud detection. Moreover, big data technologies allow the integration of heterogeneous data drawn from multiple sources to determine the role of fraud more efficiently.

Data parallelism entails partitioning a large data set across numerous nodes in a cluster. Each node is responsible for processing a small chunk of data, after which the results are combined to produce the final result. This is the most efficient option to allow for rapid response and quick decision-making. Google's MapReduce (Dean and Ghemawat, 2010) processes data in parallel applying two stages: Map and Reduce. In the map phase, data is dispersed within separate jobs allocated to various nodes. Each map task produces a set of key-value pairs that are fed to the reduce tasks, aka reducers. These reducers aggregate these key-value pairs into a smaller set that forms the final output. Hadoop is the main framework that enables big data analytics, and which is based on Google's MapReduce in addition to the distributed file system HDFS (Shvachko et al., 2010) and resources orchestrator, Yarn (Vavilapalli et al., 2013). However, Hadoop introduces additional overheads as data is read and written twice on the disk when performing the mapping and reducing tasks. This led to the proposal of a new framework, Spark. The main principle of Spark is that it can be used to process data in memory, delivering significant improvements in terms of efficiency and scalability relative to Hadoop. Spark is similar to MapReduce in that it provides both mapping and reduction functions. However, it supports more operations to take place over large and distributed data sets, such as filtering and SQL-like operations over distributed and resilient data structures, RDDs and data frames.

As the scale of financial transaction data continuously and dramatically increases, it becomes progressively more challenging to detect fraudulent operations reliably and in real-time using traditional techniques. Therefore, the adoption of a big data analytics approach that learns patterns derived from large sets of historical data and then deploys a distributed infrastructure to alleviate the heavy computations is needed. Despite the importance of a distributed approach, this study has shown very few research works have investigated utilizing big data analytics within a distributed architecture to detect and/or predict credit card fraud (Chen et al., 2020; Zhou et al., 2021). Thus, additional research is required to provide real-time systems and investigate the challenges that may impede the effective use of such technologies while preserving the privacy of users. Moreover, it is important to integrate several data sets from multiple heterogeneous sources to attain more accurate results. Furthermore, there is still a need for open big datasets to support researchers in investigating big data analytic. A possible research direction would be to develop valuable synthetic big datasets.

8.2. Cloud computing

Cloud computing supports the delivery of computing as a service over the Internet, rather than as a product (Armbrust et al., 2010). This includes shared resources, software, and information that can be delivered to computers and other devices remotely. Cloud computing is characterized by five essential aspects: on-demand self-service, broad network access, resource pooling (Wischik et al., 2008), rapid elasticity (Galante and de Bona, 2012), and measured service (Mell, 2009). In summary, these characteristics allow organizations to share resources, scale up and down in a flexible fashion, and acquire and pay for resources on-demand, which provides many benefits for business management and minimizes cost (Soltani Halvaiee and Akbari, 2014). This means that cloud computing can help organizations continuously improve upon their strategic skills while decreasing the complexity of their business and IT functions, facilitating competitiveness in today's ever-changing market (Wang et al., 2016). Gartner estimated that global cloud revenue in 2022 to 474 billion, up from 408 billion in 2021 (Gartner, 2021). Moreover, Gartner's analysts predict that in excess of 85% of organizations will have adopted a cloud-first principle by 2025. According to Milind Govekar, a vice president at Gartner, "cloud has enabled new digital experiences such as mobile payment systems where banks have invested in startups, energy companies using the cloud to improve their customers' retail experiences" (Gartner, 2021). This rapid growth in cloud computing markets has attracted attention from both academia and industry.

In particular, cloud computing can offer notable benefits to allow credit card fraud detection from different perspectives, such as cost savings and offering computation power. In cloud computing there are no limitations on memory, computation or storage, as it utilizes large data centers as resources. The increased usage of small devices for credit card fraud detection raise interesting research questions about the value of migrating computation to the cloud while designing new lightweight solutions and algorithms for remote devices. It is also interesting to adopt a cloud based architecture and reuse cloud intelligent services and infrastructures to empower credit card fraud detection. Very few research studies have investigated cloud computing in the field of credit card detection, and several open research directions can be explored in this field namely, the use of the cloud to collect and store customers' data, along with computation and AI models hosted in the cloud. Moreover, the aggregating of heterogeneous data sources and interoperability between several organizations can assist in enhancing real-time fraud detection.

Future research may investigate federated machine learning and the use of cloud/edge computing to model the credit card fraud detection problem as a distributed machine learning system that involves multiple heterogeneous datasets coming from several banks while preserving the privacy of card holders.

8.3. IoT and credit card transactions

IoT technology is powering a global process in the credit card payment field. The global financial services sector is experiencing significant mutation. In particular, digital payment is in the depth of a rapid evolution, led by increasing consumer demand for comfortable, connected payment solutions, and driven by a rising wave of strong contemporary and innovative technologies. In addition to

the increasingly cutting-edge data analytics, artificial intelligence and cloud-based architectures, one of the most disruptive technologies with an impact on the credit card field is the IoT. The evolution of IoT has been rapid, prompting enormous growth in multiple fields (Kumari and Mishra, 2019). It is expected to grow larger and faster over the next decade. According to statista.com, it is estimated that, by 2023, global spending on IoT technology could be as high as $1.1trn, with financial services being one of the domains most readily associated with its adoption.

Over the last ten years, every smartphone has become a potential purchasing tool. Similarly, it is expected that any and all devices will become platforms for purchasing goods and services in the very near future, as announced by Mastercard directors. This gave rise to the concept of the IoP (Internet Of Payments). Although this terminology has not yet been investigated within academia, it is increasingly used in industrial contexts when collaborating with IoT manufacturers. Mastercard and Visa are pioneering IoP services, introducing novel solutions to financial organisations aimed at providing seamless card payments using IoT connected devices. This has been facilitated through wearable IoT devices which can substitute the credit card. For example, Fitbit can enable payments with the wave of a hand. Moreover, many car manufacturers are providing in-car apps to allow seamless payments. It is now even possible to shop for groceries using Samsung refrigerators, etc. and IoP is penetrating multiple aspects of our daily economic and social activities. The COVID-19 pandemic increased the volume of online transactions and encouraged many customers to move to online service providers, leading to a higher number of payments via IoT devices and mobile phones (Wiścicka-Fernando, 2021). However, the rapid development of mobile and IoT financial payment services has not only provided convenience and efficiency for consumers but has also brought additional hidden fraud risk. The concealment of complex networks, can be a breeding ground for fraudulent activities by criminals. Controlling and managing fraud risk is ever more complex, as fraud arises more and more frequently, resulting in high monetary losses for commercial banks and financial institutions.

Possible future research directions could include the consideration of hybrid datasets that combine both transactional data and external data collected from IoT devices or the development of new system designs and learning methods involving data collected from IoT devices. Besides, behaviour biometrics collected from IoT devices may be used to authenticate users and avoid fraudulent transactions. Thus, new detection and authentication models should be investigated.

8.4. Security and privacy concerns

Recently, research has concentrated on providing techniques to protect against fraud in the financial services arena. However, there are several security concerns that remain to be considered. First, to build an accurate fraud detection model, we need to collect a large number of user behaviour data (past-current) to detect any deviation from normal. Securing this information is critical, as revealing it violates users' privacy, meaning an attacker can take advantage of them and try to impersonate them. Additionally, this data could help fraudsters create innovative patterns to bypass the system. Notable, when DDoS attacks become very frequent, the service is prone to becoming unavailable. Furthermore, fraudsters' activities develop rapidly, so building a mechanism using only traditional machine learning techniques is inadequate, as they are not updated and do not require variations in customer's behaviour, such as regions, holidays, etc. Benchaji et al. (2021). Researchers might also take advantage of new technologies, such as blockchain to build secure financial systems. Only a few researchers have considered this area. For example, blockchain technology can be used to prevent credit card fraud (Balagolla et al., 2021) and chargeback fraud (Liu and Lee, 2021).

To relieve clients from the burden of computation, some digital transactions rely on delegated authentication to allow communication via a trusted server. However, an attacker can compromise the authentication server to impersonate clients and weaken the security of digital transactions. In Yang et al. (2021), the authors propose protocols based on mutual authentication to detect identity fraud. Despite the lightweight delegated authentication protocols, the performance rating is unrealistic. Indeed, any solution designed to detect fraud should be practical and involve minimal computational overhead. It must also be supported by an empirical performance study addressing highly impactful parameters, such as latency and network bandwidth. Fraud detection protocols are all driven by good intentions such as protecting clients, but often lack a human dimension in their design and usage. Accordingly, qualitative and quantitative studies require users to become involved in digital transactions that shed light on their security perceptions and usage behaviour. While machine learning techniques are the cornerstone of fraud detection protocols, privacy concerns slow their adoption in the industry, as regulations have become increasingly stringent (e.g. GDPR: General Data Protection Regulation). Although emerging privacy-preserving machine learning techniques use homomorphic encryption for protecting personal data, increasing the computational overheads of fraud detection unfortunately remain a serious obstacle. Collaborative learning without data sharing is thus increasingly a promising research direction.

It is evident that blockchain technology has the potential to become the new engine of growth in the digital economy. Making it possible to eliminate fraud and increase transparency regarding flow in digital transactions. However, using such technology requires the compliance of banks to link blockchain tokens with traditional monetary systems. While its main advantages include decentralization, transparency, immutability, and automation, it is also subject to limitations that relate to transaction throughput, latency, size, and bandwidth. Moreover, due to the unlimited storage of all transactions to support tracking, blockchain-based systems can be prone to privacy issues created by user profiling and pseudonymization breaches.

Credit card security violations are generally outside the scope of the industry's current laws and regulations, and industrial regulations are being weakened by the innovative development of IoP, revealing insufficient laws and regulations to deal with credit card violations in a timely manner. Regulations such as the Federal Trade Commission (FTC), General Data Protection Regulation (GDPR), and Payment Card Industry Data Security Standard (PCI DSS) need adapting to comply with to secure Personally Identifiable Information (PII).

Undoubtedly, although IoP may makes security more challenging, its capabilities can also enhance the security of online transactions. For example, IoP wearables permit customers to connect with banking apps for better tracking of their financial behaviour, while also performing a wide range of other mobile banking operations. These devices not only facilitate banking operations for the user but can also bestow advanced security features, such as reliance on customer heart rate for biometric authentication. Tokenization can be a sure industry-standard admissible by leading payment networks to guarantee the privacy and security of banking transactions. Furthermore, with card-on-file EMV tokenization, vendors no longer need to store card details in their databases, only tokens. An initial version of the Tokenization framework was published in March 2014 by EMV Co. This framework was intended to promote a new generation of payment products, while empowering the existing payments infrastructure. Research works are required to investigate these advances and devise new fraud detection techniques to take into account new technologies.

9. Conclusion

The rise of IoT technologies and data-driven finance seriously influences the daily life and behaviour of customers. Digital payments are spreading rapidly and gaining in importance, mainly since the COVID-19 pandemic. The use of recent advances, namely AI, big data and cloud computing are attracting huge interest from researchers. Although an abundance of research has been undertaken in the field in recent years, extremely limited comprehensive surveys have also been conducted. Therefore, this research aims to offer the latest and most relevant research on credit card fraud detection, impacting new technologies in the field. Our study analyzed 40 works extracted from the Scopus and Web of Science databases. The literature sample was also explored, focusing on research trends, theoretical achievements, methodology, strengths and weaknesses, modeling and testing. Evidence shows a few state-of-the-art models have focused on the use of deep learning models, big data analytics and technologies, IoT, real-time and security aspects, whereas the majority of previous works have focused on comparing traditional detection models without a clear effort in providing new designs and models better suited to the credit card fraud detection problem. Besides, it is worth noting that a large majority of the analysed work relies on test metrics that are unsuitable for credit card datasets. Some relevant directions for future work are also proposed. Each topic in this article contains an overview of research done in the area of credit card fraud and key concepts are considered each category. Thus, a thorough analysis of prior findings allows researchers to avoid duplicating empirical studies and to identify relevant research gaps. This research is original, setting out new directions for future work to fill existing research gaps. Our comprehensive literature review collates many scholarly articles, establishing a solid foundation to accurately represent the significant contributions in this field to date. We believe that our research provides a starting point for an in-depth understanding of the current state of the customer perspective with regard to digital payment.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research work was funded by Institutional Fund Projects under Grant No. (IFPRC-032-612-2020). Therefore, authors gratefully acknowledge technical and financial support from the Ministry of Education and King Abdulaziz University, Jeddah, Saudi Arabia.

References

Akila, S., Srinivasulu Reddy, U., 2018. Cost-sensitive Risk Induced Bayesian Inference Bagging (RIBIB) for credit card fraud detection. J. Comput. Sci. 27, 247–254. https://doi.org/10.1016/j.jocs.2018.06.009. URL https://www.sciencedirect.com/science/article/pii/S1877750317311729.

Akosa, J.S., 2017. Predictive accuracy: A misleading performance measure for highly imbalanced data.

Al-Hashedi, K.G., Magalingam, P., 2021. Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019. Comput. Sci. Rev. 40, 100402. https://doi.org/10.1016/j.cosrev.2021.100402. URL https://www.sciencedirect.com/science/article/pii/S1574013721000423.

Ali Yeşilkanat, A., Bayram, B., Köroğlu, B., Arslan, S., 2020. An Adaptive Approach on Credit Card Fraud Detection Using Transaction Aggregation and Word Embeddings — SpringerLink. https://link.springer.com/chapter/10.1007/978-3-030-49161-1_1.

Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., Zaharia, M., 2010. A view of cloud computing. Commun. ACM 53 (4), 50–58. https://doi.org/10.1145/1721654.1721672.

Awad, M., Khanna, R., 2015. Machine Learning, Apress, Berkeley, CA, pp. 1–18. https://doi.org/10.1007/978-1-4302-5990-9_1.

Baabdullah, T., Alzahrani, A., Rawat, D.B., 2020. On the comparative study of prediction accuracy for credit card fraud detection with imbalanced classifications. In: Proceedings of the 2020 Spring Simulation Conference, SpringSim '20, Society for Computer Simulation International, San Diego, CA, USA.

Babu, A.M., Pratap, A., 2020. Credit Card Fraud Detection Using Deep Learning. In: 2020 IEEE Recent Advances in Intelligent Computational Systems (RAICS), pp. 32–36. https://doi.org/10.1109/RAICS51191.2020.9332497.

Bagga, S., Goyal, A., Gupta, N., Goyal, A., 2020. Credit card fraud detection using pipeling and ensemble learning. Procedia Comput. Sci. 173, 104–112, international Conference on Smart Sustainable Intelligent Computing and Applications under ICITETM2020. https://doi.org/10.1016/j.procs.2020.06.014. URL https://www.sciencedirect.com/science/article/pii/S1877050920315167.

Bahnsen, A.C., Aouada, D., Ottersten, B., 2014. Example-dependent cost-sensitive logistic regression for credit scoring. In: 2014 13th International Conference on Machine Learning and Applications, pp. 263–269. https://doi.org/10.1109/ICMLA.2014.48.

Bahnsen, A.C., Stojanovic, A., Aouada, D., Ottersten, B., 2014. Improving Credit Card Fraud Detection with Calibrated Probabilities, pp. 677–685. arXiv:https://epubs.siam.org/doi/pdf/10.1137/1.9781611973440.78, https://doi.org/10.1137/1.9781611973440.78. https://epubs.siam.org/doi/abs/10.1137/1.9781611973440.78.

Balagolla, E., Fernando, W., Rathnayake, R., Wijesekera, M., Senarathne, A.N., Abeywardhana, K., 2021. Credit card fraud prevention using blockchain. in: 2021 6th International Conference for Convergence in Technology (I2CT), pp. 1–8. https://doi.org/10.1109/I2CT51068.2021.9418192.

Balaji, K., Shipeng, Y., Bharat, R.R., 2011. Cost-Sensitive Machine Learning, 1st ed. vol. 1, CRC Press.

Baldi, P., 2011. Autoencoders, unsupervised learning and deep architectures. In: Proceedings of the 2011 International Conference on Unsupervised and Transfer Learning Workshop - Volume 27, UTLW'11, JMLR.org, pp. 37–50.

Becker, A., Becker, J., 2021. Dataset shift assessment measures in monitoring predictive models. Procedia Comput. Sci. 192 (2021) 3391–3402, knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 25th International Conference KES2021. https://doi.org/10.1016/j.procs.2021.09.112. URL https://www.sciencedirect.com/science/article/pii/S1877050921018512.

Benchaji, I., Douzi, S., ElOuahidi, B., 2018. Using Genetic Algorithm to Improve Classification of Imbalanced Datasets for Credit Card Fraud Detection. In: 2018 2nd Cyber Security in Networking Conference (CSNet), pp. 1–5. https://doi.org/10.1109/CSNET.2018.8602972.

Benchaji, I., Douzi, S., El Ouahidi, B., Jaafari, J., 2021. Enhanced credit card fraud detection based on attention mechanism and LSTM deep model. J. Big Data 8 (1), 151. https://doi.org/10.1186/s40537-021-00541-8.

Breiman, L., 2001. Random Forests. Machine Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.

Breunig, M.M., Kriegel, H.-P., Ng, R.T., Sander, J., 2000. Lof: Identifying density-based local outliers. In: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, SIGMOD '00, Association for Computing Machinery, New York, NY, USA, p. 93–104. https://doi.org/10.1145/342009.335388.

Carcillo, F., Le Borgne, Y.-A., Caelen, O., Kessaci, Y., Oblé, F., Bontempi, G., 2021. Combining unsupervised and supervised learning in credit card fraud detection. Inf. Sci. 557, 317–331. https://doi.org/10.1016/j.ins.2019.05.042. URL https://www.sciencedirect.com/science/article/pii/S0020025519304451.

Chakraborty, B., Chatterjee, A., Malakar, S., Sarkar, R., 2022. An iterative approach to unsupervised outlier detection using ensemble method and distance-based data filtering. Complex Intell. Syst. (02) https://doi.org/10.1007/s40747-022-00674-0.

Chambers, B., Zaharia, M., 2018. Spark: The Definitive Guide Big Data Processing Made Simple. O'Reilly Media Inc.

Chandola, V., Banerjee, A., Kumar, V., 2009. Anomaly detection: A survey, ACM Comput. Surv. 41 (3). https://doi.org/10.1145/1541880.1541882.

Chaquet-Ulldemolins, J., Gimeno-Blanes, F.-J., Moral-Rubio, S., Muñoz-Romero, S., Rojo-Álvarez, J.-L., 2022. On the black-box challenge for fraud detection using machine learning (ii): Nonlinear analysis through interpretable autoencoders. Appl. Sci. 12 (8). https://doi.org/10.3390/app12083856. URL https://www.mdpi.com/2076-3417/12/8/3856.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. Smote: Synthetic minority over-sampling technique. J. Artif. Int. Res. 16 (1), 321–357.

Chen, H., Ai, H., Yang, Z., Yang, W., Ye, Z., Dong, D., 2020. An improved xgboost model based on spark for credit card fraud prediction, in: 2020 IEEE 5th International Symposium on Smart and Wireless Systems within the Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS), IEEE, 2020, pp. 1–6.

Cheng, D., Wang, X., Zhang, Y., Zhang, L., 2020. Graph neural network for fraud detection via spatial-temporal attention. IEEE Trans. Knowl. Data Eng., 1 https://doi.org/10.1109/TKDE.2020.3025588.

Cho, K., Merrienboer, B., Bahdanau, D., Bengio, Y., 2014. On the properties of neural machine translation: Encoder-decoder approaches 9. https://doi.org/10.3115/v1/W14-4012.

Choi, D., Lee, K., 2018. An artificial intelligence approach to financial fraud detection under IoT environment: A survey and implementation 2018 5483472, publisher: Hindawi. https://doi.org/10.1155/2018/5483472.

Cochrane, N., Gomez, T., Warmerdam, J., Flores, M., Mccullough, P., Weinberger, V., Pirouz, M., 2021. Pattern Analysis for Transaction Fraud Detection. In: 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), pp. 0283–0289. https://doi.org/10.1109/CCWC51732.2021.9376045.

Correa Bahnsen, A., Aouada, D., Stojanovic, A., Ottersten, B., 2016. Feature engineering strategies for credit card fraud detection. Expert Syst. Appl. 51, 134–142. https://doi.org/10.1016/j.eswa.2015.12.030. URL https://www.sciencedirect.com/science/article/pii/S0957417415008386.

Cui, J., Yan, C., Wang, C., 2021. ReMEMBeR: Ranking Metric Embedding-Based Multicontextual Behavior Profiling for Online Banking Fraud Detection. IEEE Trans. Comput. Social Syst. (2021) 1–12. Conference Name: IEEE Transactions on Computational Social Systems. https://doi.org/10.1109/TCSS.2021.3052950.

Dablain, D., Krawczyk, B., Chawla, N.V., 2022. Deepsmote: Fusing deep learning and smote for imbalanced data. IEEE Trans. Neural Networks Learn. Syst., 1–15 https://doi.org/10.1109/TNNLS.2021.3136503.

Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., Bontempi, G., 2015. Credit card fraud detection and concept-drift adaptation with delayed supervised information. In: 2015 International Joint Conference on Neural Networks (IJCNN), pp. 1–8. https://doi.org/10.1109/IJCNN.2015.7280527.

Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., Bontempi, G., 2018. Credit card fraud detection: A realistic modeling and a novel learning strategy. IEEE Trans. Neural Networks Learn. Syst. 29 (8), 3784–3797. https://doi.org/10.1109/TNNLS.2017.2736643.

Dandekar, A., Zen, R.A.M., Bressan, S., 2018. A comparative study of synthetic dataset generation techniques. In: Hartmann, S., Ma, H., Hameurlain, A., Pernul, G., Wagner, R.R. (Eds.), Database and Expert Systems Applications. Springer International Publishing, Cham, pp. 387–395.

Daniel, G.G., 2013. Artificial Neural Network. Springer, Netherlands, Dordrecht. https://doi.org/10.1007/978-1-4020-8265-8_200980. 143–143.

Dean, J., Ghemawat, S., 2010. Mapreduce: A flexible data processing tool. Commun. ACM 53 (1), 72–77. https://doi.org/10.1145/1629175.1629198.

Denil, M., Trappenberg, T., 2010. Overlap versus imbalance. In: Canadian conference on artificial intelligence. Springer, pp. 220–231.

Doko, F., Miskovski, I., 2019. An overview of big data analytics in banking: Approaches, challenges and issues. URL https://knowledgecenter.ubt-uni.net/conference/2019/events/270.

Domingos, P., 2018. The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World. Basic Books Inc, USA.

Dornadula, V.N., Geetha, S., 2019. Credit card fraud detection using machine learning algorithms, Procedia Comput. Sci. 165 (2019) 631–641, 2nd International Conference on Recent Trends in Advanced Computing ICRTAC - DISRUP - TIV INNOVATION, 2019 November 11–12, 2019. doi: https://doi.org/10.1016/j.procs.2020.01.057. https://www.sciencedirect.com/science/article/pii/S187705092030065X.

Drechsler, J., 2011. Synthetic Datasets for Statistical Disclosure Control. Springer.

El hlouli, F.Z., Riffi, J., Mahraz, M.A., 2020. Credit Card Fraud Detection Based on Multilayer Perceptron and Extreme Learning Machine Architectures. https://ieeexplore.ieee.org/document/9204185/.

Elrahman, S.M.A., Abraham, A., 2014. A review of class imbalance problem. J. Network Innovative Comput. 1.

Fan, C., Xiao, F., Zhao, Y., Wang, J., 2018. Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in building energy data. Appl. Energy 211, 1123–1135. https://doi.org/10.1016/j.apenergy.2017.12.005. URL https://www.sciencedirect.com/science/article/pii/S0306261917317166.

Forough, J., Momtazi, S., 2021. Ensemble of deep sequential models for credit card fraud detection. Appl. Soft Comput. 99, 106883. https://doi.org/10.1016/j.asoc.2020.106883. URL https://www.sciencedirect.com/science/article/pii/S1568494620308218.

Fu, K., Cheng, D., Tu, Y., Zhang, L., 2016. Credit Card Fraud Detection Using Convolutional Neural Networks. In: Hirose, A., Ozawa, S., Doya, K., Ikeda, K., Lee, M., Liu, D. (Eds.), Neural Information Processing, Lecture Notes in Computer Science, Springer International Publishing, Cham, pp. 483–490. https://doi.org/10.1007/978-3-319-46675-0_53.

Gadi, M.F.A., Wang, X., do Lago, A.P., 2008. Credit card fraud detection with artificial immune system. In: Bentley, P.J., Lee, D., Jung, S. (Eds.), Artificial Immune Systems. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 119–131.

Galante, G., de Bona, L.C.E., 2012. A survey on cloud computing elasticity. In: 2012 IEEE Fifth International Conference on Utility and Cloud Computing, pp. 263–270. https://doi.org/10.1109/UCC.2012.30.

https://www.gartner.com/en/newsroom/press-releases/2021-11-10-gartner-says-cloud-will-be-the-centerpiece-of-new-digital-experiences, accessed: 2021-12-10.

Gerbaix, M.P.S., 2010. The complexity of security studies in nfc payment system, School of Computer and Information Science, Edith Cowan University, Perth, Western Australia, https://doi.org/10.4225/75/57b674cb34783.

Glossary of Terms, 1998. Machine Learn. 30 (2), 271–274. https://doi.org/10.1023/A:1017181826899.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press. URL http://www.deeplearningbook.org.

Grover, A., Leskovec, J., 2016. Node2vec: Scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, Association for Computing Machinery, New York, NY, USA, pp. 855–864. https://doi.org/10.1145/2939672.2939754.

Han, S., Zhu, K., Zhou, M., Cai, X., 2021. Information-Utilization-Method-Assisted Multimodal Multiobjective Optimization and Application to Credit Card Fraud Detection. IEEE Trans. Comput. Social Syst. 1–14. Conference Name: IEEE Transactions on Computational Social Systems. https://doi.org/10.1109/TCSS.2021.3061439.

Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (roc) curve. Radiology 143, 29–36. https://doi.org/10.1148/radiology.143.1.7063747.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Ikram Ul Haq, I., Gondal, P.V., Robert, L., 2016. Generating synthetic datasets for experimental validation of fraud detection, vol. 170.

Ingole, S., Kumar, A., Prusti, D., Rath, S.K., 2021. Service-Based Credit Card Fraud Detection Using Oracle SOA Suite. SN Comput. Sci. 2 (3), 161. https://doi.org/10.1007/s42979-021-00539-2.

Ishwarappa, J., 2015. Anuradha, A brief introduction on big data 5vs characteristics and hadoop technology. Procedia Comput. Sci. 48, 319–324.

Itoo, F., Meenakshi, Singh, S., 2021. Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection. Int. J. Informat. Technol. 13(4), 1503–1511. https://doi.org/10.1007/s41870-020-00430-y. URL https://doi.org/10.1007/s41870-020-00430-y.

Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P.-E., He-Guelton, L., Caelen, O., 2018. Sequence classification for credit-card fraud detection. Expert Syst. Appl. 100, 234–245. https://doi.org/10.1016/j.eswa.2018.01.037. URL https://www.sciencedirect.com/science/article/pii/S0957417418300435.

Kanika, Singla, J., 2020. A Survey of Deep Learning based Online Transactions Fraud Detection Systems. In: 2020 International Conference on Intelligent Engineering and Management (ICIEM), pp. 130–136. https://doi.org/10.1109/ICIEM48762.2020.9160200.

Kazemi, B., Abhari, A., 2020. Content-based node2vec for representation of papers in the scientific literature. Data Knowledge Eng. 127, 101794. https://doi.org/10.1016/j.datak.2020.101794. URL https://www.sciencedirect.com/science/article/pii/S0169023X1830185X.

Kewei, X., Peng, B., Jiang, Y., Lu, T., 2021. A Hybrid Deep Learning Model For Online Fraud Detection. In: 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), pp. 431–434. https://doi.org/10.1109/ICCECE51280.2021.9342110.

Kim, E., Lee, J., Shin, H., Yang, H., Cho, S., Nam, S.-K., Song, Y., Yoon, J.-A., Kim, J.-I., 2019. Champion-challenger analysis for credit card fraud detection: Hybrid ensemble and deep learning. Expert Syst. Appl. 128, 214–224. https://doi.org/10.1016/j.eswa.2019.03.042. URL https://www.sciencedirect.com/science/article/pii/S0957417419302167.

Kotu, V., Deshpande, B., 2019. Chapter 13 - anomaly detection, in: V. Kotu, B. Deshpande (Eds.), Data Science (Second Edition), second edition Edition, Morgan Kaufmann, pp. 447–465. https://doi.org/10.1016/B978-0-12-814761-0.00013-7. https://www.sciencedirect.com/science/article/pii/B9780128147610000137.

Kumari, P., Mishra, S.P., 2019. Analysis of credit card fraud detection using fusion classifiers. In: Behera, H.S., Nayak, J., Naik, B., Abraham, A. (Eds.), Computational Intelligence in Data Mining. Springer Singapore, Singapore, pp. 111–122.

Leng, Q., Qi, H., Miao, J., Zhu, W., Su, G., 2015. One-class classification with extreme learning machine 2015 412957, publisher: Hindawi Publishing Corporation. https://doi.org/10.1155/2015/412957.

Li, X., Han, S., Zhao, L., Gong, C., Liu, X., 2017. New Dandelion Algorithm Optimizes Extreme Learning Machine for Biomedical Classification Problems. Comput. Intell. Neurosci. 2017, 4523754. https://doi.org/10.1155/2017/4523754, publisher: Hindawi.

Li, Y., Zhang, S., Yin, Y., Xiao, W., Zhang, J., 2018. Parallel one-class extreme learning machine for imbalance learning based on bayesian approach. https://doi.org/10.1007/s12652-018-0994-x.

Li, L., Fan, Y., Tse, M., Lin, K.-Y., 2020. A review of applications in federated learning. Comput. Ind. Eng. 149, 106854. https://doi.org/10.1016/j.cie.2020.106854. URL https://www.sciencedirect.com/science/article/pii/S0360835220305532.

Li, Z., Huang, M., Liu, G., Jiang, C., 2021. A hybrid method with dynamic weighted entropy for handling the problem of class imbalance with overlap in credit card fraud detection. Expert Syst. Appl. 175, 114750. https://doi.org/10.1016/j.eswa.2021.114750. URL https://www.sciencedirect.com/science/article/pii/S0957417421001913.

Liu, D., Lee, J.-H., 2021. Cfledger: Preventing chargeback fraud with blockchain. ICT Express. https://doi.org/10.1016/j.icte.2021.06.001. URL https://www.sciencedirect.com/science/article/pii/S2405959521000771.

Liu, F.T., Ting, K.M., Zhou, Z.-H., 2008. Isolation forest. In: 2008 Eighth IEEE International Conference on Data Mining, pp. 413–422. https://doi.org/10.1109/ICDM.2008.17.

Liu, F.T., Ting, K.M., Zhou, Z.-H., 2012. Isolation-based anomaly detection. ACM Trans. Knowl. Discov. Data 6 (1). https://doi.org/10.1145/2133360.2133363.

Liu, W., Wang, X., Peng, W., 2020. State of the art: Secure mobile payment. IEEE Access 8, 13898–13914. https://doi.org/10.1109/ACCESS.2019.2963480.

Lucas, Y., Jurgovsky, J., 2020. Credit card fraud detection using machine learning: A survey, arXiv:2010.06479 [cs]ArXiv: 2010.06479. http://arxiv.org/abs/2010.06479.

Lucas, Y., Portier, P.-E., Laporte, L., Calabretto, S., He-Guelton, L., Oblé, F., Granitzer, M., 2019. Dataset Shift Quantification for Credit Card Fraud Detection. In: 2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), pp. 97–100. https://doi.org/10.1109/AIKE.2019.00024.

Lucas, Y., Portier, P.-E., Laporte, L., He-Guelton, L., Caelen, O., Granitzer, M., Calabretto, S., 2020. Towards automated feature engineering for credit card fraud detection using multi-perspective hmms. Future Generat. Comput. Syst.

102, 393–402. https://doi.org/10.1016/j.future.2019.08.029. URL https://www.sciencedirect.com/science/article/pii/S0167739X19300664.
Mahmoudi, N., Duman, E., 2015. Detecting credit card fraud by modified fisher discriminant analysis. Expert Syst. Appl. 42 (5), 2510–2516. https://doi.org/10.1016/j.eswa.2014.10.037. URL https://www.sciencedirect.com/science/article/pii/S0957417414006617.
Mell, T.G.P., 2009. The nist definition of cloud computing, national institute of standards and technology.
Mirjalili, S., Lewis, A., 2016. The whale optimization algorithm. Adv. Eng. Softw. 95, 51–67. https://doi.org/10.1016/j.advengsoft.2016.01.008. URL https://www.sciencedirect.com/science/article/pii/S0965997816300163.
Mittal, S., Tyagi, S., 2020. Computational Techniques for Real-Time Credit Card Fraud Detection. In: Gupta, B.B., Perez, G.M., Agrawal, D.P., Gupta, D. (Eds.), Handbook of Computer Networks and Cyber Security: Principles and Paradigms. Springer International Publishing, Cham, pp. 653–681. https://doi.org/10.1007/978-3-030-22277-2_26.
Mohammed, R.A., Wong, K.-W., Shiratuddin, M.F., Wang, X., 2018. Scalable Machine Learning Techniques for Highly Imbalanced Credit Card Fraud Detection: A Comparative Study. In: Geng, X., Kang, B.-H. (Eds.), PRICAI 2018: Trends in Artificial Intelligence, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 237–246. https://doi.org/10.1007/978-3-319-97310-4_27.
Mohebali, B., Tahmassebi, A., Meyer-Baese, A., Gandomi, A.H., 2020. Chapter 14 - probabilistic neural networks: a brief overview of theory, implementation, and application. In: Samui, P., Tien Bui, D., Chakraborty, S., Deo, R.C. (Eds.), Handbook of Probabilistic Models. Butterworth-Heinemann, pp. 347–367. https://doi.org/10.1016/B978-0-12-816514-0.00014-X.
Olowookere, T.A., Adewale, O.S., 2020. A framework for detecting credit card fraud with cost-sensitive meta-learning ensemble approach. Scientific African 8, e00464. https://doi.org/10.1016/j.sciaf.2020.e00464. URL https://www.sciencedirect.com/science/article/pii/S2468227620302027.
Padmanabhuni, S.S.H., Kandukuri, A.S., Prusti, D., Rath, S.K., 2019. Detecting Default Payment Fraud in Credit Cards. In: 2019 IEEE International Conference on Intelligent Systems and Green Technology (ICISGT), pp. 15–153. https://doi.org/10.1109/ICISGT44072.2019.00018.
Perozzi, B., Al-Rfou, R., Skiena, S., 2014. Deepwalk: Online learning of social representations. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, Association for Computing Machinery, New York, NY, USA, p. 701–710. https://doi.org/10.1145/2623330.2623732.
Popat, R.R., Chaudhary, J., 2018. A Survey on Credit Card Fraud Detection Using Machine Learning. In: 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), 2018, pp. 1120–1125. https://doi.org/10.1109/ICOEI.2018.8553963.
Raghuwanshi, B.S., Shukla, S., 2018. Class-specific extreme learning machine for handling binary class imbalance problem. Neural Networks 105, 206–217. https://doi.org/10.1016/j.neunet.2018.05.011. URL https://www.sciencedirect.com/science/article/pii/S0893608018301734.
Rai, A.K., Dwivedi, R.K., 2020. Fraud Detection in Credit Card Data Using Machine Learning Techniques. In: Bhattacharjee, A., Borgohain, S.K., Soni, B., Verma, G., Gao, X.-Z. (Eds.), Machine Learning, Image Processing, Network Security and Data Sciences, Communications in Computer and Information Science. Springer, Singapore, pp. 369–382. https://doi.org/10.1007/978-981-15-6318-8_31.
RB, A., KR, S.K., 2021. Credit card fraud detection using artificial neural network. Global Trans. Proc. 2 (1), 35–41, 1st International Conference on Advances in Information, Computing and Trends in Data Engineering (AICDE - 2020). https://doi.org/10.1016/j.gltp.2021.01.006. https://www.sciencedirect.com/science/article/pii/S2666285X21000066.
Robin, D.R., 1993. J. Off. Stat. 9, 461–468.
Roy, A., Sun, J., Mahoney, R., Alonzi, L., Adams, S., Beling, P., 2018. Deep learning detecting fraud in credit card transactions. In: 2018 Systems and Information Engineering Design Symposium (SIEDS), pp. 129–134. https://doi.org/10.1109/SIEDS.2018.8374722.
Rtayli, N., Enneya, N., 2020. Enhanced credit card fraud detection based on SVM-recursive feature elimination and hyper-parameters optimization. J. Informat. Sec. Appl. 55, 102596. https://doi.org/10.1016/j.jisa.2020.102596. URL https://www.sciencedirect.com/science/article/pii/S221421262030764X.
Saito, T., Rehmsmeier, M., 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PloS One 10(3), e0118432–e0118432, publisher: Public Library of Science. https://doi.org/10.1371/journal.pone.0118432. https://pubmed.ncbi.nlm.nih.gov/25738806.
Sammut, C., Webb, G.I., 2010. Encyclopedia of Machine Learning. Springer, Boston, MA.
Schafer, J.B., Frankowski, D., Herlocker, J., Sen, S., 2007. Collaborative Filtering Recommender Systems. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp. 291–324. https://doi.org/10.1007/978-3-540-72079-9_9.
Schmidhuber, J., 2015. Deep learning in neural networks: An overview. Neural Networks 61, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003. URL https://www.sciencedirect.com/science/article/pii/S0893608014002135.
Sepa report, 2022. https://www.ecb.europa.eu/pub/pdf/cardfraud/ecb.cardfraudreport202110cac4c418e8.en.pdf, accessed: 2022-01-25.
Shapira, F.R.R., 2015.
Shvachko, K.V., Kuang, H., Radia, S.R., Chansler, R.J., 2010. The hadoop distributed file system. In: 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST), pp. 1–10.
Soltani Halvaiee, N., Akbari, M.K., 2014. A novel model for credit card fraud detection using Artificial Immune Systems. Appl. Soft Comput. 24, 40–49. https://doi.org/10.1016/j.asoc.2014.06.042. URL https://www.sciencedirect.com/science/article/pii/S1568494614003160.
Specht, D., 1990. Probabilistic neural networks and the polynomial adaline as complementary techniques for classification. IEEE Trans. Neural Networks 1 (1), 111–121. https://doi.org/10.1109/72.80210.
Specht, D.F., 1990. Probabilistic neural networks. Neural Networks 3 (1), 109–118. https://doi.org/10.1016/0893-6080(90)90049-Q. URL https://www.sciencedirect.com/science/article/pii/089360809090049Q.
Speiser, J.L., Miller, M.E., Tooze, J., Ip, E., 2019. A comparison of random forest variable selection methods for classification prediction modeling. Expert Syst. Appl. 134, 93–101. https://doi.org/10.1016/j.eswa.2019.05.028. URL https://www.sciencedirect.com/science/article/pii/S0957417419303574.
Srivastava, A., Yadav, M., Basu, S., Salunkhe, S., Shabad, M., 2016. Credit card fraud detection at merchant side using neural networks. In: 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 667–670.
Sudha, C., Akila, D., 2021. Majority vote ensemble classifier for accurate detection of credit card frauds. Mater. Today: Proc. https://doi.org/10.1016/j.matpr.2021.01.616. URL https://www.sciencedirect.com/science/article/pii/S2214785321007112.
Sudha, C., Akila, D., 2021. Credit Card Fraud Detection System based on Operational Transaction features using SVM and Random Forest Classifiers. In: 2021 2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM), pp. 133–138. https://doi.org/10.1109/ICCAKM50778.2021.9357709.
Sun, Y., Wong, A.K.C., Kamel, M.S., 2009. Classification of imbalanced data: A review. Int. J. Pattern Recognit Artif Intell. 23 (04), 687–719. https://doi.org/10.1142/S0218001409007326.
Thennakoon, A., Bhagyani, C., Premadasa, S., Mihiranga, S., Kuruwitaarachchi, N., 2019. Real-time Credit Card Fraud Detection Using Machine Learning. In: 2019 9th International Conference on Cloud Computing, Data Science Engineering (Confluence), pp. 488–493. https://doi.org/10.1109/CONFLUENCE.2019.8776942.
Tran, T.C., Dang, T.K., 2021. Machine Learning for Prediction of Imbalanced Data: Credit Fraud Detection. In: 2021 15th International Conference on Ubiquitous Information Management and Communication (IMCOM), pp. 1–7. https://doi.org/10.1109/IMCOM51814.2021.9377352.
Vavilapalli, V.K., Murthy, A.C., Douglas, C., Agarwal, S., Konar, M., Evans, R., Graves, T., Lowe, J., Shah, H., Seth, S., Saha, B., Curino, C., O'Malley, O., Radia, S., Reed, B., Baldeschwieler, E., 2013. Apache hadoop yarn: Yet another resource negotiator. In: Proceedings of the 4th Annual Symposium on Cloud Computing, SOCC '13, Association for Computing Machinery, New York, NY, USA, https://doi.org/10.1145/2523616.2523633.
Vishwakarma, P.P., Tripathy, A.K., Vemuru, S., 2021. Fraud detection in nfc-enabled mobile payments: A comparative analysis. In: Raj, J.S., Iliyasu, A.M., Bestak, R., Baig, Z.A. (Eds.), Innovative Data Communication Technologies and Application. Springer Singapore, Singapore, pp. 397–403.
Vuttipittayamongkol, P., Elyan, E., 2020. Neighbourhood-based undersampling approach for handling imbalanced and overlapped data. Inf. Sci. 509, 47–70.
Wang, H., Zheng, H., 2013. True Positive Rate. Springer, New York, New York, NY, pp. 2302–2303. https://doi.org/10.1007/978-1-4419-9863-7_255.
Wang, N., Liang, H., Jia, Y., Ge, S., Xue, Y., Wang, Z., 2016. Cloud computing research in the is discipline: A citation/co-citation analysis. Decis. Support Syst. 86, 35–47. https://doi.org/10.1016/j.dss.2016.03.006. URL https://www.sciencedirect.com/science/article/pii/S0167923616300409.
Wang, C., Wang, Y., Ye, Z., Yan, L., Cai, W., Pan, S., 2018. Credit Card Fraud Detection Based on Whale Algorithm Optimized BP Neural Network. In: 2018 13th International Conference on Computer Science Education (ICCSE), pp. 1–4. https://doi.org/10.1109/ICCSE.2018.8468855. iSSN: 2473-9464.
Wischik, D., Handley, M., Braun, M.B., 2008. The resource pooling principle. SIGCOMM Comput. Commun. Rev. 38 (5), 47–52. https://doi.org/10.1145/1452335.1452342.
Wiścicka-Fernando, M., 2021. The use of mobile technologies in online shopping during the covid-19 pandemic - an empirical study. Procedia Comput. Sci. 192, 3413–3422, knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 25th International Conference KES2021. doi: https://doi.org/10.1016/j.procs.2021.09.114. https://www.sciencedirect.com/science/article/pii/S1877050921018536.
Xiao, W., Zhang, J., Li, Y., Zhang, S., Yang, W., 2017. Class-specific cost regulation extreme learning machine for imbalanced classification. Neurocomput. 261, 70–82, advances in Extreme Learning Machines (ELM 2015). https://doi.org/10.1016/j.neucom.2016.09.120. URL https://www.sciencedirect.com/science/article/pii/S0925231217302199.
Xu, C., Shen, J., Du, X., Zhang, F., 2018. An intrusion detection system using a deep neural network with gated recurrent units. IEEE Access 6, 48697–48707. https://doi.org/10.1109/ACCESS.2018.2867564.
Yamashita, R., Nishio, M., Do, R.K.G., Togashi, K., 2018. Convolutional neural networks: an overview and application in radiology. Insights Imaging 9 (4), 611–629. https://doi.org/10.1007/s13244-018-0639-9.
Yang, W., Zhang, Y., Ye, K., Li, L., Xu, C.-Z., 2019. FFD: A Federated Learning Based Method for Credit Card Fraud Detection. In: Chen, K., Seshadri, S., Zhang, L.-J. (Eds.), Big Data - BigData 2019, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 18–32. https://doi.org/10.1007/978-3-030-23551-2_2.

Yang, Z., Yin, C., Jin, C., Ning, J., Zhou, J., 2021. Lightweight delegated authentication with identity fraud detection for cyber-physical systems. In: Proceedings of the 7th ACM on Cyber-Physical System Security Workshop, CPSS '21, Association for Computing Machinery, New York, NY, USA, p. 17–28. https://doi.org/10.1145/3457339.3457984.

Yen, S.-J., Lee, Y.-S., 2006. Under-Sampling Approaches for Improving Prediction of the Minority Class in an Imbalanced Dataset. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp. 731–740. https://doi.org/10.1007/978-3-540-37256-1_89.

Yue, B., Fu, J., Liang, J., 2018. Residual recurrent neural networks for learning sequential representations. Information 9 (3). https://doi.org/10.3390/info9030056. URL https://www.mdpi.com/2078-2489/9/3/56.

Yue, C., Qu, B., Liang, J., 2018. A multiobjective particle swarm optimizer using ring topology for solving multimodal multiobjective problems. IEEE Trans. Evol. Comput. 22 (5), 805–817. https://doi.org/10.1109/TEVC.2017.2754271.

Zepeda-Mendoza, M.L., Resendis-Antonio, O., 2013. Hierarchical Agglomerative Clustering. Springer, New York, New York, NY, pp. 886–887. https://doi.org/10.1007/978-1-4419-9863-7_1371.

Zhang, X., Han, Y., Xu, W., Wang, Q., 2021. Hoba: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture. Inf. Sci. 557, 302–316. https://doi.org/10.1016/j.ins.2019.05.023. URL https://www.sciencedirect.com/science/article/pii/S002002551930427X.

Zheng, P., 2020. Dynamic Fraud Detection via Sequential Modeling Master's thesis. University of Arkansas, Fayetteville, Fayetteville, USA.

Zheng, L., Liu, G., Yan, C., Jiang, C., 2018. Transaction fraud detection based on total order relation and behavior diversity. IEEE Trans. Comput. Social Syst. 5 (3), 796–806. https://doi.org/10.1109/TCSS.2018.2856910.

Zhou, H., Sun, G., Fu, S., Wang, L., Hu, J., Gao, Y., 2021. Internet Financial Fraud Detection Based on a Distributed Big Data Approach With Node2vec, IEEE Access 9, 43378–43386, conference Name: IEEE Access. https://doi.org/10.1109/ACCESS.2021.3062467.

Zhu, H., Liu, G., Zhou, M., Xie, Y., Kang, Q., 2019. Dandelion algorithm with probability-based mutation. IEEE Access 7, 97974–97985. https://doi.org/10.1109/ACCESS.2019.2927846.

Zhu, H., Liu, G., Zhou, M., Xie, Y., Abusorrah, A., Kang, Q., 2020. Optimizing Weighted Extreme Learning Machines for imbalanced classification and application to credit card fraud detection. Neurocomputing 407, 50–62. https://doi.org/10.1016/j.neucom.2020.04.078. URL https://www.sciencedirect.com/science/article/pii/S0925231220306639.

Zong, W., Huang, G.-B., Chen, Y., 2013. Weighted extreme learning machine for imbalance learning. Neurocomputing 101, 229–242. https://doi.org/10.1016/j.neucom.2012.08.010. URL https://www.sciencedirect.com/science/article/pii/S0925231212006479.