Journal of Future Artificial Intelligence and Technologies

E-ISSN: 3048-3719

Review Article

Recent Advances in Credit Card Fraud Detection: An Analytical Review of Frameworks, Methodologies, Datasets, and Challenges

Terseer Andrew Gaav, Haruna Umar Adoga *, and Timothy Moses

Department of Computer Science, Federal University of Lafia, PMB 146 Lafia, Nasarawa State 950001, Nigeria;
e-mail : gaavterseer@gmail.com; haruna.umar@science.fulafia.edu.ng; moses.timothy@science.fulafia.edu.ng
* Corresponding Author: Haruna Umar Adoga

Abstract: Credit card fraud detection (CCFD) remains a critical research domain due to the dynamic, adversarial, and highly imbalanced nature of fraudulent activities in financial systems. This study employs a systematic mapping review guided by the PRISMA 2020 guidelines. It analytically synthesizes 40 peer-reviewed and open-access studies, focusing on methodological trends, machine learning techniques, datasets, optimization strategies, and evaluation metrics. Supervised learning (SL) models, including Random Forest, Decision Trees, Support Vector Machine (SVM), and XGBoost, accounted for nearly half of the reviewed studies and consistently demonstrated strong performance. Deep learning (DL) frameworks, including Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and their variants, have demonstrated strong capabilities in capturing sequential and high-dimensional patterns of fraud. However, their effectiveness is constrained by class imbalance and dataset bias. Ensemble and hybrid models further enhanced predictive accuracy but introduced higher computational costs and lower interpretability. A key finding is the heavy reliance on the ECCT 2013 dataset (used in more than half of the reviewed studies), which supports reproducibility but limits generalizability to modern fraud contexts. Optimization strategies, such as the Synthetic Minority Oversampling Technique (SMOTE), hyperparameter tuning, and dimensionality reduction, have proven effective in improving recall and reducing false negatives; however, they have been inconsistently applied. Similarly, evaluation metrics were uneven, with accuracy dominating (reported in 75% of studies), while more informative measures such as recall, F1-score, Precision-Recall curves (AUPRC), and Matthews Correlation Coefficient (MCC) received less emphasis despite their relevance to imbalanced data. Overall, while many models achieve high accuracy in controlled environments, their scalability, adaptability, and trustworthiness in real-world deployment remain limited. Future research should prioritize cross-dataset evaluations, standardized metrics, and emerging paradigms such as federated learning, self-supervised approaches, and explainable AI to guide the development of robust and deployable fraud detection systems.

Keywords: Class Imbalance; Credit card; Deep learning; Fraud detection; Machine learning; Model interpretability; Optimization; Supervised learning.

Received: July, 23rd 2025
Revised: August, 27th 2025
Accepted: August, 28th 2025
Published: September, 3rd 2025
Curr. Ver.: September, 3rd 2025

1. Introduction

The global shift toward a cashless economy has positioned credit cards as the backbone of digital commerce, valued for their convenience, efficiency, and widespread acceptance. However, this increasing reliance has amplified the risks of credit card fraud (CCF), which remains one of the most pervasive threats in the financial sector [1]. For example, financial institutions in the United Kingdom recorded losses exceeding €574.2 million in 2020 due to

fraudulent card activities [2]. Detecting such fraud is challenging due to evolving attack patterns, data imbalance, and the need to distinguish genuine transactions from fraudulent ones in real-time [3], [4].

To address these challenges, machine learning (ML) and artificial intelligence (AI) methods have been widely adopted, with evidence suggesting that they offer superior adaptability compared to traditional rule-based systems. ML models have demonstrated strong predictive power on large, complex datasets [5], while deep learning (DL) methods, such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and hybrid architectures, have shown improved capacity to capture nonlinear fraud patterns [6], [7], [8], [9]. More recent directions include AutoML for automated model selection and tuning [10] and explainable AI (xAI) frameworks that balance predictive accuracy with interpretability [11]. Together, these approaches highlight a growing ecosystem of methods, yet a systematic synthesis is needed to consolidate their strengths, weaknesses, and real-world implications.

2. Background and Motivation

Several researchers have conducted surveys on CCFD. The summary of those surveys is presented in Table 1. This section highlights the key contributions made by existing surveys, providing the basis for our study.

Recent surveys on CCFD have explored diverse methodological directions, as summarized in Table 1. These range from traditional ML methods such as Artificial Neural Networks (ANN), Support Vector Machine (SVM), and K-nearest neighbor KNN [12] to advanced ensemble and gradient boosting techniques (LightGBM, XGBoost, CatBoost) [13], [14]. Some reviews emphasize feature extraction and classification [15], while others benchmark ML models on real-world datasets [14], [16]. Hybrid frameworks that integrate ML with metaheuristic algorithms have been introduced for adaptive detection [17], alongside DL models (CNNs, RNNs, LSTMs) that capture complex fraud patterns [18]. Additional work has investigated privacy-preserving federated learning (FL) [2], ensemble models with sampling strategies [19], and links between fraud detection and risk management [20].

Despite these contributions, prior surveys share common limitations. Many focus narrowly on isolated techniques, such as LightGBM [13] or CatBoost [14], without conducting systematic cross-framework comparisons. Others emphasize emerging AI trends or financial risks [20], [21] but overlook optimization strategies, dataset suitability, or metric trade-offs issues, which are critical in imbalanced fraud detection contexts. While studies such as [15] and [12] highlighted feature engineering and the transition from rule-based to intelligent systems, they did not assess interactions between frameworks, datasets, and evaluation methods. Similarly, DL and hybrid-oriented surveys [17], [18] advanced accuracy but neglected interpretability, scalability, and optimization depth. Even privacy-aware surveys [2], [22] remained limited to input decentralization without linking evaluation metrics or datasets.

This fragmentation underscores the need for a comprehensive review that unifies perspectives across methods, datasets, optimization strategies, and evaluation metrics. Earlier surveys have described these components independently, but none have systematically connected them into a holistic synthesis. Furthermore, newer approaches, such as AutoML, xAI, and FL, are only beginning to emerge and remain underrepresented. To address this gap, the present study reviews 40 recent works (2021–2025), integrating ML, DL, hybrid, AutoML, xAI, and privacy-preserving frameworks into a single analytical framework. The novelty of this review lies in linking frameworks, datasets, optimization strategies, and metrics to provide practice-oriented insights for interpretability, scalability, and compliance in real-world fraud detection systems. The main objectives of this review are to:

- Examine the key challenges and research problems that necessitate the adoption of frameworks for CCFD.
- Identify and analyze the methodologies employed for CCFD, with a focus on their effectiveness.
- Investigate the ML techniques and other frameworks used in CCFD.
- Review datasets commonly used in CCFD research, focusing on their accessibility, applicability, and adoptability.
- Evaluate the optimization techniques integrated into CCFD models and their impact on performance.
- Assess the relevance of evaluation metrics employed in CCFD research.

- Understand the limitations of current CCFD models.

Therefore, the key contributions of this study to the current body of knowledge in CCFD are:

Table 1. Summary of existing surveys on credit card fraud detection (CCFD).

Ref/Year	Survey Scope	Findings	Revealed Existing Challenges
[21] 2023	Class imbalance and ML/AI advances.	Traditional models dominate, largely due to the lack of adoption of real-time, IoT, and advanced AI-based techniques.	Limited exploration of deep learning, reliance on unsuitable metrics, and inadequate big data/cloud usage.
[15] 2021	Feature extraction and classification.	ML models outperform others; existing classification schemes struggle with feature relationships	Inefficiency in extracting meaningful features, difficulty linking diverse features, and complexity of datasets.
[16] 2021	Comparative study of ML models using real-world US datasets.	Random Forest outperformed Logistic Regression and XGBoost; real-time ML-based detection shown to be effective.	The ineffectiveness of traditional methods and the lack of real-time prevention tools.
[12] 2025	Reviewed traditional ML methods (ANN, SVM, K-NN, etc.).	Conventional methods are useful but lack adoptability; rule-based approaches are limited.	Lack of benchmark datasets, weak detection of rare/new patterns, and trade-offs between interpretability and usability.
[13] 2024	Reviewed ML models with a focus on LightGBM	LightGBM is effective on imbalanced and high-dimensional data.	Lack of real-time adoptability, narrow algorithm focus, and limited metric analysis.
[17] 2025	Systematic review of ML, DL, and Metaheuristic Optimization (MHO) techniques.	DL and MHO approaches show strong results; hybrid and ensemble models are promising but costly.	Scalability, data imbalance, hyperparameter tuning, explainability, and underuse of hybrid models.
[2] 2022	Comparison of ML techniques and federated learning-ANN hybrid for CCFD	Federated ANN models enhance accuracy while preserving privacy, making them effective for collaborative fraud detection.	Large and real-time datasets, data imbalance, and institutional policy barriers.
[20] 2023	Broad review of credit card (fraud detection, risk assessment, and data management)	ML improves fraud detection and risk assessment; big data frameworks are critical.	Data inconsistencies, data security, and ethical concerns in model training.
[22] 2025	Evaluation of the Luhn algorithm and DFA-based enhancement for credit card number validation	DFA-enhanced Luhn achieves high accuracy on string length validation.	Inability of the Luhn to detect string length errors; risks of system failure and unauthorized transactions.
[19] 2024	Ensemble ML model for CCFD using data balance.	Ensemble models, SVM, KNN, RF, Bagging, and Boosting outperform traditional models.	Data imbalance, concept drift, real-time detection, false positives/negatives, model efficiency.
[18] 2024	DL-based models for CCFD	DL models demonstrate strong accuracy and robustness, guiding selecting appropriate architectures.	Lack of interpretability, training complexity, and limited adoption of DL in practice.
[14] 2025	Evaluating CatBoost, XGBoost, and LightGBM for CCFD using a large-scale dataset.	CatBoost outperforms others; top 10 features identified.	Lacks real-time monitoring, feature selection, and model adaptability to evolving threats.

- Provides a holistic comparative synthesis of ML, DL, hybrid, AutoML, and privacy-preserving approaches.

- Critically evaluates benchmark datasets, distinguishing between mature and still-challenging ones.v
- Integrates and analyzes optimisation strategies for enhancing model robustness and scalability.
- Clarifies trade-offs among evaluation metrics, with emphasis on recall in imbalanced fraud detection.
- Establishes the urgency for an updated review beyond outdated models.
- Offers forward-looking recommendations for adaptive, interpretable, and compliance-ready CCFD systems.

This paper is structured into seven (7) sections: Section 1 is the introduction. Section 2 presents the background and motivation; Section 3 outlines the review methodology; Section 4 discusses the Overview of CCFD frameworks. Section 5 presents an overview of the frameworks adopted for CCFD. Section 6: Present result and discussion. Section 7 concludes the study and underscores future research directions.

3. Overview of Credit Card Fraud Detection Frameworks

There are four frameworks identified for CCFD. They include ML, DL, automated ML (AutoML), and explainable AI (xAI).

3.1 Machine Learning Techniques

Machine learning algorithms are trained by utilizing labeled, unlabeled, or hybrid datasets. Consequently, the nature of the requisite data for a particular task dictates the archetype of the machine learning model that is formulated. Consequently, the fundamental categories of machine learning have been established, which are supervised, unsupervised, semi-supervised, and reinforcement [23].

- Supervised Learning (SL): Supervised Machine Learning (SML) has continued to play a pivotal role in shaping the development of designed detection systems for Credit Card Fraud. Its strength lies in its ability to derive insights from historical transaction data where the expected outcomes are known (labeled). Supervised algorithms operate only on labeled datasets; that is, they learn from the relationship between input features and expected outputs to form predictive or detective models applicable to new, unseen instances. The essence of the technique involves supplying an ML model with structured example datasets, typically numerical representations of various transaction attributes, then allowing the machine to iteratively refine its internal parameters to enhance its prediction or detection accuracy. Depending on the labeled output, models are generally classified into either classification, where predictions fall into defined categories, or regression, where outcomes vary along a continuous scale [24]. The classification methods have seen widespread adoption in fraud detection, with researchers examining numerous approaches to address issues such as class or imbalanced datasets, overfitting, and the opacity of model decisions.
- Unsupervised Learning (UL): This is a type of ML that works with unlabeled datasets. Unlike supervised ML, it accepts datasets that do not have a predefined output. That is, it does not rely on known categories or feedback to make sense of the data. It instead tries to find patterns, groupings, or anomalies on its own. This approach is commonly used for clustering, anomaly detection, and estimating underlying distributions. Its application can never be overemphasized in fraud detection, where there is often very little labeled data to train on or the data is heavily imbalanced [24]. In this case, unusual transactions, such as fraudulent ones, can sometimes stand out simply by how different they appear from the usual data.
- Deep Learning (DL): DL is a branch of ML that is concerned with pattern identification in complex and high-volume datasets. Its architecture is based on an ANN, which encompasses CNNs and LSTM network models. These models operate through layers that learn features at various levels, enabling them to handle data from diverse sources, such as IoT devices, mobile systems, surveillance tools, and security logs. It provides an opportunity for automatic feature extraction. Also, it supports classification and regression tasks, making it useful in areas such as fraud detection, Natural Language Processing (NLP), and image recognition [23].

3.2. Other Frameworks for Credit Card Fraud Detection

This review identified automated ML (AutoML) and explainable AI (xAI) as the emerging frameworks employed for CCFD tasks.

- Automated Machine learning (AutoML): AutoML is defined as the automation of the predictive analytics workflow, encompassing various stages such as data processing, feature engineering, model selection, and hyperparameter tuning. It is worth mentioning that, despite the promise of comprehensive automation, Current solutions with AutoML often fall short in an attempt to ensure a full handling of the processing phase. The H20 AutoML framework is identified as a tool that stands out as the most advanced and consistently delivers superior performance in classification and regression tasks. Additionally, hyperparameter optimization is typically carried out through a random grid search, where parameter values are selected at random from a predefined range. This feature presented users with the opportunity to define stopping criteria, including the number of models or time limits, thereby enhancing computational efficiency and resource management [10].
- Explainable Artificial Intelligence (xAI): In terms of predictive accuracy, the rapid improvement of statistical ML has significantly enhanced AI capabilities; however, progress often comes at the cost of transparency, as increasingly complex models become less interpretable for humans. The xAI framework emerged as a weak response in bridging the gap between high-performing algorithms and human interpretability. xAI architecture is tailored to make AI systems transparent, traceable, and understandable; as such, it shifts the emphasis from mere prediction to explaining model behavior in ways that align with human cognition. The heatmap-based visual explanations for deep neural networks are an early record of xAI; however, the framework advancement aims to address more advanced scenarios, such as the interpretation of unsupervised models, as well as enhance explanations for human decision-making. Many researchers suggest ways to achieve truly interpretable AI, which include integrating usability, developing evaluation metrics for explanation quality, and designing interfaces that facilitate human-AI collaboration. The idea of interactive ML, where domain experts contribute contextual or causal insights, further exemplifies the evolving synergy between artificial and human intelligence. These efforts have garnered widespread attention across academia, industry, and the public sector, creating a fertile ground for practical xAI innovations [11].

3.3. Frameworks for Credit Card Fraud Detection

The review of the related studies revealed that ML algorithms, dimensionality reduction, automated ML (AutoML), and explainable AI (xAI) are the frameworks trending in the detection of credit card fraud. Figure 1 depicts an architectural representation of the identified frameworks adopted.

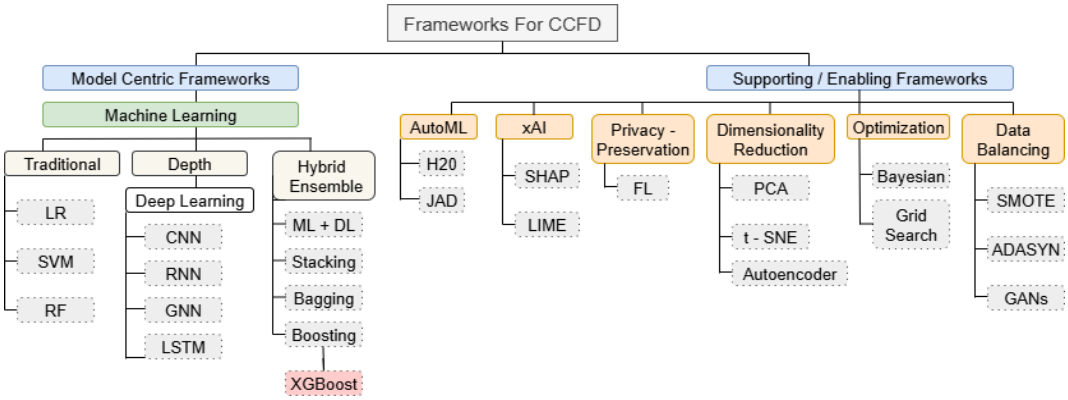

Figure 1. Architecture of MLT and frameworks employed for CCFD.

4. Review Methodology

In this review, we employed a systematic mapping approach, a recognized effective method for reviewing and synthesizing research trends and study types within a specific field.

This process is typically conducted in five key phases and is instrumental in categorizing studies, identifying gaps, and visualizing trends by mapping publication frequencies and thematic distributions over time [25]. In this paper, systematic mapping was employed to provide an overview of the current state of research on CCFD.

4.1. Overview of Systematic Mapping

Systematic mapping provides a high-level summary of research by classifying and visually analyzing existing studies. It helps to identify what has been studied, where research gaps exist, and what areas warrant future exploration [26]. This method is particularly suitable for understanding the distribution of research topics, methodologies, and outcomes within the field of CCFD.

4.2. Methodology Adopted

This study followed the updated systematic mapping guidelines proposed by recent scholars, such as [25] and [26]. The entire process was structured into several major phases: The process was divided into several key phases:

- Phase 1: Definition of Research Questions and Objectives
- Phase 2: Search Strategy and Study Selection
- Phase 3: Screening and Selection

4.2.1. Research Questions and Corresponding Research Objectives

Formulating well-defined research questions is critical for determining the scope of a systematic mapping study [26]. This study aims to analyze literature on CCFD frameworks, emphasizing research trends, existing gaps, and future research opportunities. Therefore, the research questions align with each research objective outlined in the background and motivation of this study. These questions determine the inclusion criteria, mapping categories, and analysis approach adopted in this review.

4.2.2. Search Strategies and Study Selection Process

The search strategy for this review was developed in several stages, involving the identification of keywords, formulation of search queries, and the selection of appropriate databases. The process was conducted iteratively to ensure a comprehensive exploration of the relevant literature. Keywords and search terms were derived from the research questions and were refined as the search progressed.

4.2.3. Keywords and Search Strategy Formulation

The initial search concepts include "frameworks for credit card fraud detection," "credit card fraud detection," "machine learning for credit card fraud detection," "Credit card security," and "intrusion detection system." Synonyms and related terms were also incorporated to ensure that the search captured a wide range of related studies. For instance, terms such as "cyber threats," "Model Interpretability," and "machine learning" were added in subsequent iterations. This approach ensured that the search query was inclusive, increasing the likelihood of identifying a diverse set of studies. The final search query combined these keywords, as demonstrated below:

- Query_1: ("credit card fraud detection" OR "fraud detection" OR "cyber intrusion detection" OR "credit card security").
- Query_2: ("machine learning" OR "deep learning" OR "supervised learning" OR "model interpretability" OR "automated machine learning" OR "Explainable artificial intelligence").
- Query_3: ("credit card fraud detection frameworks" OR "pattern recognition" OR "fraud monitoring").

The comprehensive search query combined these elements to identify studies related to frameworks used for detecting credit card fraud.

4.2.4. Screening and Selection Criteria

This review adopted a systematic approach to screen and select relevant studies based on predefined criteria. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were followed to ensure transparency in reporting the study selection process. After applying the research strategy across popular databases (IEEE Xplore, Elsevier, ResearchGate, and Google Scholar), a total of 635 articles were identified.

Inclusion and Exclusion Criteria: To ensure that only relevant and highly quality studies were included, the following inclusion and exclusion criteria were applied:

- Inclusion Criteria*:*
  1. Publications made from 2021 to 2025 inclusive, that is, the consideration of the most recent (up-to-date) works;
  2. Papers that explicitly applied detection frameworks for CCFD;
  3. Papers published in peer-reviewed journals or presented at reputable conferences.
- Exclusion Criteria:
  1. Papers focusing on ML applications in areas not related to CCFD;
  2. Papers that were not peer-reviewed (e.g., unpublished preprints, ResearchGate-only uploads) or that lacked sufficient methodological rigor.
  3. Papers written in languages other than English.
- Snowballing Technique*:* In addition to querying databases. The snowballing technique was employed to expand the pool of relevant studies further. Snowballing involved examining the reference lists of selected articles (backward snowballing) and identifying newer studies that may not have been retrieved in the initial search. Both forward and backward snowballing were performed to ensure a comprehensive exploration of the literature.
- Search Results*:* The final search, including snowballing and database queries, yielded a total of 635 articles. After applying the inclusion and exclusion criteria, 40 studies were selected for in-depth review. These studies form the basis of the systematic mapping analysis, providing insights into the problem and challenges that necessitate the choice of framework for CCFD, the adopted methodology, the integrated optimization methods, and their effectiveness, among other factors. The procedure for the selection of the articles for this study is presented in Table 3.

Table 3. The selection and screening process from five academic databases.

Database	Initial Search	After Duplicate Removal	Inclusion Criteria	Exclusion Criteria	Final Selected Articles
IEEE	180	40	25	15	15
Elsevier	150	33	18	16	13
ResearchGate	100	30	14	15	7
Google Scholar	205	33	21	13	5
Total	635	137	78	59	40

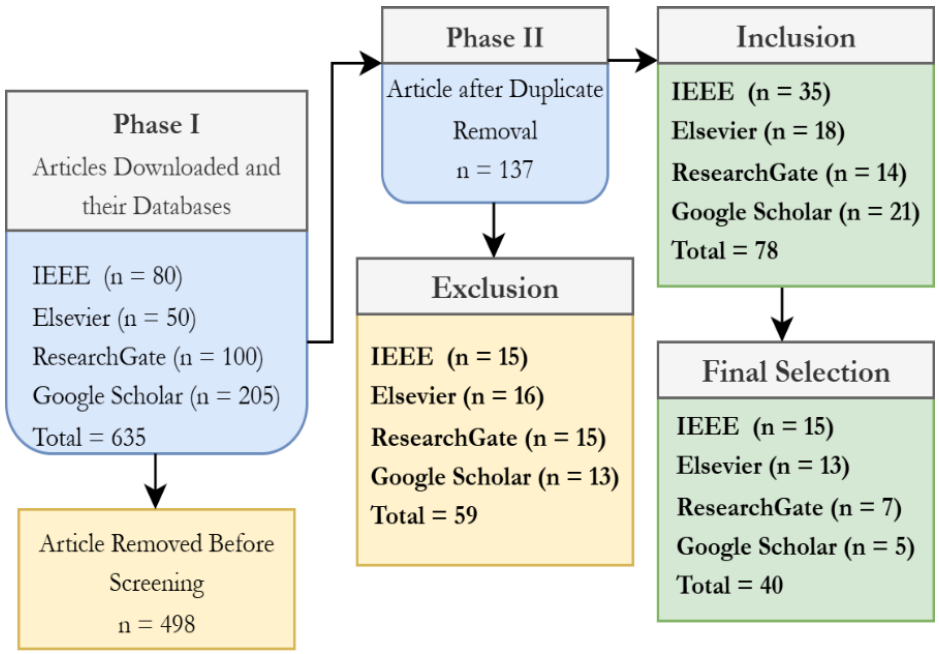

Figure 2. PRISMA 2020 flow diagram for study selection

To enhance transparency and reproducibility, the study selection process was further illustrated using a PRISMA 2020 flow diagram in Figure 2. This diagram outlines the number of records identified through database searches and snowballing, the number of duplicates removed, the screening process, the assessment of full-text eligibility, and the final number of studies included in the review. It provides a clear visual representation of how the initial 635 records were filtered down to the 40 studies analyzed in this research. It should be noted that ResearchGate is not a formal database, but rather a repository of author-uploaded papers. The official published versions were traced to their respective journals when available.

The selected databases, as depicted in Table 3 were chosen for their relevance, scope, and high-quality impact. IEEE Xplore was included for its peer-reviewed research on ML and fraud detection [27]. Elsevier journals were referenced for their broad interdisciplinary coverage in the applied sciences, which offer valuable insights into fraud analytics [27]. ResearchGate provided access to preprints, rare studies, and niche research areas, supporting a comprehensive literature review [28]. Finally, Google Scholar was utilized for its extensive scope across peer-reviewed and grey literature [29]. This combination of databases ensured a thorough analysis of frameworks for CCFD by covering a wide spectrum of scientific literature, from peer-reviewed journal articles to preprints and interdisciplinary studies.

Source Limitation: The majority of the reviewed studies were sourced from open-access (OA) journals and repositories. This reliance on OA sources ensured accessibility and transparency but may introduce bias by underrepresenting non-OA literature.

5. Literature Mapping

This section provides a structured mapping of existing research on CCFD. The organization is organized into five major themes: Traditional ML Approaches, Deep Learning Approaches, Comparative Studies of ML and DL Models, Data Augmentation and Balancing Techniques, and Deployment and Real-Time Systems.

5.1. Theme 1: Traditional Machine Learning Approaches

Table 4. The review summary of traditional ML approaches implemented for CCFD.

Ref.	Problem Addressed	Methods Adopted	MLT Used	Dataset Used	Optimization Approach	Evaluation Metrics	Effectiveness and Challenges
[30]	▪ Improving feature selection ▪ Classification robustness	Soft voting ensemble	SL: Extra Trees (ET), XGBoost, CatBoost	Balance Dataset with 28 anonymised features	Energy Valley Optimisation (EVO)	Acc, Precision, Recall, F1-Score,	▪ Outperformed 10 reviewed related works ▪ A slight drop in post-feature selection ▪ Lacks mention of a real-time application
[31]	Classification comparison	Supervised Learning	SL: SVM, DT, Naïve Bayes, RF	ECCTR 2013	Not Specified	Acc, Precision, Recall, F1-Score, MCC	▪ RF outperformed other ML models. ▪ Hyperparameter sensitivity, complex training for generative models
[32]	Improving accuracy	Supervised Learning	SL: LR, RF, KNN, Decision Tree, SVM, NB	ECCF Dataset	Hyperparameter tuning, Resampling	Acc, Precision, Recall, F1-Score, ROC-AUC	▪ RF achieved high accuracy; ▪ LR performed well. ▪ Class imbalance.
[33]	Limitations of individual classifiers in an imbalanced dataset	Supervised learning with an ensemble	SL: Adaboost, RF	Not stated	Note applied	Acc, Precision, Recall, F1-Score	▪ Both algorithm shows a good performance. ▪ Exhibition of narrow scope by the employed models.

Table 4 (cont.)

Ref.	Problem Addressed	Methods Adopted	MLT Used	Dataset Used	Optimization Approach	Evaluation Metrics	Effectiveness and Challenges
[34]	Improving classification accuracy	Sacking ensemble	SL: XGBoost, DT, RF, LightGBM, LR	PaySim Dataset	K-Means, SMOTE, ENN	F1-Score, Precision, Recall, AUPRC, ROC-AUC	▪ Excellent generalization and robustness. ▪ Overfitting, complexity, and real-time adoptability.
[35]	Enhancing CCFD accuracy and reliability)	Ensemble modeling with preprocessing	SL: RF, LR, Adaboost (soft voting)	Customer dataset and PaySim dataset	Soft-voting, feature engineering	Acc, F1-Score,Precision, AUC	▪ ensemble model outperforms individual models, achieving an accuracy. ▪ Generalizability and robustness; ▪ high computational cost.
[36]	Improving Fraud Detection	FFD and SVDD	SL: SVDD	ECCT Dataset 2013	Resampling (Under-sampling), Feature selection. Particle; Swarm Optimization	Acc	▪ Feature selection improved model accuracy. ▪ Difficulty in predicting with feature selection.

5.2. Theme 2: Deep Learning Approaches

Table 5. The review summary on deep learning models employed for CCFD

Ref.	Problem Addressed	Methods Adopted	MLT Used	Dataset Used	Optimization Approach	Evaluation Metrics	Effectiveness and Challenges
[37]	Feature selection and pattern recognition.	Deep Learning	DL: ANN, CNN	ECCTR 2013	Not Used	Acc, F1-Score, TPR, FNR	▪ CNN achieved high acc. And F1-Score ▪ Decrease in model performances.
[38]	Imbalanced Dataset	Federated Learning	DL: LSTM, CNN	Decentralized Financial Institution Dataset	SMOTE	Acc, Precision, Recall, F1-Score	▪ CNN was outperformed ▪ Sensitivity of CNN to batch; ▪ Privacy limitations in data sharing
[39]	▪ Addressing data scarcity ▪ Comparison	Ensemble Learning	DL: CN, MLP, RNN; SL: XGBoost, DT, RF	Kaggle Credit Card Fraud Dataset	▪ Over-sampling; ▪ Under-sampling; ▪ Dropout regularisation	Acc, AUC, F1-Score	▪ Outperformed all baseline models across all metrics; ▪ Robust and generalizable. ▪ Data privacy concerns; ▪ Overfitting.
[40]	Class imbalance	GAN-Based Generative Model	DL: AE, VAE, AE-GAN	ECCTR 2013	SMOTE and ADASYN	BEFS metric	▪ Generative models outperformed traditional over-sampling detection. ▪ Hyperparameter sensitivity; ▪ Complex training for generative models
[41]	Tracking sequential FD through model's ability improvement	Hybrid model	DL: LSTM, SL: XGBoost	BankG hurners	SMOTE	ACC, FI-Score, Precision	▪ LSTM outperformed XGBoost. ▪ Requires more resources, lower interpretability.

Table 5 (cont.)

Ref.	Problem Addressed	Methods Adopted	MLT Used	Dataset Used	Optimization Approach	Evaluation Metrics	Effectiveness and Challenges
[42]	Fraud detection improvement	Brain-inspired models	DL: Continuous Coupled NN (CCNN)	Kaggle CCFD Dataset	SMOTE	Acc, Precision, Recall, F1-Score	▪ CCNN outperforms traditional ML models. ▪ Real-world generalization and computational validation are needed
[43]	Learning the patterns and relationships within the CC Dataset	Integration of DL with hyperparameter tuning	DL: AE, CNN, LSTM	ECCT Dataset 2013	Hyperparameter Tuning: Random search, Bayesian optimization	Acc, Detection Rate, AUC	▪ LSTM outperformed AE and CNN. ▪ DR remains a challenge.

5.3. Theme 3: ML and DL Models Comparison

Table 6. The review summary of models' comparison methodology for CCFD

Ref.	Problem Addressed	Methods Adopted	MLT Used	Dataset Used	Optimization Approach	Evaluation Metrics	Effectiveness and Challenges
[44]	Determination of the best approach	Comparative analysis of ML and DL	SL: SVM, RF. DL: CNN, LSTM	ECCT 2013	SMOTE, ADASYN, under-sampling, cost-sensitive learning	Acc, F1-Score, Precision, Recall	▪ CNN outperformed all models. ▪ SVM improved but remained behind CNN and RF.
[45]	Imbalance dataset	Comparison of RF and SVM	SL: RF and SVM	ECCT 2013	Random under-sampling, SMOTE, Hyperparameter Tuning, Ensemble	Acc, Precision, Recall, F1-Score	▪ RF achieved a strong balance of the metrics. ▪ SVM is overfitting; ▪ class imbalance
[46]	Evaluating model performances beyond accuracy	Comparison	SL: SVM, LR, RF, AdoBoost. DL: ANN.	Kaggle Credit Card Dataset	Not specified	Acc, Precision, Recall, FI-Score	▪ SVM achieved a high recall value; ANN had high precision. ▪ Dataset imbalance; high acc misleading due to class skew.
[47]	Boost classification performance	ensemble classifier with Hybrid Resampling	DL: LSTM. SL: AdaBoost	ECCT Dataset 2013	SMOTE-ENN (Hybrid Resampling)	Sensitivity, Specificity, AUC	▪ LSTM outperformed other models. ▪ Class imbalance.
[48]	Features normalization	Comparison of SML vs DL	SL: SVM, KNN. DL: ANN	ECCT Dataset 2013 - 14	Normalization and under-sampling	Acc, Precision, Recall	▪ ANN outperformed other ML and DL models. ▪ Model training and an imbalanced dataset.
[49]	Fraud detection	Comparison of DL & SL	SL: LR, SVM. DL: ANN.	Not specified	Resampling techniques	Acc, Precision, Recall, F1-Score, MCC	▪ SVM outperformed LR and ANN. ▪ Class imbalance
[3]	Feature engineering, unbalanced dataset, and scalability	Comparison of the proposed model with existing systems	SL: RF, LR. DL: Multi-Layer Perceptron (MLP)	Real World Credit Card Fraud Dataset	Feature engineering.	Average Precision, Rank Precision, and Recall Chart.	▪ RF outperformed the existing system. ▪ Feature engineering, scalability, unbalanced data, and concept drift.

Table 6 (cont.)

Ref.	Problem Addressed	Methods Adopted	MLT Used	Dataset Used	Optimization Approach	Evaluation Metrics	Effectiveness and Challenges
[50]	Credit card fraud detection	Comparative analysis of ML Models	SL: LR, RF, Extra Trees, LGBM, XGBoost, CatBoost	ECCT Dataset 2013	None specified	Acc, Recall, F1-Score, Confusion Matrix, Training Time (TT)	▪ All models had high performances. ▪ LR and LGBM had fast TT. ▪ Trade-off between training time and model performance.
[51]	Models comparison	Categorical encoding	SL: SVM, RF, Bagging, XGBoost, DT	ECCT Dataset	None	Acc, Precision, Recall, F1-Score	▪ XGBoost achieved the highest accuracy. ▪ Class imbalance.
[48]	Features normalization	Models Comparison	SL: SVM, KNN. DL: ANN	ECCT Dataset 2013 - 14	Normalization and undersampling	Acc, Precision, Recall	▪ ANN outperformed other ML and DL models. ▪ Model training and an imbalanced dataset.

5.4 Theme 4: Data Augmentation and Balancing Techniques

Table 7. The review summary on improvement, data augmentation, and balancing techniques, adopted for CCFD

Ref.	Problem Addressed	Methods Adopted	MLT Used	Dataset Used	Optimization Approach	Evaluation Metrics	Effectiveness and Challenges
[52]	Reducing FP and FN in an imbalanced dataset	Novel ensemble-based unsupervised	USL: Isolation Forest	ECCT and Medicare Part D claims dataset	Confidence-based filtering and iterative refinement	Jaccard index (JI), precision, MCC	▪ Model outperformed iForest. ▪ Unlabeled dataset and extreme class imbalance.
[53]	Dimensional reduction	Supervised Learning with Ensemble	SL: LR, Boosting. DL: ANN,	CC Clients in Taiwan, 2005.	PCA	FP Rate, FN Rate	▪ LR with PAC outperformed others. ▪ Use of an outdated dataset
[54]	Fraud detection	Supervised Machine Learning with Resampling Techniques	SL:CatBoost, XGBoost, RF, LR, K-NN, DT, Naive Bayes, GBM, Light GBM	ECCT Dataset 2013	AIIKNN Undersampling, Stratified K-Fold cross-validation	AUC, Recall, F1-Score.	▪ The AIIKNN-CatBoost model outperformed previous models. ▪ Class imbalance
[55]	CC-not-Present Fraud Detection (CCFDP) and a method for dealing with CNP Fraud.	Hybrid of Dimensionality Reduction (DR) and SL	DR: t-SNE, PCA, SVD. SL: LR	ECCT Dataset 2013	Random Undersampling (reduction of the majority class instance)	Acc, RMSE, RRMSE, MBE, MDA.	▪ CCFDP outperforms all others on diverse sample sets. ▪ Managing high-dimensional and imbalanced datasets.
[18]	Class imbalance and dynamic fraud pattern	Hybrid GAN-RNN framework for synthetic data generation and classification	DL: GAN+ RNN (GRU, LSTM, Simple RNN)	ECCT and Brazilian Datasets	Dual-Phase training	Sensitivity, Specificity, Precision, F-Measure.	▪ GAN-GRU best for ECCT Dataset; ▪ GAN-LSTM is best for the Brazilian Dataset. ▪ Real-world implementation, computational overhead.

Table 7 (cont.)

Ref.	Problem Addressed	Methods Adopted	MLT Used	Dataset Used	Optimization Approach	Evaluation Metrics	Effectiveness and Challenges
[56]	Class imbalance, relational dependencies in transaction data	Encoder – Decoder-based GNN with feature engineering	DL: Graph Neural Networks (GNNs)	Sparkov Dataset	Graph converter; batch normalization	Precision, Recall, F1-Score, ROC	▪ The proposed system outperformed other models; the GNN captures complex merchant–customer dependencies. ▪ Difficulties in acquiring real-time location.
[9]	Capturing sequence patterns and relevant features	Sequential Modeling	DL: LSTM. Unclassified: Attention Mechanism	ECCT Dataset and Synthetic dataset using BankSim Software	UMAP for feature selection and SMOTE for dataset balancing.	Acc, Precision, Recall	▪ The model outperformed recent approaches in sensitivity. ▪ Relying on multiple techniques increases model complexity.
[7]	Fraud detection	Hybrid (DL, ML & Classification models)	DL: CNN. SL: KNN	ECCT & Text2LMG Converted Dataset	Inverse frequency method for class weighting	Acc, Sensitivity, Fi-Score	▪ Achieved high accuracy with text2IMG. ▪ Class imbalance
[57]	Capture both temporal and spatial patterns in transaction data	Hybrid BiLSTM	DL: BiLSTM, Transformer	ECCTR 2013	Ablation architectural tuning	Acc, Recall, AUC, F1-Score	▪ Outperformed traditional ML models; ▪ Excelled in unbalanced and complex time-series data. ▪ Class imbalance; ▪ Data quality issues
[58]	▪ Improving Accuracy ▪ Class Imbalance	Supervised ML	SL: Logistic Regression, Decision Tree, Random Forest	CCT Western USA	Under sampling	Acc, AUC, F1-Score, Recall, Precision, Roc	▪ Random Forest achieved high Acc and AUC. ▪ Class Imbalance
[59]	Selecting relevant features in a highly imbalanced dataset	GA-Based	SL: DT, RF, LR, Naive Bayes. DL: ANN.	ECCT, Synthetic Dataset	GA-Based for feature selection	Acc, AUC	▪ GA-RF, GA-DT, and GA-ANN outperformed traditional models. ▪ Class imbalance.
[60]	Handling a highly imbalanced dataset	Resampling	SL: SVM, LR, RF, XGBoost, DT, Extra Tree (ET)	ECCT Dataset 2013	AdoBoast	Acc, Precision, Recall, MCC, AUC	▪ XGB-AdaBoost achieved high performance. ▪ Several class imbalances. ▪ Overfitting

5.5 Theme 5: Deployment and Real-Time Systems

Table 8. The review summary on deployment and the real-time system employed for CCFD

Ref.	Problem Addressed	Methods Adopted	MLT Used	Dataset Used	Optimization Approach	Evaluation Metrics	Effectiveness and Challenges
[61]	CCFD	AutML SaaS platform	AutoML-Selected models via JAD System	ECCT Dataset 2013	Auto Hyperparameter Turning	Acc, AUC, F-Measure. Precision, Recall, Sensitivity, Specificity.	▪ Performance either matches or exceeds that of the existing system. ▪ Missed low-value frauds; ▪ reliance on AutoML decisions

Table 8 (cont.)

Ref.	Problem Addressed	Methods Adopted	MLT Used	Dataset Used	Optimization Approach	Evaluation Metrics	Effectiveness and Challenges
[8]	Real-Time detection of complex and large – Scale Financial Fraud	Comparison of DCNN, ML and DL	DL: DCNN, RNN. SL: LR, SVM, RF	Real World Credit Card Fraud Dataset	Adam, RMSprop, Adagrad	Acc, Loss Rate, Training Time	▪ Proposed system outperformed existing models like AE. ▪ RF slow with large data; SVM lacks transparency, LR prone to overfitting
[11]	Real-Time CCFD Monitoring System	i. Ensemble; ii. Federated Learning	DL: NN, GCN; Unclassified: Graph Based anomaly detection; xAI: (SHAP, LIME)	Real-World Financial Dtatsets	Enhanced via adversarial training	AUC-ROC, Precision, Recall, F1-Score	▪ High Acc, low FP, Real-Time adoptability, interpretable results. ▪ Black-box nature of deep learning; ▪ Deployment scalability, real-time operationalisation still sub-optimal
[62]	Real-Time CCFD on high imbalanced data	AI-based anomaly detection system	SL: KNN, SVM, DT, LR	Kaggle CCFD dataset	None	Precision, Recall, F1-score, ROC-AUC	▪ DT achieved best balance of precision & recall for fraud cases. ▪ SVM showed strong ROC & PR performance ▪ Issue of false negatives, and Overfitting risk in DT.

6. Analytical Synthesis and Future Directions

This section synthesizes the reviewed literature guided by the research objectives and highlights key strengths, gaps, and emerging trends in CCFD. It also outlines directions for future research. It is important to clarify that the totals reported in some subsections represent the frequency of occurrences rather than the number of distinct studies. Several studies contributed multiple methods, datasets, optimization strategies, or evaluation metrics, while others reported none in certain categories. This explains why some totals exceed or fall below the 40 studies reviewed.

6.1 Examination of the Key Challenges and Research Problems that Necessitate the Adoption Of Frameworks for CCFD

The survey of numerous studies identified a range of technical and operational challenges that necessitate the adoption of frameworks for effective CCFD. As presented in Figure 3, class imbalance emerged as the predominant research gap, representing 25% of the sampled studies [38], [58], [63]. The skewed distribution of legitimate and fraudulent transactions hinders the model's learning capacity, making it challenging to identify minority fraud cases accurately. To address this, frameworks often employ class balancing mechanisms, such as SMOTE and ADASYN [40], [41], [42].

Another frequently reported challenge was the ineffective selection and representation of fraud patterns, as observed in six studies, including [9], [37], which accounted for 15% of the reviewed articles. Redundant or irrelevant features reduce classification accuracy and increase model complexity, particularly in high-dimensional spaces. Closely related to this is the issue of classification robustness, as emphasized in works such as [30], [47], which have shown that models often fail to remain stable under adversarial or evolving fraud settings.

Additional challenges include the scarcity of labeled or real-world datasets [3], [42], the need to capture temporal and spatial fraud patterns [6], [57], difficulties with real-time fraud detection under constrained computing environments [8], [11], and limited interpretability and sequential tracking of fraud behaviors [46]. Figure 3 illustrates the distribution of these challenges, reinforcing class imbalance, feature redundancy, and evolving fraud patterns as the dominant issues across the literature.

Overall, these findings highlight that challenges in CCFD extend beyond the pursuit of high accuracy. The prominence of class imbalance underscores the inadequacy of raw ML models, which are easily biased toward majority classes—hence the widespread integration of resampling strategies and synthetic data generation in modern frameworks. Similarly, the problem of feature redundancy highlights the importance of integrating feature engineering, dimensionality reduction, and pattern recognition within frameworks to minimize computational overhead and enhance generalizability. Robustness and adaptability are equally critical, as fraud evolves dynamically, requiring ensemble or hybrid architectures that can withstand adversarial behaviors.

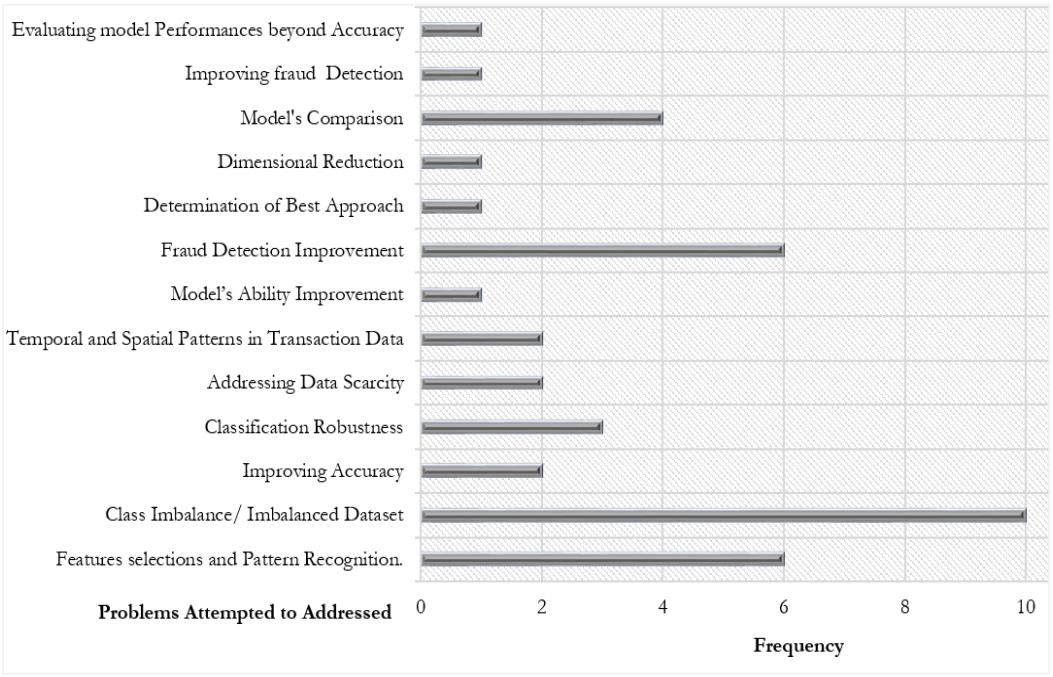

Figure 3. Revealed challenges and research problems attempted in CCFD

The scarcity of real-world datasets and latency issues in real-time detection further suggest that effective frameworks must strike a balance between predictive reliability and operational efficiency, often through FL, data augmentation, and lightweight model architectures. At the same time, interpretability and sequential tracking remain strategically important: frameworks must not only detect fraud but also explain their decisions and trace the progression of fraud in ways that are comprehensible to practitioners.

Taken together, these insights demonstrate that the adoption of frameworks in CCFD is driven not merely by technical improvements in accuracy but by the need to address interconnected problems of data imbalance, robustness, scalability, and transparency. This holistic perspective explains why framework-based approaches are becoming indispensable in modern fraud detection research and practice.

6.2. Identification of the Methodologies Adopted for CCFD and their Effectiveness.

As shown in Figure 4, a wide range of methodological frameworks have been explored for CCFD, each demonstrating varied levels of performance depending on design, dataset, and implementation contexts. Among these, models comparison method accounted for 23% of the studied surveys, while supervised ML (SML) remains one of the most frequently applied approaches, representing about 10% of the reviewed studies and appeared frequently in other methodologies adopted such as models comparison and ensemble methods,. Models such as Random Forest (RF), Decision Tree (DT), SVM, and Logistic Regression (LR) were widely implemented, with RF consistently emerging as a strong performer across accuracy, AUC, and F1-score [58]. Other supervised techniques, such as CatBoost with AIKNN undersampling, also outperformed several baselines [54]**.** Ensemble learning was equally prominent, also appearing in 23% of the surveyed articles. For instance, [30] employed soft voting

ensembles of SML such as XGBoost, Extra Trees, and CatBoost, achieving results that surpass those of related works. Meanwhile, [34] demonstrated the advantages of stacking multiple learners, noting substantial improvements in both precision and recall.

Deep learning (DL) approaches represented another major focus, owing to their ability to capture complex and high-dimensional patterns of fraud. Research [37] applied CNN and ANN to the European Credit Card Transaction Record (ECCTR) dataset, reporting strong accuracy and F1-scores despite variations in feature values. Hybrid DL models further enhanced sequential fraud detection, with BiLSTM-based combinations [7] producing high detection rates; [7] in particular showed that CNN–KNN hybrids offered robust performance despite interpretability challenges. FL has also been explored in recent studies, with a focus on privacy-preserving detection. For example, [11] integrated FL with adversarial training and Graph Neural Networks (GNN), achieving lower false positives and improved interpretability; however, real-time deployment remained challenging. Generative approaches such as GANs were further explored to address class imbalance [40], [59], generating synthetic minority samples that improved classification performance, albeit with sensitivity to hyperparameters and training stability.

Beyond these, less conventional methods have emerged. A brain-inspired Continuous Coupled Neural Network (CCNN) was proposed in [42], which surpasses traditional classifiers in accuracy but struggles with real-world generalization. Similarly, [52] investigated unsupervised isolation forests (iForest), which improved precision through iterative refinement but faced challenges with unlabeled data. Methodological comparison studies constituted approximately 22% of the surveyed articles, with examples such as [8], which benchmarked deep convolutional and recurrent networks against ML models, concluding that DCNN achieved the best performance but at a significant computational cost.

Other exploratory strategies included AutoML platforms [61], which automated model selection and tuning, and graph-based encoder–decoder frameworks [56], which exploited relational data structures. These novel methods demonstrated adaptability and accuracy gains but also raised concerns regarding transparency, scalability, and training complexity.

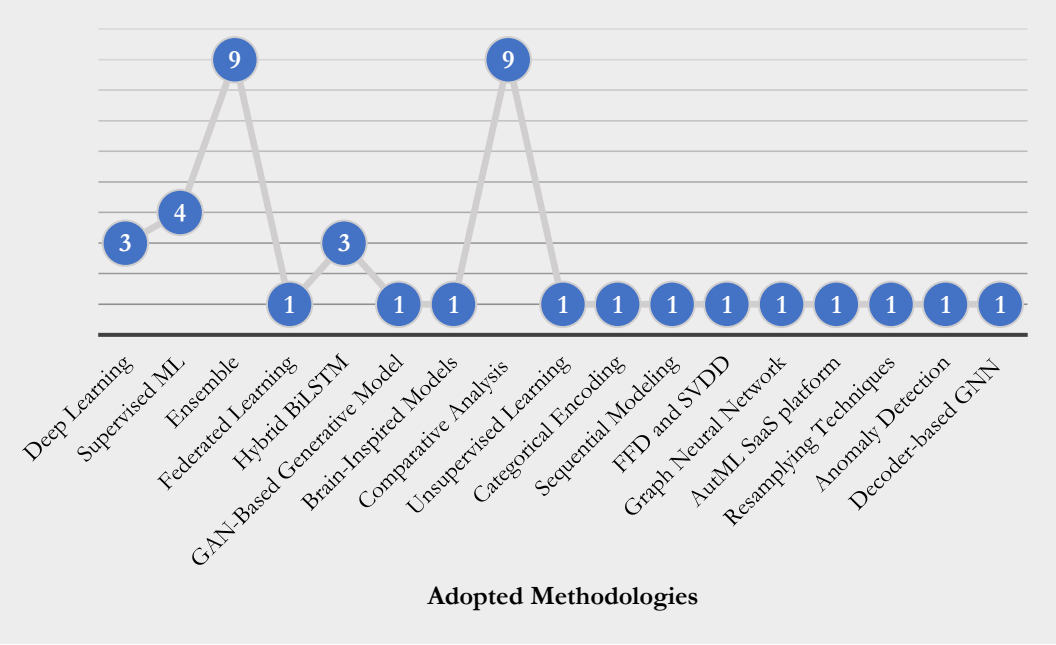

Figure 4. The revealed methodologies adopted for CCFD

The methodological mapping highlights a growing diversification of approaches in CCFD; however, the choice of methodology is often constrained by trade-offs among predictive power, interpretability, and scalability. The strong showing of RF and ensemble methods suggests that models benefiting from variance reduction and aggregated decision-making are well-suited for the imbalanced and noisy nature of fraud datasets. Their prevalence in Figure 4 indicates an increasing preference for robustness over reliance on single classifiers. Deep learning approaches, although less frequent than classical ML, reveal an important shift

in the field toward capturing sequential and high-dimensional fraud patterns. References in Table 5 show that CNNs and BiLSTMs improve sequential fraud detection, but their fluctuating performance across features underscores the sensitivity of DL to dataset quality and representation. This suggests that frameworks relying on DL must integrate strong feature engineering and interpretability components to remain practical.

The introduction of FL and GAN-based synthetic oversampling represents an attempt to bridge the data-centric challenges identified in the objective discussed in 6.1. By decentralizing model training and generating minority samples, these methodologies address both privacy concerns and class imbalance. However, their complexity, as reflected in studies such as [11] and [40], highlights why frameworks must embed such methods into larger hybrid pipelines rather than adopting them as standalone solutions.

Novel models, such as CCNNs and AutoML frameworks, though promising, highlight an emerging gap between laboratory performance and operational deployment. Table 5 and 8 respectively indicates that while they improve detection accuracy, they often struggle with generalization, interpretability, or computational feasibility. These limitations underscore the need for frameworks that can integrate unconventional methodologies selectively, striking a balance between innovation and real-world constraints.

Overall, the synthesis suggests that no single methodology provides a complete solution to fraud detection. Instead, Figure 4 reflects a methodological ecosystem where classical ML, ensemble learning, deep learning, federated approaches, and generative models complement one another. Their coexistence highlights the central role of frameworks as integrative structures, ensuring that the strengths of each methodology can be harnessed while mitigating their individual weaknesses.

6.3. Investigation of the Machine Learning Techniques (MLTs) and other Frameworks Employed for CCFD

Across the studies surveyed and summarized in Figure 5, various ML Techniques (MLTs) and complementary frameworks have been applied to CCFD, each offering unique strengths and trade-offs. A significant dominance of Supervised Learning (SL) and Deep Learning (DL) models was observed, while unsupervised learning (USL), explainable AI (xAI), AutoML, and dimensionality reduction approaches received comparatively less attention. Among the SL models, techniques such as RF, DT, SVM, LR, NB, and boosting methods, including XGBoost and CatBoost, accounted for 48% of the reviewed studies. For instance, [58] demonstrated that RF achieved higher accuracy and AUC compared to its counterparts, though class imbalance remained a persistent issue. Similarly, [64] showed that ensemble-based SL models, such as XGBoost, RF, and LightGBM, achieved excellent generalization capabilities, albeit at the expense of increased model complexity. Deep Learning frameworks, reported in 23 studies, demonstrated strong capabilities in learning complex patterns, especially in sequential and time-series fraud detection tasks. Study [37] highlighted the effectiveness of CNN and ANN, with CNN achieving high accuracy and F1-score. Likewise, [57] introduced a BiLSTM model that captured temporal dynamics more effectively than traditional models, although challenges like class imbalance and data quality issues persisted. More recent innovations include [42]**,** who employed a CCNN framework that surpassed conventional ML models but still required validation under real-world constraints. In addition, hybrid strategies integrating DL and dimensionality reduction were also identified. For example, [55] combined DL with t-SNE, PCA, and SVD to achieve superior results in credit card-not-present (CNP) fraud detection.

The percentage distribution in Figure 5 revealed several important insights. The dominance of supervised learning models underscores their continued reliability and ease of adoption in fraud detection; yet, their limitations become evident when handling high-dimensional or severely imbalanced datasets. RF and boosting algorithms, such as XGBoost, appear consistently strong across studies; however, their performance is often contingent upon careful preprocessing or resampling strategies, which were not uniformly applied across the surveyed literature. This suggests that although SL remains foundational, its standalone application may not be sufficient for modern fraud patterns that evolve rapidly. Deep learning frameworks provide an advanced alternative, particularly for temporal and sequential fraud detection, where BiLSTM and CNN models demonstrate clear advantages over traditional ML. The analysis, however, also reveals that while DL models improve predictive power, they often encounter interpretability issues and require computationally expensive infrastructures. For

example, the superior performance of CCNNs in [42] illustrates the potential of biologically inspired architectures, but their limited validation across diverse datasets highlights the risks of overfitting to specific benchmarks.

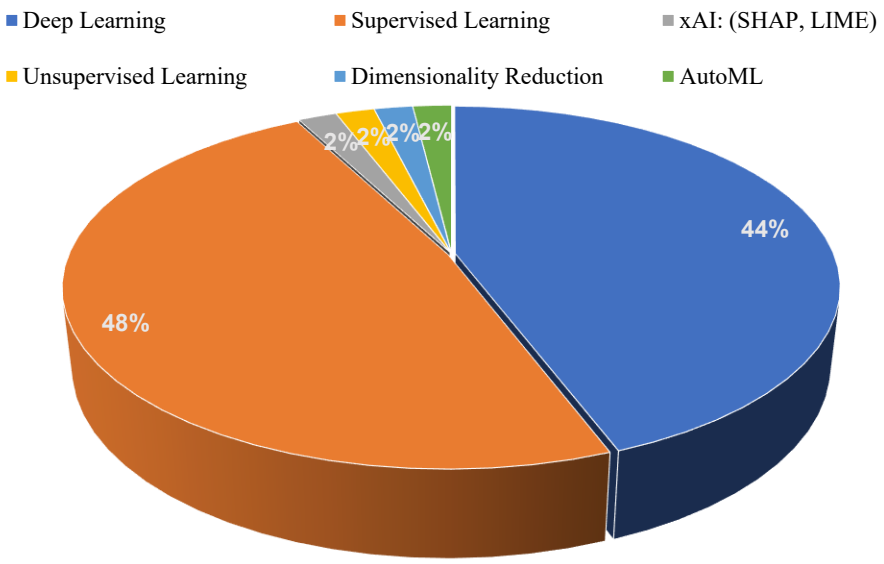

Figure 5. The percentage of adopted ML algorithms and frameworks in CCFD

Hybrid approaches, such as combining dimensionality reduction with DL as in [55], demonstrate that adaptability improves when multiple frameworks are integrated. Yet, the trade-off here lies in transparency and increased pipeline complexity, making real-world deployment more challenging. This observation is consistent with the distribution patterns in Figure 5, which, while reflecting methodological variety, also emphasize the relative neglect of unsupervised learning, AutoML, and xAI. These underexplored directions could provide more scalable and interpretable solutions, but their low adoption rates suggest that the field remains highly conservative in its methodology.

6.4. Determination of the Datasets Commonly used in CCFD Research, and How Accessibility, Applicability, and Adoptability they Are.

Across the reviewed studies as presented in Figure 6, a wide range of datasets has been utilized in training and evaluating CCFD models, each offering varying degrees of accessibility, applicability, and adaptability. Among these, the ECCT dataset of 2013 stands out as the most widely adopted, being used in 25 out of 40 studies. Its structured, tabular design with anonymized numerical features supported the application of traditional ML methods such as SVM and RF in [57], while its manageable size and clear labeling also enabled DL architectures including ANN and CNN in [37] as well as LSTM/AE in [65]. Despite this versatility, ECCT remains constrained by its limited scale, outdated feature space, and imbalance, reducing adaptability to modern fraud patterns. In contrast, the Kaggle CCFD dataset has been employed in studies such as [42] and [46], offering broader anonymized features that facilitate both ensemble ML models (e.g., RF, SVM, LR, AdaBoost) and DL variants, including CNN. While this dataset provides stronger grounds for benchmarking, it similarly reflects outdated transaction behaviors with limited representation of emerging fraud dynamics.

Beyond these benchmark datasets, real-world financial datasets have been used in [8] and [11], which enhanced model realism by aligning evaluation with heterogeneous transaction streams. These allowed deployment of sequence-aware DL approaches (RNN/DCNN) and graph-based methods (GCN), though their restricted accessibility undermines reproducibility. Decentralized institutional datasets under federated setups, such as those in [38], support privacy-preserving DL models but further limit their wider adoption. Synthetic datasets like BankSim, employed by [9], provided flexibility for simulating fraud behaviors and supported sequential DL models such as LSTM with attention. Yet, their artificial nature introduced patterns that may not generalize to real-world fraud.

Other localized datasets, such as the CCT Western USA [58], the Brazilian dataset [47], and the Taiwanese Client Credit Data [53], provided contextual profiling of fraud through supervised and hybrid ML/DL approaches, although transferability across regions remained limited. Emerging datasets such as Text2LMG [7], which enabled CNNs by transforming textual records into image-like representations, and Sparkov [56], which was tailored for GNNs to capture relational structures between merchants and customers, demonstrated strong potential but required extensive preprocessing. Meanwhile, several studies, including [30] and [33], relied on unspecified datasets, which hinders replicability and cross-study comparison. Furthermore, to ensure model generalizability, many studies for instance, [18], [59] utilized more than one datasets for the detection of CCF.

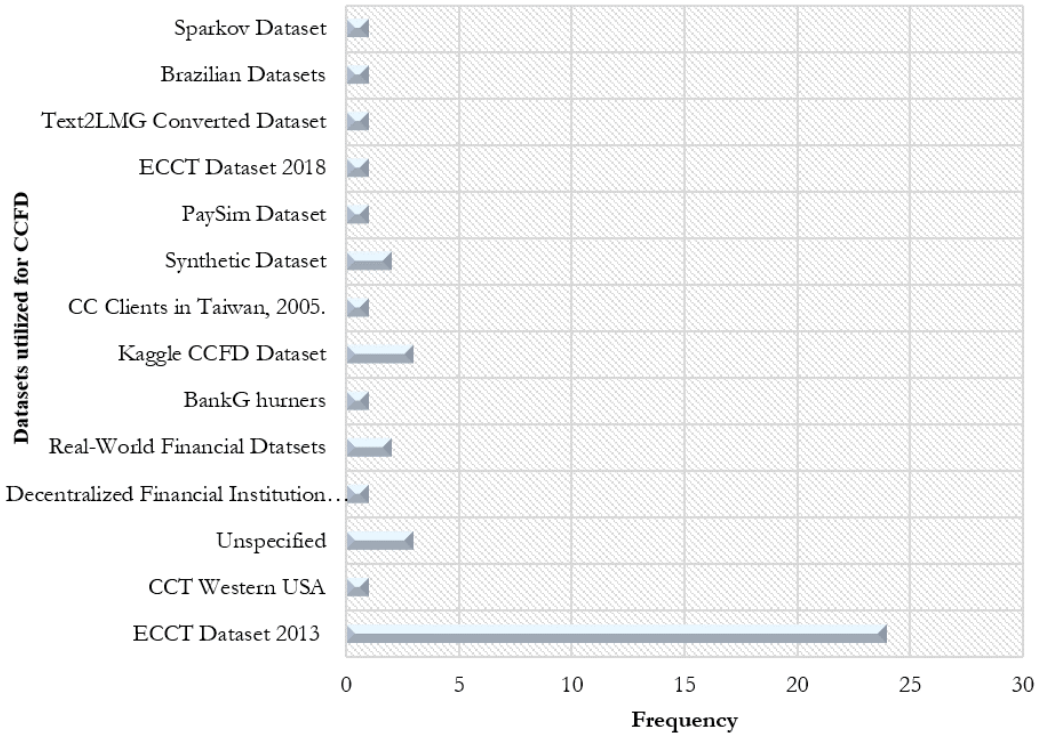

Figure 6. The frequency distribution of 43 datasets used for CCFD

The distribution in Figure 6 highlights a pronounced dependence on legacy benchmark datasets, particularly ECCT 2013, which was used in 56% of studies. This reliance provides a common ground for reproducibility, yet simultaneously creates a bottleneck in generalizability, as the dataset is not representative of contemporary fraud dynamics and exhibits severe imbalance. The Kaggle dataset, though offering additional anonymized features, is similarly constrained by outdated transaction patterns, meaning that much of the literature continues to validate models on environments that no longer reflect current fraud behavior. In contrast, real-world datasets, while offering higher external validity as seen in [8] and [11], remain underrepresented due to access restrictions. This imbalance indicates a tension between accessibility and applicability: easily available datasets drive methodological innovation but fail to ensure adaptability in real deployments.

Evidently, Synthetic datasets such as BankSim have partially addressed the adaptability gap by simulating flexible fraud scenarios; however, the artificial nature of the generated data limits the external realism of the results. Emerging resources like Sparkov and Text2LMG signal promising directions for next-generation fraud detection, as they provide richer sequential and relational features necessary for DL and GNN frameworks. However, their underutilization in the reviewed studies reflects both technical barriers and insufficient documentation, leaving much of their potential untapped. Furthermore, localized datasets, such as those from Brazil and Taiwan, demonstrate the value of context-aware fraud detection; however, their limited cross-regional applicability constrains scalability.

The synthesis suggests that dataset selection in CCFD research is heavily influenced by accessibility rather than representativeness, leading to a strong bias toward ECCT and Kaggle benchmarks. Consequently, models evaluated predominantly on these datasets risk overfitting to static, outdated fraud patterns and may underperform in dynamic, real-world scenarios. To overcome this, broader adoption of real-world, federated, and synthetic–hybrid datasets is essential, as these offer richer diversity while supporting privacy and security. Future methodological advances must therefore be coupled with shifts in dataset strategies to ensure robustness, transferability, and practical relevance for deployment.

6.5. Evaluation of the Optimisation Techniques Integrated into CCFD Models, and their Contribution to Improving Model Performance

The reviewed studies highlighted a broad spectrum of optimization techniques integrated into CCFD models to address persistent challenges, including class imbalance, overfitting, and feature redundancy. As presented in Figure 7, the most frequently employed approach is the SMOTE, which accounted for 17% of the sampled studies. For instance, [38], [40], and [64] applied SMOTE to improve recall and sensitivity toward minority class fraud cases. Complementary applications included hybrid strategies, such as SMOTE with ADASYN and cost-sensitive learning [44], as well as SMOTE with UMAP-based feature selection [9]. Undersampling was another commonly applied technique, as recorded in [36], [48], [58], often in combination with ensemble or hybrid approaches. More advanced adaptations included undersampling with dimensionality reduction, as in [55], which preserved critical feature structures. Recent contributions also introduced novel methods such as Energy Valley Optimization (EVO) in an ensemble framework [30], adversarial training for robustness against noisy environments [11], and combinatorial optimization pipelines that integrate resampling, dropout, and regularization [42].

Other optimization directions involved hyperparameter tuning, as implemented in [63] and [60], with AutoML-based optimization standing out for achieving strong performance without requiring handcrafted adjustments. Dimensionality reduction and feature selection techniques such as PCA and classifier-based sampling were incorporated in [36], [53], [60]. Optimizers such as Adam, RMSprop, and Adagrad [8], as well as batch normalization and graph converters [56], further enhanced model convergence in deep architectures. However, approximately 13% of the studies, including [31], [37], did not integrate any optimization strategies, potentially limiting reproducibility and scalability.

From the mapping summary in Figure 7, it becomes clear that optimization plays a pivotal role in shaping the performance trajectory of CCFD models, yet its application remains uneven and fragmented. The dominance of SMOTE across studies underscores the research community's continued reliance on data-level interventions to address imbalance. While its popularity reflects accessibility and ease of integration, the synthetic generation of data has repeatedly been critiqued for introducing artificial noise, which risks overfitting in sparse fraud contexts. This suggests that its continued prevalence may reflect inertia rather than genuine superiority. Under-sampling, although effective for computational efficiency, consistently highlights the trade-off between model simplicity and information loss. When paired with dimensionality reduction, however, it illustrates the potential of hybrid optimizations to overcome individual shortcomings. Emerging approaches, such as EVO and adversarial training, demonstrate a gradual shift toward more sophisticated optimization, signaling the field's responsiveness to real-world complexities, including noisy environments and dynamic fraud behaviors.

Equally significant is the growing uptake of hyperparameter tuning and AutoML frameworks, which reduce reliance on manual configuration while delivering competitive results. This reflects a methodological evolution toward automation, though it also introduces dependence on opaque optimization pipelines that may hinder interpretability. Similarly, the adoption of PCA, feature selection, and advanced optimizers indicates that optimization is no longer confined to handling class imbalance but is increasingly aligned with enhancing generalizability and convergence across large-scale datasets. However, a critical limitation lies in the 13% of studies that omitted optimization altogether, raising questions about the validity of their reported performances. Without optimization, models risk inflating baseline accuracy while failing to perform well under skewed distributions common in fraud detection. Collectively, these findings highlight that while traditional resampling methods remain dominant, the future of optimization in CCFD research is likely to rely on composite strategies that

combine data-level, model-level, and algorithm-level techniques to balance robustness, interpretability, and scalability.

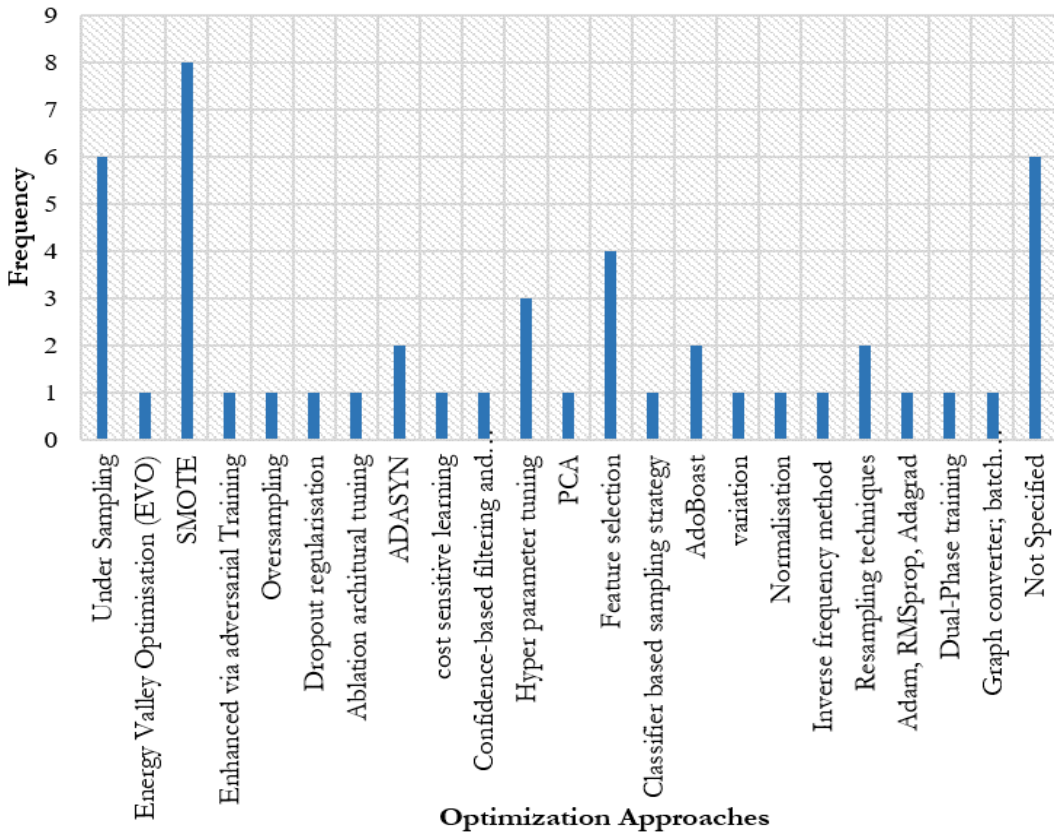

Figure 7. The frequency of the optimization approaches used in CCFD

The mapping of optimization approaches as presented in Tables 4, 5, 6,7 and 8 revealed that, will many studies like [40]**,** [45] employed more than one optimization approaches accounted for the frequency of 42, about six (6) studies [50], [51] either did not specified or used any optimization approach.

6.6. Assessment of the Evaluation Metrics Employed in CCFD Research, and their Relevance in Fraud Detection Scenarios

Evaluating the effectiveness of CCFD models requires performance metrics that not only reflect predictive accuracy but also capture the unique challenges posed by fraud detection, particularly class imbalance and the disproportionate costs of misclassifications. As illustrated in Figure 8, a total of 19 evaluation metrics were identified across the reviewed studies, which are often applied in combination to enhance the reliability of model validation.

Accuracy emerged as the most frequently reported metric, appearing in 30 studies such as [8], [37], [38], [65]. However, reliance on accuracy is problematic in skewed datasets, since predicting all transactions as legitimate may yield accuracy levels above 99% even when fraud remains undetected. To counter this, precision, recall, and F1-score were widely integrated. Precision was examined in 25 studies, including [42], which focused on the accuracy of fraud predictions and the minimization of false alarms. Recall was employed in 20 studies, emphasizing sensitivity to fraud detection where undetected cases incur high costs, while the F1-score was featured in 24 works, such as [7], [11], [44], providing a balance between precision and recall. Additionally, ROC-AUC and related AUC measures were reported in 14 studies, including [32], [58], [61], providing threshold-independent perspectives on classification performance. Precision-Recall curves (AUPRC), used by [34], addressed imbalances more effectively by focusing on fraud-relevant trade-offs, while individual measures such as True Positive Rate (TPR) and False Negative Rate (FNR), as in [53], highlighted sensitivity to missed

detections. Specificity, as applied in [18] and [47], complemented recall by ensuring that legitimate transactions were correctly identified. More advanced metrics, including the Jaccard Index and Matthews Correlation Coefficient (MCC), were utilized in [52], both of which are effective in imbalanced contexts, with MCC capturing the full confusion matrix.

Additional performance measures extended beyond classification accuracy. The BEFS metric [40] focused on balanced evaluation, while runtime efficiency and model reliability were gauged through training time, loss rate, and error measures (RMSE, RRMSE, MBE) in works such [8], [50], [55]. The confusion matrix itself, also reported in [50], continued to serve as a diagnostic tool for visualizing classification outcomes.

The mapping of metrics in Figure 8 reveals a field that is highly diverse yet uneven in its evaluation practices. The heavy reliance on accuracy, despite its shortcomings under class imbalance, suggests that some studies still prioritize simplicity over robustness, even though its inadequacy in fraud detection is well-documented. The frequent use of precision, recall, and F1-score indicates a recognition of fraud-specific trade-offs, with recall emerging as the most critical metric, as missed fraud transactions result in direct financial losses. Precision, on the other hand, is equally important for operational efficiency, as it reduces false alerts that overwhelm analysts. The F1-score provides a more balanced benchmark.

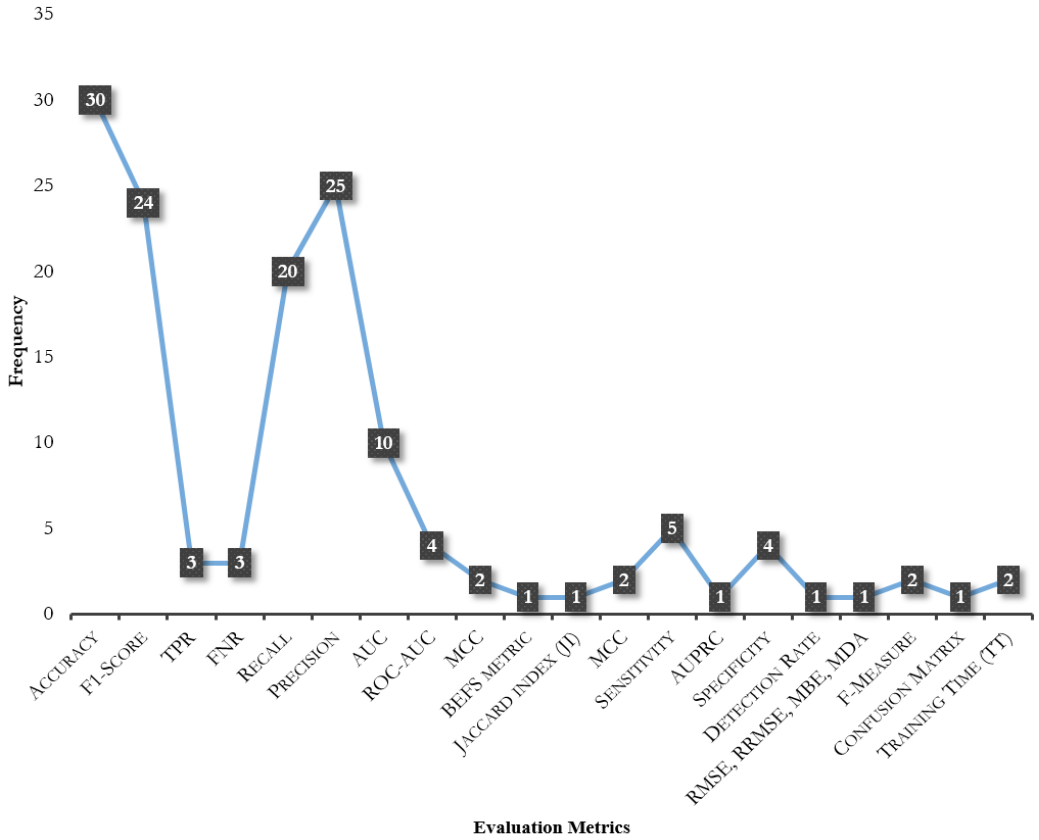

Figure 8. The frequency distribution of the performance metrics used in CCFD with ML

The introduction of ROC-AUC and AUPRC reflects a methodological shift toward threshold-independent metrics, yet the preference for ROC in many studies raises concern, given its tendency to present overly optimistic results in skewed datasets. AUPRC offers a more faithful reflection of fraud detection capability in imbalanced data, but its adoption remains limited compared to ROC-AUC. Similarly, the selective use of TPR, FNR, and specificity highlights an increasing awareness of cost-sensitive detection; however, the lack of consistency in applying these metrics reduces comparability across studies. Advanced measures, such as MCC and Jaccard Index, are especially suited to addressing imbalance; yet, their sparse adoption suggests that researchers still lean heavily on conventional metrics. This reliance could undermine the evaluation of models in realistic fraud detection environments where extreme disparities between legitimate and fraudulent transactions prevail. Furthermore, the

inclusion of metrics such as training time, error rates, and BEFS highlights an emerging focus on computational efficiency and scalability, which are increasingly relevant for real-world deployment but remain underexplored relative to predictive performance.

Overall, Figure 9 shows that while current research demonstrates awareness of the limitations of single metrics, such as accuracy, the evaluation landscape is fragmented, with some studies adopting holistic, multi-metric frameworks and others relying on minimal benchmarks. The progression toward comprehensive evaluations combining recall, precision, F1, ROC-AUC, and MCC suggests a recognition that no single metric adequately reflects fraud detection performance, especially in highly imbalanced and cost-sensitive scenarios. Furthermore, as revealed in figure 8, the 19 performance evaluation matrices appeared 142 times (frequency) across the 40 articles surveyed.

6.7 Acertainment of the effectiveness and limitations of existing CCFD models, and the potential prospects and directions for future research

The evolution of CCFD models reflects a complex interplay between increasing fraud sophistication and the growing capabilities of data-driven methods. While significant advances have been made, especially with the adoption of DL, ensemble techniques, and Hybrid models, a critical assessment of current approaches reveals that effectiveness is often context-dependent, and limitations persist across various fronts.

6.7.1 Effectiveness of Existing Models

Several studies have demonstrated that ensemble approaches, e.g., [30], [47]**,** and DL models, e.g., [42], [66] consistently outperform traditional methods in capturing nonlinear fraud patterns and learning from large-scale transactional data. Hybrid models that integrate BiLSTM with CNNs or attention mechanisms e.g., [7]) also recorded superior performance in both static and sequential contexts due to their ability to learn temporal dependencies in transaction sequences. Moreover, the application of GANs, e.g., [40] and brain-inspired systems [42], provided novel avenues for generating synthetic fraud patterns and mimicking human cognitive strategies in detection. These models were often validated using multiple performance metrics, such as F1-score, AUC, recall, and specificity, which demonstrate reliable fraud detection under simulated environments. Optimization strategies, such as SMOTE, ADASYN, under-sampling, and hyperparameter tuning, further boosted predictive accuracy and generalization, e.g., [44], [53].

6.7.2 Limitations and Challenges Faced by the Existing Models

Despite the effectiveness of some models employed, major limitations were evident across the reviewed studies.

(a) Limited Generalizability: A primary concern is the limited generalizability of models trained on static, publicly available datasets. Notably, the ECCT dataset from 2013 was used by 24 out of the 40 selected studies. While it offers accessibility, its outdated nature, anonymization, and specific feature structure limit the applicability of models to modern fraud contexts and real-world deployment. Only a handful of studies, e.g., [8], [38], have explored decentralized or real-time datasets, indicating a gap in adaptive learning from live data streams.

(b) Class Imbalance and Dataset Issues: This remains a prominent issue as fraud instances constitute a minuscule portion of real-world data. Although many researchers have applied various resampling techniques, such as SMOTE, under sampling, and hybrid approaches (e.g., ENN, AIIKNN), many models still struggle with performance degradation due to skewed data distributions [32], [59], [65]. Additionally, the use of outdated datasets (from 2013 and 2018) presented a limitation to the generalizability of the models, as observed by [53], [55], raising concerns about the relevance of these data in detecting modern-day fraud, particularly with evolving fraud patterns. Additionally, many studies have prioritized accuracy over more context-sensitive metrics, such as recall or precision, which risks inflating effectiveness in imbalanced environments [37].

(c) Model Complexity and Overfitting: Many of the studies reviewed, especially those incorporated in DL like LSTM, CNN, and GANs as presented by [8], [64], face the challenges of model overfitting and high computational overhead. Although those models demonstrated a potential for good accuracy, a lack of interpretability and the need for large computational

resources still exist, thereby affecting their scope for adoption in real-world scenarios. In addition, certain models, such as RF and XGBoost, while being effective in terms of accuracy [32], [59] still face issues of overfitting when dealing with imbalanced datasets.

(d) Performance Metrics and Trade-Offs: The findings of the review revealed accuracy as the predominant key performance metric; however, it may be misleading due to class imbalance. Some models, especially those evaluated by [46], showed high accuracy but failed to perform well in terms of recall and precision for fraud detection tasks, where false positives and false negatives can have significant financial implications. The study revealed that Models that emphasize one metric (e.g., accuracy or F1-score) sometimes compromise other evaluation metrics, such as training time, model complexity, and resource usage [6] and [50]. Those challenges suggest the need for better model selection strategies that consider a balanced trade-off among all relevant metrics.

(e) Explainability and Interpretability: While [11] integrated xAI techniques like SHAP and LIME, the model functions as a black-box system, making it less suitable for financial institutions requiring transparency for compliance and trust-building. Interpretability is important where understanding the rationale behind a decision is essential for trust and regulatory compliance. Some Models, like LSTM and CNN, offer high performance but often lack explainability. Although some studies, such as [7], [41]**,** have demonstrated the superiority of DL models in terms of fraud detection accuracy, these models remain black boxes.

(f) Synthetic Data Generation and Domain Adaptation: The use of synthetic data especially through GANs, has shown promise in tackling class imbalance [18]**;** however, the challenge of adapting synthetic data may not fully represent the complexity and evolving nature of real-world fraud patterns. Therefore, there is a need for robust synthetic data generation methods that better capture fraud characteristics across diverse domains (e.g., online and in-store transactions), as evident in [36], [51].

(g) Hybrid Models and Multi-Technique Approaches: Hybrid models, a combination of Traditional ML and DL approaches, as adopted by [36], [54], and others, have shown promise in addressing limitations of individual methods; however, models combination often leads to an increase in model complexity and interpretability challenges. Thus, there is room for further research in designing simple, interpretable hybrid models that can maintain high performance without excessive complexity.

A structured recommendation matrix is presented in Table 9, highlighting the suitability of each framework with respect to data characteristics, deployment needs, and interpretability.

6.8. Future Directions and Prospects

In the future in this domain, researchers should adopt a more ecosystem-oriented approach that combines advanced ML/DL models with stream-based architectures, federated frameworks, and explainable AI to better handle dynamic fraud patterns, as well as the adaptation of a self-supervised learning approach that does not introduce noise. Specifically, further research should prioritize:

- Real-time and incremental Learning: These are models that continuously adopt to emerging fraud behaviors as proposed by [41];
- Cross-institutional and FL Environment: This secure training on distributed datasets without centralizing sensitive information [38];
- Multi-Objective Optimization Techniques: This will balance accuracy, recall, runtime efficiency, and robustness under adversarial conditions [56];
- Benchmarking with diverse datasets: Adoption of the most current dataset beyond the ECCT dataset 2013 to improve applicability across varying financial landscapes.

In essence, while existing models for CCFD have shown substantial promise in terms of accuracy and robustness under constrained scenarios, their limitations in real-world deployment, transparency, and adaptability underscore the urgent need for a paradigm shift. Future research must prioritize hybrid and explainable approaches, integration with real-time infrastructure, and the development of standardized benchmarks to advance the field beyond experimental settings and into trustworthy, scalable financial systems.

Table 9. Recommendation metrics of frameworks/approaches for CCFD

Framework / Approach	Structured Data	High-Dimensional Data	Imbalanced Data	Sequential / Temporal Patterns	Real-Time Deployment	Interpretability	Computational Cost	Recommendation
Traditional ML	Excellent	Moderate	Needs Balancing	Poor	Limited	High	Low	Best for structured datasets, fast training, and interpretable models. Use with resampling if imbalance exists.
Deep Learning (DL)	Moderate	Excellent	Good with Augmentation	Excellent	Challenging	Low	High	Best for complex patterns, sequential data, and high-dimensional data. Requires GPU/strong computational resources.
ML and DL Comparison	ML: Good; DL: Moderate	DL: Excellent	Varies	DL: Good	Varies	Mixed	Varies	Use comparisons to select the best model for the dataset type and application context. Highlights trade-offs.
Data Augmentation / Balancing	Moderate	Good	Excellent	Moderate	Not real-time	Low	High	Enhances ML/DL performance, reduces class imbalance, and handles high-dimensional data. Adds complexity.
Deployment / Real-Time / AutoML / xAI	Excellent	Moderate	Good with Augmentation	Moderate	Excellent	High (xAI)	High	Best for automated model selection, real-time monitoring, and interpretable results. Consider computational and scalability constraints.

7. Conclusions

This survey synthesized recent advances in CCFD by systematically examining the challenges, methodologies, datasets, optimization strategies, evaluation practices, and the effectiveness of existing models. The findings reveal that class imbalance, feature redundancy, and limited access to real-world datasets remain significant obstacles, underscoring the importance of frameworks in ensuring robustness, scalability, and interpretability. In addressing these challenges, the study mapped out how methodological choices, ranging from supervised learning and ensemble models to deep learning and federated learning, offer different trade-offs between predictive accuracy, interpretability, and adaptability.

The review also revealed that datasets such as ECCT 2013 and Kaggle remain dominant in CCFD research, providing benchmarks but restricting generalizability to real-world fraud dynamics. Optimization strategies, particularly SMOTE, undersampling, and hyperparameter tuning, have proven essential for enhancing model sensitivity and reducing false negatives meanwhile, emerging approaches like adversarial training and AutoML point to more adaptive solutions. Evaluation practices demonstrated both progress and gaps: although accuracy is still the most frequently reported metric, recall, F1-score, ROC-AUC, and MCC better capture the imbalanced and cost-sensitive nature of fraud detection. Collectively, these results confirm that while existing models demonstrate considerable progress, especially with ensemble and hybrid frameworks, effectiveness remains firmly tied to dataset representativeness, optimization design, and evaluation rigor.

Therefore, future surveys should go beyond cataloguing models to critically examine the interplay between datasets, optimization techniques, and evaluation metrics in shaping detection outcomes. Greater emphasis is needed on systematically reviewing studies that employ federated, synthetic–hybrid, and domain-specific datasets, as these offer more realistic grounds for assessing scalability and generalizability. Future surveys should also advocate for standardized evaluation protocols that move beyond accuracy toward fraud-relevant measures such as recall, AUPRC, and MCC. By adopting these directions, subsequent reviews

can provide deeper, practice-oriented insights, ensuring that the next generation of survey work contributes not only to academic benchmarking but also to building trustworthy, deployable fraud detection systems.

Author Contributions: T. A. G., H. U. A., and T. M.; Literature Search: T. A. G. and H. U. A.; Manuscript Writing: T. A. G.; editing: H. U. A. and T. M.; Supervision: H. U. A. All authors have read and agreed to the published version of the manuscript.

Funding: This work received no funding support.

Conflicts of Interest: The authors have no conflict of interests related to this publication.

References

[1] R. Rajan and S. Rajest, "Revolutionizing Credit Card Fraud Detection : Harnessing Machine Learning Revolutionizing Credit Card Fraud Detection : Harnessing Machine Learning and Data Science for Enhanced Security," no. October, 2024.

[2] R. Bin Sulaiman, V. Schetinin, and P. Sant, "Review of Machine Learning Approach on Credit Card Fraud Detection," *Human-Centric Intell. Syst.*, vol. 2, no. 1–2, pp. 55–68, 2022, doi: 10.1007/s44230-022-00004-0.

[3] I. Mekterović, M. Karan, D. Pintar, and L. Brkić, "Credit card fraud detection in card-not-present transactions: Where to invest?," *Appl. Sci.*, vol. 11, no. 15, 2021, doi: 10.3390/app11156766.

[4] L. Bonde and A. K. Bichanga, "Improving Credit Card Fraud Detection with Ensemble Deep Learning-Based Models: A Hybrid Approach Using SMOTE-ENN," *J. Comput. Theor. Appl.*, vol. 2, no. 3, pp. 383–394, Feb. 2025, doi: 10.62411/jcta.12021.

[5] F. O. Aghware *et al.*, "Enhancing the Random Forest Model via Synthetic Minority Oversampling Technique for Credit-Card Fraud Detection," *J. Comput. Theor. Appl.*, vol. 1, no. 4, pp. 407–420, Mar. 2024, doi: 10.62411/jcta.10323.

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

[7] A. Alharbi *et al.*, "A Novel text2IMG Mechanism of Credit Card Fraud Detection: A Deep Learning Approach," *Electron.*, vol. 11, no. 5, pp. 1–18, 2022, doi: 10.3390/electronics11050756.

[8] J. I.-Z. and Chen and K.-L. Lai, "Deep Convolution Neural Network Model for Credit-Card Fraud Detection and Alert," *J. Artif. Intell. Capsul. Networks*, vol. 3, no. 2, pp. 101–112, 2021, doi: 10.36548/jaicn.2021.2.003.

[9] I. Benchaji, S. Douzi, B. El Ouahidi, and J. Jaafari, "Enhanced credit card fraud detection based on attention mechanism and LSTM deep model," *J. Big Data*, vol. 8, no. 1, 2021, doi: 10.1186/s40537-021-00541-8.

[10] M. Schmitt, "Intelligent Systems with Applications Automated machine learning : AI-driven decision making in business analytics," *Intell. Syst. with Appl.*, vol. 18, no. January, p. 200188, 2023, doi: 10.1016/j.iswa.2023.200188.

[11] O. R. Polu, "AI-Based Fake Transaction Detection in Credit Card Payments," vol. 12, no. 12, pp. 2205–2210, 2023.

[12] K. Shing Lim, L. Hong Lee, and Y.-W. Sim, "A Review of Machine Learning Algorithms for Fraud Detection in Credit Card Transaction," *Int. J. Comput. Sci. Netw. Secur.*, vol. 21, no. 9, pp. 31–40, 2021, [Online]. Available: https://doi.org/10.22937/IJCSNS.2021.21.9.4

[13] F. Aslam, "Advancing Credit Card Fraud Detection: A Review of Machine Learning Algorithms and the Power of Light Gradient Boosting," *Am. J. Comput. Sci. Technol.*, no. February, 2024, doi: 10.11648/ajcst.20240701.12.

[14] Y. Xiao, L. Tan, and J. Liu, "Application of Machine Learning Model in Fraud Identification : A Comparative Study of CatBoost , XGBoost and LightGBM Application of Machine Learning Model in Fraud Identification : A Comparative Study of CatBoost , XGBoost and LightGBM," pp. 0–7, 2025, doi: 10.20944/preprints202503.1199.v1.

[15] S. S. Sohal, "A Review of Credit Card Fraud Detection Techniques," *Lect. Notes Electr. Eng.*, vol. 832, no. 8, pp. 485–496, 2022, doi: 10.1007/978-981-16-8248-3_40.

[16] M. M. H. Sizan *et al.*, "Advanced Machine Learning Approaches for Credit Card Fraud Detection in the USA: A Comprehensive Analysis," *J. Ecohumanism*, vol. 4, no. 2, pp. 883–905, 2025, doi: 10.62754/joe.v4i2.6377.

[17] I. Y. Hafez, A. Y. Hafez, A. Saleh, A. A. Abd El-Mageed, and A. A. Abohany, "A systematic review of AI-enhanced techniques in credit card fraud detection," *J. Big Data*, vol. 12, no. 1, 2025, doi: 10.1186/s40537-024-01048-8.

[18] I. D. Mienye and N. Jere, "Deep Learning for Credit Card Fraud Detection: A Review of Algorithms, Challenges, and Solutions," *IEEE Access*, vol. 12, pp. 96893–96910, 2024, doi: 10.1109/ACCESS.2024.3426955.

[19] S. Sruthi, S. Emadaboina, and C. Jyotsna, "Enhancing Credit Card Fraud Detection with Light Gradient-Boosting Machine: An Advanced Machine Learning Approach," in *2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS)*, IEEE, Apr. 2024, pp. 1–6. doi: 10.1109/ICKECS61492.2024.10616809.

[20] K. Patel, "Credit Card Analytics: A Review of Fraud Detection and Risk Assessment Techniques," *Int. J. Comput. Trends Technol.*, vol. 71, no. 10, pp. 69–79, 2023, doi: 10.14445/22312803/ijctt-v71i10p109.

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

[22] A. K. Aguilar, "A Comparative Analysis of Credit Card Validation : Luhn Algorithm vs . A Comparative Analysis of Credit Card Validation : Luhn Algorithm vs . Deterministic Finite Automaton-Based Approach," no. May 2024, 2025, doi: 10.13140/RG.2.2.25392.47365.

[23] D. Ghobadi, F.; Kang, "Application of machine learning in dementia diagnosis: A systematic literature review," *Multidiscip. Digit. Publ. Inst.*, vol. 15, no. 4, p. 620, 2023, doi: 10.3390/w15040620 Academic.

[24] A. Trisal and D. Mandloi, "Machine Learning: an Overview," *Int. J. Res. -GRANTHAALAYAH*, vol. 9, no. 7, pp. 343–348, 2021, doi: 10.29121/granthaalayah.v9.i7.2021.4120.

[25] E. Christou, A. Parmaxi, and P. Zaphiris, "A systematic exploration of scoping and mapping literature reviews," *Univers. Access Inf. Soc.*, vol. 24, no. 1, pp. 941–951, 2025, doi: 10.1007/s10209-024-01120-3.

[26] F. Campbell *et al.*, "Mapping reviews, scoping reviews, and evidence and gap maps (EGMs): the same but different— the 'Big Picture' review family," *Syst. Rev.*, vol. 12, no. 1, pp. 1–8, 2023, doi: 10.1186/s13643-023-02178-5.

[27] S. Bagga, A. Goyal, N. Gupta, and A. Goyal, "ScienceDirect Credit Card Fraud Detection ICITETM2020 using Pipeling and Ensemble Learning Credit Card Fraud Detection using Ensemble a Pipeling and Goyal c Learning," *Procedia Comput. Sci.*, vol. 173, no. 2019, pp. 104–112, 2020, doi: 10.1016/j.procs.2020.06.014.

[28] M. Thelwall and K. Kousha, "ResearchGate: Disseminating, communicating, and measuring Scholarship?," *J. Assoc. Inf. Sci. Technol.*, vol. 66, no. 5, pp. 876–889, 2015, doi: 10.1002/asi.23236.

[29] G. Halevi, H. Moed, and J. Bar-Ilan, "Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation—Review of the Literature," *J. Informetr.*, vol. 11, no. 3, pp. 823–834, 2017, doi: 10.1016/j.joi.2017.06.005.

[30] A. S. I. Al-Dulaimi, I. R. Abdelmaksoud, S. Abdelrazek, and H. M. El-Bakry, "An intelligent credit card fraud detection model using data mining and ensemble learning," *Edelweiss Appl. Sci. Technol.*, vol. 9, no. 2, pp. 1391–1405, 2025, doi: 10.55214/25768484.v9i2.4651.

[31] E. Tank and M. Das, "On Credit Card Fraud Detection Using Machine Learning Techniques," in *Lecture Notes in Networks and Systems*, vol. 966 LNNS, 2024, pp. 293–303. doi: 10.1007/978-981-97-2004-0_21.

[32] H. J. Kim and J. S. Rhee, "Navigating the Fraud Frontier: Machine Learning Solutions for Credit Card Security," *Teh. Vjesn.*, vol. 32, no. 2, pp. 730–738, 2025, doi: 10.17559/TV-20241013002057.

[33] C. Yosepu, D. S. Kiran, K. Rammohan, and G. G. Babu, "Using Adaboost and Majority Voting to Identify Credit Card Fraudulent Activity," *Int. J. Eng. Res. Sci. Technol.*, vol. 21, no. 1,2025, pp. 193–199, 2025, [Online]. Available: www.ijerst.com

[34] N. Damanik and C.-M. Liu, "Advanced Fraud Detection: Leveraging K-SMOTEENN and Stacking Ensemble to Tackle Data Imbalance and Extract Insights," *IEEE Access*, vol. 13, pp. 10356–10370, 2025, doi: 10.1109/ACCESS.2025.3528079.

[35] A. A. Al-Maari, M. Abdulnabi, Y. Nathan, A. Ali, U. Ali, and M. Khan, "Optimized Credit Card Fraud Detection Leveraging Ensemble Machine Learning Methods," *Eng. Technol. Appl. Sci. Res.*, vol. 15, no. 3, pp. 22287–22294, 2025, doi: 10.48084/etasr.10287.

[36] A. Mniai, M. Tarik, and K. Jebari, "A Novel Framework for Credit Card Fraud Detection," *IEEE Access*, vol. 11, no. September, pp. 112776–112786, 2023, doi: 10.1109/ACCESS.2023.3323842.

[37] S. Al Balawi and N. Aljohani, "Credit-card Fraud Detection System using Neural Networks," *Int. Arab J. Inf. Technol.*, vol. 20, no. 2, pp. 234–241, 2023, doi: 10.34028/iajit/20/2/10.

[38] W. Mohamedhen, M. Charfeddine, and Y. H. Kacem, "Enhanced Credit Card Fraud Detection Using Federated Learning, LSTM Models, and the SMOTE Technique," *Int. Conf. Agents Artif. Intell.*, vol. 3, no. Icaart, pp. 368–375, 2025, doi: 10.5220/0013135100003890.

[39] J. Wang, "Credit Card Fraud Detection via Hierarchical Multi-Source Data Fusion and Dropout Regularization," no. 1, 2025.

[40] M. Tayebi and S. El Kafhali, "Generative Modeling for Imbalanced Credit Card Fraud Transaction Detection," *J. Cybersecurity Priv.*, vol. 5, no. 1, pp. 1–36, 2025, doi: 10.3390/jcp5010009.

[41] K. Kandi and A. García-Dopico, "Enhancing Performance of Credit Card Model by Utilizing LSTM Networks and XGBoost Algorithms," *Mach. Learn. Knowl. Extr.*, vol. 7, no. 1, pp. 1–21, 2025, doi: 10.3390/make7010020.

[42] Y. Wu, L. Wang, H. Li, and J. Liu, "A Deep Learning Method of Credit Card Fraud Detection Based on Continuous-Coupled Neural Networks," *Mathematics*, vol. 13, no. 5, pp. 1–18, 2025, doi: 10.3390/math13050819.

[43] S. S. Sulaiman, I. Nadher, and S. M. Hameed, "Credit Card Fraud Detection Challenges and Solutions: A Review," *Iraqi J. Sci.*, vol. 65, no. 4, pp. 2287–2303, 2024, doi: 10.24996/ijs.2024.65.4.42.

[44] W. salah salem, I. el- hasnony, A. Abu Elfetouh, and A. Rezk, "Enhancing Fraud Detection in Imbalanced Datasets: A Comparative Study of Machine Learning and Deep Learning Algorithms with SMOTE Preprocessing," *Mansoura J. Comput. Inf. Sci.*, vol. 20, no. 1, pp. 1–21, Jun. 2025, doi: 10.21608/mjcis.2025.313097.1007.

[45] A. Nuthalapati, "Smart Fraud Detection Leveraging Machine Learning For Credit Card Security," *Educ. Adm. Theory Pract.*, vol. 29, no. 2, pp. 433–443, 2023, doi: 10.53555/kuey.v29i2.6907.

[46] N. G. Md Rokibul Hasan, Md Sumon Gazi, "Explainable AI in Credit Card Fraud Detection: Interpretable Models and Transparent Decision-making for Enhanced Trust and Compliance in the USA Md," *J. Comput. Sci. Technol. Stud.*, no. April, pp. 104–111, 2024, doi: 10.32996/jcsts.2024.6.2.1.

[47] E. Esenogho, I. D. Mienye, T. G. Swart, K. Aruleba, and G. Obaido, "A Neural Network Ensemble with Feature Engineering for Improved Credit Card Fraud Detection," *IEEE Access*, vol. 10, pp. 16400–16407, 2022, doi: 10.1109/ACCESS.2022.3148298.

[48] R. Asha and K. uresh Kumar, "Credit card fraud detection using artificial neural network," *Glob. Transitions Proc.*, vol. 2, no. 1, pp. 35–41, 2021, doi: 10.1016/j.gltp.2021.01.006.

[49] S. Khan, A. Alourani, B. Mishra, A. Ali, and M. Kamal, "Developing a Credit Card Fraud Detection Model using Machine Learning Approaches," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 3, pp. 411–418, 2022, doi: 10.14569/IJACSA.2022.0130350.

[50] A. Aslam and A. Hussain, "A Performance Analysis of Machine Learning Techniques for Credit Card Fraud Detection," *J. Artif. Intell.*, vol. 6, no. 1, pp. 1–21, 2024, doi: 10.32604/jai.2024.047226.

[51] Himanshu Sinha, "An examination of machine learning-based credit card fraud detection systems," *Int. J. Sci. Res. Arch.*, vol. 12, no. 2, pp. 2282–2284, 2024, doi: 10.30574/ijsra.2024.12.2.1456.

[52] M. A. Walauskis and T. M. Khoshgoftaar, "Unsupervised label generation for severely imbalanced fraud data," *J. Big Data*, vol. 12, no. 1, 2025, doi: 10.1186/s40537-025-01120-x.

[53] A. Hassan, A. Khader, J. Saudagar, S. Bhanja, and A. Das, "Data-Driven Methods for Credit Card Fraud Detection Using Machine Learning Data-Driven Methods for Credit Card Fraud Detection Using Machine Learning," no. March, 2025.
[54] N. S. Alfaiz and S. M. Fati, "Enhanced Credit Card Fraud Detection Model Using Machine Learning," *Electron.*, vol. 11, no. 4, 2022, doi: 10.3390/electronics11040662.
[55] A. Razaque *et al.*, "Credit Card-Not-Present Fraud Detection and Prevention Using Big Data Analytics Algorithms," *Appl. Sci.*, vol. 13, no. 1, 2023, doi: 10.3390/app13010057.
[56] A. Cherif, H. Ammar, M. Kalkatawi, S. Alshehri, and A. Imine, "Encoder–decoder graph neural network for credit card fraud detection," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 36, no. 3, p. 102003, 2024, doi: 10.1016/j.jksuci.2024.102003.
[57] Y.-F. Zhang, H.-L. Lu, H.-F. Lin, X.-C. Qiao, and H. Zheng, "The Optimized Anomaly Detection Models Based on an Approach of Dealing with Imbalanced Dataset for Credit Card Fraud Detection," *Mob. Inf. Syst.*, vol. 2022, pp. 1–10, Apr. 2022, doi: 10.1155/2022/8027903.
[58] J. K. Afriyie *et al.*, "A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions," *Decis. Anal. J.*, vol. 6, no. December 2022, p. 100163, 2023, doi: 10.1016/j.dajour.2023.100163.
[59] E. Ileberi, Y. Sun, and Z. Wang, "A machine learning based credit card fraud detection using the GA algorithm for feature selection," *J. Big Data*, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00573-8.
[60] E. Ileberi, Y. Sun, and Z. Wang, "Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost," *IEEE Access*, vol. 9, pp. 165286–165294, 2021, doi: 10.1109/ACCESS.2021.3134330.
[61] V. Plakandaras, P. Gogas, T. Papadimitriou, and I. Tsamardinos, "Credit Card Fraud Detection with Automated Machine Learning Systems," *Appl. Artif. Intell.*, vol. 36, no. 1, 2022, doi: 10.1080/08839514.2022.2086354.
[62] S. Nehe and P. Devale, "Ai Based Real-time Fraud Detection System for Credit Card Transaction Anomaly Identification," *Int. J. Sci. Technol.*, vol. 16, no. 3, pp. 1–16, 2025, doi: 10.71097/ijsat.v16.i3.7443.
[63] I. D. Mienye and Y. Sun, "A Machine Learning Method with Hybrid Feature Selection for Improved Credit Card Fraud Detection," *Appl. Sci.*, vol. 13, no. 12, 2023, doi: 10.3390/app13127254.
[64] N. Damanik and C. M. Liu, "Advanced Fraud Detection: Leveraging K-SMOTEENN and Stacking Ensemble to Tackle Data Imbalance and Extract Insights," *IEEE Access*, vol. 13, no. December 2024, pp. 10356–10370, 2025, doi: 10.1109/ACCESS.2025.3528079.
[65] S. S. Sulaiman, I. Nadher, and S. M. Hameed, "Credit Card Fraud Detection Using Improved Deep Learning Models," *Comput. Mater. Contin.*, vol. 78, no. 1, pp. 1049–1069, 2024, doi: 10.32604/cmc.2023.046051.
[66] Y. F. Zhang, H. L. Lu, H. F. Lin, X. C. Qiao, and H. Zheng, "The Optimized Anomaly Detection Models Based on an Approach of Dealing with Imbalanced Dataset for Credit Card Fraud Detection," *Mob. Inf. Syst.*, vol. 2022, 2022, doi: 10.1155/2022/8027903.