IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, February 2026
Impact Factor: 8.2

Credit Card Fraud Detection using Machine Learning Algorithms

Mamta Indrasing Girase[1], Jagruti S. Patil[2], Dr. Rahul M. Patil[3]
PG Student, Department of E&TC Engineering, Gangamai College of Engineering, Nagaon, Dhule, India[1]
Assistant Professor, Department of E&TC Engineering, Gangamai College of Engineering, Nagaon, Dhule, India[2]
Assistant Professor, Department of E&TC Engineering, Gangamai College of Engineering, Nagaon, Dhule, India[3]

Abstract: *The exponential growth of digital payment systems and electronic commerce has positioned credit card fraud as one of the most pressing challenges facing modern financial institutions. While traditional rule-based detection systems have demonstrated effectiveness in controlled environments, they exhibit fundamental limitations in adaptability when confronted with the dynamic and adversarial nature of contemporary fraud schemes. This comprehensive review examines the evolution and current state of machine learning approaches for credit card fraud detection, spanning from classical supervised learning algorithms to cutting-edge deep learning architectures and graph-based methodologies. This paper addresses critical challenges inherent in fraud detection, including extreme class imbalance (fraudulent transactions typically represent less than 1% of total volumes), concept drift caused by evolving fraud strategies, asymmetric misclassification costs, and real-time processing constraints. We systematically analyze supervised learning methods such as Logistic Regression, Decision Trees, Random Forests, Support Vector Machines, and Gradient Boosting Machines (XGBoost), alongside deep learning architectures including Artificial Neural Networks, Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks, and hybrid CNN-LSTM models. A comparative analysis of existing studies from 2010 to 2025 reveals that ensemble methods and deep learning models consistently outperform traditional classifiers, with graph neural networks and attention-based architectures achieving state-of-the-art performance.*

Despite significant advancements, challenges related to model interpretability, adaptive learning in non-stationary environments, computational efficiency for real-time deployment, and data privacy compliance remain unresolved. This review identifies emerging research directions—including semi-supervised learning, adaptive feature selection, and explainable AI—as promising avenues for developing robust, scalable, and transparent fraud detection systems suitable for real-world financial environments.

Keywords: Credit card fraud detection, machine learning, deep learning, class imbalance, ensemble methods, concept drift, graph neural networks, feature engineering, real-time detection, financial

I. INTRODUCTION

The rapid expansion of electronic commerce and digital payment infrastructures has fundamentally transformed the landscape of modern financial transactions. Credit cards have emerged as one of the most widely used payment instruments globally, valued for their convenience, transaction speed, and international interoperability. According to recent industry reports, global credit card transaction volumes exceeded $42 trillion in 2024, representing a substantial portion of worldwide commerce [1]. However, this widespread adoption has simultaneously increased vulnerability to fraudulent activities, establishing credit card fraud as a persistent and economically devastating challenge for financial institutions, merchants, and consumers alike. The Federal Trade Commission reported that consumers lost approximately $10 billion to fraud in 2023, with credit card fraud accounting for a significant proportion of these losses [2].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 2, February 2026 Impact Factor: 8.2

Fraudulent transactions impose substantial direct financial costs and indirect consequences, including erosion of consumer trust in digital payment ecosystems, increased operational expenses for fraud investigation and resolution, and potential regulatory penalties for inadequate fraud prevention measures [3]. These multifaceted impacts necessitate the development and deployment of robust, reliable, and adaptive fraud detection mechanisms capable of identifying fraudulent activities with high precision while minimizing disruption to legitimate transactions.

Conventional fraud detection systems have predominantly relied upon rule-based and expert-driven approaches that employ predefined thresholds, static heuristics, and manually crafted decision rules to identify suspicious transactions [4]. While such methodologies can demonstrate effectiveness within controlled environments characterized by stable fraud patterns, they exhibit fundamental limitations in adaptability and struggle to accommodate the dynamic and adversarial nature of modern fraud. Fraudsters continuously evolve their strategies, techniques, and tactics to evade detection systems, rendering static rules progressively ineffective and contributing to elevated false positive rates that burden both financial institutions and customers [5]. Furthermore, the unprecedented scale, velocity, and complexity of contemporary transaction data exacerbate the limitations of traditional rule-based approaches, which cannot efficiently process and analyze massive data streams in real-time.

Machine learning has emerged as a transformative alternative paradigm by enabling automated learning from historical transaction data and facilitating adaptive identification of complex, non-linear fraud patterns [6]. Machine learning-based fraud detection systems possess the capability to model intricate relationships among transaction attributes, capture subtle behavioral characteristics that distinguish fraudulent from legitimate activities, and generalize effectively to previously unseen fraud scenarios [7]. Over the past fifteen years, an extensive array of supervised, unsupervised, semi-supervised, ensemble, and deep learning algorithms have been proposed, evaluated, and deployed for credit card fraud detection applications [8].

Despite notable advances in algorithmic sophistication and detection performance, persistent challenges continue to constrain the practical deployment and long-term effectiveness of machine learning-based fraud detection systems. These challenges include extreme class imbalance where fraudulent transactions represent less than 1% of total volumes, concept drift caused by evolving fraud patterns and changing consumer behavior, stringent real-time processing requirements that demand sub-second latency, the need for model interpretability to satisfy regulatory compliance and build stakeholder trust, and data privacy constraints that limit access to comprehensive training datasets [9], [10].

Credit Card Fraud Detection: Problem Overview

Credit card fraud detection is formally characterized as a binary classification problem wherein each transaction must be categorized as either legitimate (class 0) or fraudulent (class 1) [11]. In fraud detection applications, the cost function is inherently asymmetric, as false negatives (failing to detect fraud) typically incur substantially higher costs than false positives (incorrectly flagging legitimate transactions), though the latter negatively impacts customer experience and operational efficiency [12].

One of the most significant challenges in credit card fraud detection is extreme class imbalance, where fraudulent transactions typically represent less than 1% of total transaction volumes, and in many real-world datasets, the imbalance ratio can be as severe as 1:1000 or even 1:10000 [13]. This severe imbalance causes standard machine learning classifiers to exhibit a strong bias toward the majority class, as minimizing overall classification error can be trivially achieved by predicting all transactions as legitimate [14, 15].

The asymmetric cost structure of fraud detection further complicates the classification problem [16]. False negatives, where fraudulent transactions are incorrectly classified as legitimate, result in direct financial losses that must be absorbed by financial institutions or cardholders. The average cost of an undetected fraudulent transaction can range from tens to thousands of dollars, depending on transaction characteristics and fraud type [17]. Conversely, false positives, where legitimate transactions are incorrectly flagged as fraudulent, impose indirect costs including customer dissatisfaction, potential loss of business due to declined transactions, operational expenses associated with fraud investigation and customer service, and potential damage to the institution's reputation [18]. Research indicates that

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, February 2026
Impact Factor: 8.2

excessive false positive rates can lead to customer attrition and reduced transaction volumes, particularly in e-commerce environments where declined legitimate transactions may drive customers to competing platforms [19].

Modern payment systems demand real-time or near-real-time fraud detection to authorize or decline transactions within milliseconds [20]. This stringent latency requirement constrains the computational complexity of detection algorithms and limits the feasibility of ensemble methods or deep learning architectures that require extensive computation. A typical credit card authorization must be completed within 2-3 seconds, leaving minimal time for fraud scoring and decision-making [21].

Balancing detection accuracy with computational efficiency represents a fundamental trade-off in system design. While complex deep learning models may achieve superior detection performance, their computational requirements may exceed real-time processing constraints, necessitating model compression, pruning, or deployment on specialized hardware accelerators [22].

Dataset and Data Preprocessing

Due to confidentiality constraints and privacy regulations, access to real-world credit card transaction data remains severely limited for academic research [23]. Consequently, most published studies rely on anonymized, synthetic, or publicly available datasets that may not fully represent the complexity and characteristics of production fraud detection environments.

The most widely used benchmark dataset in fraud detection research is the European cardholders dataset, collected during two days in September 2013 and made publicly available for research purposes [24]. This dataset contains 284,807 transactions, of which 492 (0.172%) are fraudulent, exhibiting the extreme class imbalance characteristic of real-world fraud data. The dataset comprises 30 features, including 28 principal components obtained through PCA transformation to protect cardholder privacy, along with transaction time and amount as non-transformed features [25].

Alternative datasets used in fraud detection research include the IEEE-CIS Fraud Detection dataset from a Kaggle competition, synthetic datasets generated using simulators, and proprietary datasets from financial institutions used in industry research but not publicly available [26]. The development of more comprehensive, realistic, and publicly accessible fraud detection datasets remains an important priority for advancing research in this domain.

Data preprocessing is a critical stage in fraud detection pipelines, directly impacting model performance and generalization capability. Key preprocessing steps include:

Missing Value Handling: Transaction data may contain missing values due to incomplete information capture, system errors, or privacy-preserving transformations. Common strategies include deletion of records with missing values (when the proportion is small), imputation using statistical measures (mean, median, mode), and model-based imputation using algorithms such as k-nearest neighbors or matrix factorization [27].

Feature Scaling and Normalization: Many machine learning algorithms, particularly distance-based methods such as k-NN and SVM, are sensitive to feature scales [28]. For tree-based ensemble methods such as Random Forest and XGBoost, feature scaling is generally not required due to their invariance to monotonic transformations [29].

Categorical Encoding: Categorical features such as merchant category codes, country codes, and transaction types must be transformed into numerical representations suitable for machine learning algorithms [30]. One-hot encoding creates binary indicator variables for each category, while ordinal encoding assigns integer values based on category order. More sophisticated techniques include target encoding (replacing categories with aggregated target statistics) and embedding representations learned through neural networks [31].

Temporal Feature Engineering: Temporal attributes such as transaction timestamp can be decomposed into multiple informative features including hour of day, day of week, day of month, and month of year, enabling models to capture temporal patterns in transaction behavior [32]. Cyclical encoding using sine and cosine transformations can preserve the circular nature of temporal features (e.g., hour 23 is close to hour 0) [33].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, February 2026
Impact Factor: 8.2

Machine Learning Algorithms for Credit Card Fraud Detection

Machine learning has waded into several branches, each of which deals with a different type of learning task. All machine learning models automate the process of inductive inference including phenomenon observation, building a model based on observed phenomenon and making predictions using a model. Following are the classical machine learning algorithms used for fraud detections.

Logistic Regression: Logistic Regression serves as a fundamental baseline classifier in fraud detection research due to its simplicity, computational efficiency, and interpretability. The interpretability of Logistic Regression, where feature weights directly indicate the contribution of each attribute to fraud probability, makes it attractive for applications requiring model transparency and regulatory compliance. However, Logistic Regression exhibits limitations when confronted with complex non-linear relationships and feature interactions that characterize fraud patterns [34].

Decision Trees construct hierarchical rule-based models by recursively partitioning the feature space based on attribute values that maximize information gain or minimize impurity. Each internal node represents a test on a feature, each branch corresponds to a test outcome, and each leaf node assigns a class label. The primary advantages of Decision Trees include their interpretability through visualization of decision paths, ability to model non-linear relationships and feature interactions, and invariance to feature scaling [35]. However, individual Decision Trees are prone to overfitting, particularly when grown to large depths, and exhibit high variance where small changes in training data can result in substantially different tree structures [36].

Random Forests address the limitations of individual Decision Trees through ensemble learning, constructing multiple trees trained on bootstrap samples of the training data and random subsets of features. The final prediction is obtained through majority voting (classification) or averaging (regression) across all trees in the forest. Random Forests have demonstrated strong performance in fraud detection applications, consistently ranking among the top-performing algorithms in comparative studies [37]. Their advantages include robustness to overfitting through ensemble averaging, ability to handle high-dimensional data, implicit feature selection through random feature sampling, and natural handling of missing values and mixed data types [38]. Research has shown that Random Forests achieve high precision and recall on fraud detection benchmarks, with recent studies reporting F1-scores exceeding 0.85 and AUC-ROC values above 0.98 on the European cardholders dataset [39]. The algorithm's computational efficiency and scalability make it suitable for real-time fraud detection in production systems.

Support Vector Machines (SVM) construct optimal hyperplanes that maximize the margin between classes in high-dimensional feature spaces. For non-linearly separable data, SVMs employ kernel functions (polynomial, radial basis function, sigmoid) to implicitly map inputs into higher-dimensional spaces where linear separation becomes possible. SVMs have demonstrated effectiveness in fraud detection, particularly when combined with appropriate kernel selection and parameter tuning [40]. However, SVMs exhibit computational complexity that scales poorly with large training datasets, making them less suitable for applications involving millions of transactions [41].

Gradient Boosting Machines (GBM): including popular implementations such as XGBoost, LightGBM, and CatBoost, construct ensemble models by sequentially training weak learners (typically shallow decision trees) that correct errors made by previous learners. Each new tree is fitted to the residual errors of the current ensemble, with predictions combined through weighted summation. XGBoost (Extreme Gradient Boosting) has emerged as one of the most effective algorithms for fraud detection, consistently achieving state-of-the-art performance in competitive benchmarks and real-world applications [41], [42]. Key innovations include regularization terms to prevent overfitting, efficient handling of sparse data, parallel tree construction, and built-in support for class imbalance through scale_pos_weight parameter [43].

k-Nearest Neighbors (k-NN) is an instance-based learning algorithm that classifies transactions based on the majority class among the k nearest training examples in feature space. Distance metrics such as Euclidean distance, Manhattan distance, or Mahalanobis distance determine proximity between instances. While k-NN offers intuitive interpretability and requires no explicit training phase, it exhibits significant limitations for fraud detection applications [44]. The algorithm's computational complexity during prediction scales linearly with training set size, making it impractical for large-scale transaction datasets. Additionally, k-NN is highly sensitive to class imbalance, feature scaling, and the curse of dimensionality, where distance metrics become less meaningful in high-dimensional spaces [45].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, February 2026
Impact Factor: 8.2

Deep Learning Techniques in Fraud Detection

Deep learning techniques have gained prominence due to their capacity to learn hierarchical feature representations and capture complex temporal relationships some of them are discussed as follows

Artificial Neural Networks (ANNs), comprising interconnected layers of neurons with non-linear activation functions, provide a flexible framework for learning complex non-linear mappings from transaction features to fraud probabilities [46]. Feed-forward neural networks with multiple hidden layers (deep neural networks) can learn hierarchical feature representations that capture intricate fraud patterns not easily modeled by classical algorithms. Research comparing ANNs with traditional machine learning algorithms has demonstrated that deep neural networks achieve superior detection performance, particularly when provided with sufficient training data [47]. Studies report that ANNs outperform Logistic Regression, Decision Trees, and k-NN on standard fraud detection benchmarks, with accuracy improvements of 2-5% and significant gains in fraud recall [48].

Recurrent Neural Networks (RNNs) are specifically designed to process sequential data by maintaining hidden states that capture temporal dependencies across transaction sequences [49]. Long Short-Term Memory (LSTM) networks address the vanishing gradient problem of standard RNNs through gating mechanisms that regulate information flow, enabling the learning of long-range temporal dependencies [50].

Long Short-Term Memory LSTMs are particularly well-suited for fraud detection applications where transaction sequences contain valuable behavioral information [51]. For example, patterns such as transaction velocity (frequency of transactions within a time window), spending trajectory (evolution of transaction amounts over time), and location sequences can indicate fraudulent account takeover or card testing behavior [52]. Empirical studies demonstrate that LSTM networks achieve superior performance compared to feedforward neural networks and classical machine learning algorithms when modeling temporal transaction sequences [53]. Research reports LSTM accuracies exceeding 96% and AUC-ROC values above 0.97 on fraud detection benchmarks, with particular improvements in detecting sophisticated fraud patterns that unfold over multiple transactions [54].

Convolutional Neural Networks (CNNs), originally developed for image processing, have been adapted to fraud detection through transformation of transaction data into structured representations amenable to convolution operations. Approaches include treating transaction sequences as 1D signals, constructing 2D representations from transaction attributes and temporal windows, and encoding transaction graphs as adjacency matrices [55]. CNNs employ convolutional filters to automatically learn local patterns and hierarchical features from structured transaction data [56]. Pooling layers reduce dimensionality while preserving salient features, and fully connected layers perform final classification based on learned representations.

Hybrid CNN-LSTM Architectures leverage the complementary strengths of spatial feature learning and temporal sequence modeling [57]. CNNs extract local patterns and high-level features from transaction attributes, while LSTMs model temporal dependencies and sequential relationships across transactions [58]. The typical architecture consists of convolutional layers for feature extraction, followed by LSTM layers for temporal modeling, and fully connected layers for final classification [59]. This hierarchical approach enables the learning of both spatial and temporal fraud patterns from transaction sequences.

Feature Engineering and Selection

Feature engineering plays a pivotal role in fraud detection, as raw transaction attributes rarely provide sufficient discriminatory power. Behavioral features derived from transaction histories—such as spending consistency, transaction velocity, and location variability—significantly enhance model effectiveness [71]. Feature selection methods aim to eliminate redundancy and noise, improving computational efficiency and generalization.

Filter, wrapper, and embedded techniques are widely used, with genetic algorithm-based selection and recursive feature elimination demonstrating notable improvements in recent studies [72]. Adaptive feature selection is increasingly recommended to maintain relevance in evolving fraud scenarios.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, February 2026
Impact Factor: 8.2

Evaluation Metrics and Performance Analysis

Given the extreme imbalance inherent in fraud datasets, accuracy alone is inadequate. Precision, recall, F1-score, and ROC-AUC are commonly used, with recall being particularly critical to minimize undetected fraud [60]. Precision–recall curves and AUPRC provide more meaningful insights in highly skewed datasets.

Robust evaluation protocols employ stratified k-fold cross-validation and, increasingly, temporal validation to reflect real-world deployment conditions. Transparent reporting of preprocessing and evaluation strategies is essential to ensure reproducibility and fair comparison across studies [73].

Comparative Analysis of Existing Studies

This section presents a comprehensive comparative analysis of recent research in credit card fraud detection, synthesizing methodologies, datasets, performance metrics, and key contributions from studies published between 2010 and 2025. The analysis reveals a clear evolution in fraud detection methodologies, progressing from classical machine learning models to ensemble, deep learning, and graph-based approaches algorithmic approaches, with ensemble methods and deep learning architectures demonstrating superior performance compared to classical machine learning algorithms.

The evolution of fraud detection research over the past five years demonstrates following several notable trends:

Dominance of Ensemble and Deep Learning Methods: Random Forests, XGBoost, and deep neural networks consistently outperform traditional algorithms such as Logistic Regression and Decision Trees across diverse datasets and evaluation metrics.

Increased Focus on Hybrid Architectures: Combining multiple algorithmic paradigms (e.g., CNN-LSTM, CNN-SVM) leverages complementary strengths and achieves performance improvements of 2-5% over individual models.

Emphasis on Class Imbalance Handling: Nearly all high-performing systems incorporate explicit techniques for addressing class imbalance, with SMOTE variants and ensemble-based approaches demonstrating particular effectiveness.

Growing Interest in Graph-Based Methods: Graph neural networks that model relationships among transactions, cardholders, and merchants represent an emerging paradigm achieving state-of-the-art results.

Attention to Explainability: Increasing recognition of the importance of model interpretability has led to research on explainable AI techniques, attention mechanisms, and rule extraction methods.

Comparative Summary of Credit Card Fraud Detection Studies (2010–2025)

Year	Authors	Methodology	Dataset	Key Contributions
2014	Gama *et al.*[60]	Concept drift adaptation	Streaming data	Formalized drift handling for evolving fraud
2015	Dal Pozzolo *et al.* [61]	Cost-sensitive ML	European dataset	Improved fraud recall under imbalance
2016	Chen & Guestrin [65]	XGBoost ensemble	Large-scale data	High scalability and ROC-AUC
2017	Dal Pozzolo *et al.* [71]	Adaptive ML	Credit card data	Emphasized adaptive learning
2018	Roy *et al.* [68]	Deep neural networks	Public datasets	ANN outperformed classical ML
2019	Carcillo *et al.* [73]	Unsupervised detection	European dataset	Effective with limited labels
2020	Bahnsen *et al.* [63]	Cost-sensitive DT	Financial data	Reduced false negatives
2021	Al-Zoubi *et al.* [72]	Metaheuristic FS	Credit card data	Improved F1-score
2022	Chen *et al*[67].	GA + RF	European dataset	Optimized feature subsets

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, February 2026
Impact Factor: 8.2

2023	Xiang *et al.* [70]	Graph Neural Networks	IEEE-CIS	Captured relational fraud patterns
2024	El-Kenawy *et al.* [69]	CNN–LSTM	Public datasets	Improved temporal detection
2024	Assabil & Obagbuwa[74]	ML comparison	MLG-ULB	RF achieved best balance
2024	Zhu *et al.* [75]	NN + SMOTE	Imbalanced data	Enhanced recall
2025	Sha *et al.* [76]	Attention-based GNN	Transaction graphs	State-of-the-art ROC-AUC
2025	Hafez *et al*[77].	Systematic AI review	Multiple	Identified DL & ensembles as dominant

II. CONCLUSION

This review has provided a comprehensive analysis of credit card fraud detection using machine learning algorithms, highlighting the progression from traditional rule-based and classical ML methods to advanced deep learning and graph-based approaches. Ensemble and deep learning models demonstrate superior performance, particularly when supported by effective preprocessing, feature engineering, and imbalance handling strategies.

Despite significant advancements, challenges such as concept drift, interpretability, real-time deployment, and data privacy remain unresolved. Emerging research directions—including graph neural networks, semi-supervised learning, and adaptive feature selection—offer promising avenues for addressing these limitations. Future fraud detection systems must balance predictive accuracy with scalability, transparency, and regulatory compliance to ensure sustainable deployment in real-world financial environments.

REFERENCES

[1]. Federal Trade Commission, "Consumer Sentinel Network Data Book 2023," Feb. 2024. [Online]. Available: https://www.ftc.gov/reports/consumer-sentinel-network-data-book-2023

[2]. Nilson Report, "Global Card Fraud Losses Reach $28.65 Billion," Issue 1164, Oct. 2019.

[3]. A Abdallah, M. A. Maarof, and A. Zainal, "Fraud detection system: A survey," J. Netw. Comput. Appl., vol. 68, pp. 90–113, Jun. 2016.

[4]. L. Dhanabal and S. P. Shantharajah, "A study on NSL-KDD dataset for intrusion detection system based on classification algorithms," Int. J. Adv. Res. Comput. Commun. Eng., vol. 4, no. 6, pp. 446–452, Jun. 2015.

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

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

[7]. J. O. Chan, "An architecture for big data analytics," Commun. IIMA, vol. 13, no. 2, pp. 1–14, 2013.

[8]. C. Phua, V. Lee, K. Smith, and R. Gayler, "A comprehensive survey of data mining-based fraud detection research," Artif. Intell. Rev., vol. 36, no. 4, pp. 245–268, Nov. 2011.

[9]. N. Carneiro, G. Figueira, and M. Costa, "A data mining based system for credit-card fraud detection in e-tail," Decis. Support Syst., vol. 95, pp. 91–101, Mar. 2017.

[10]. Y. Sahin and E. Duman, "Detecting credit card fraud by ANN and logistic regression," in Proc. Int. Symp. Innov. Intell. Syst. Appl., Istanbul, Turkey, Jun. 2011, pp. 315–319.

[11]. V. N. Vapnik, The Nature of Statistical Learning Theory, 2nd ed. New York, NY, USA: Springer, 2000.

[12]. A C. Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, "Feature engineering strategies for credit card fraud detection," Expert Syst. Appl., vol. 51, pp. 134–142, Jun. 2016.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, February 2026
Impact Factor: 8.2

[13]. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," J. Artif. Intell. Res., vol. 16, pp. 321–357, Jun. 2002.

[14]. H. He and E. A. Garcia, "Learning from imbalanced data," IEEE Trans. Knowl. Data Eng., vol. 21, no. 9, pp. 1263–1284, Sep. 2009.

[15]. M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches," IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 42, no. 4, pp. 463–484, Jul. 2012.

[16]. A C. Bahnsen, D. Aouada, and B. Ottersten, "Example-dependent cost-sensitive decision trees," Expert Syst. Appl., vol. 42, no. 19, pp. 6609–6619, Nov. 2015.

[17]. S. Ghosh and D. L. Reilly, "Credit card fraud detection with a neural-network," in Proc. 27th Hawaii Int. Conf. Syst. Sci., Wailea, HI, USA, Jan. 1994, pp. 621–630.

[18]. D. J. Hand, C. Whitrow, N. M. Adams, P. Juszczak, and D. Weston, "Performance criteria for plastic card fraud detection tools," J. Oper. Res. Soc., vol. 59, no. 7, pp. 956–962, Jul. 2008.

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

[20]. M. Zareapoor and P. Shamsolmoali, "Application of credit card fraud detection: Based on bagging ensemble classifier," Procedia Comput. Sci., vol. 48, pp. 679–685, 2015.

[21]. A Srivastava, A. Kundu, S. Sural, and A. K. Majumdar, "Credit card fraud detection using hidden Markov model," IEEE Trans. Dependable Secure Comput., vol. 5, no. 1, pp. 37–48, Jan.–Mar. 2008.

[22]. Y. Cheng, Y. Wang, H. Cao, Y. Qian, and H. Wang, "Model compression and acceleration for deep neural networks: The principles, progress, and challenges," IEEE Signal Process. Mag., vol. 35, no. 1, pp. 126–136, Jan. 2018.

[23]. A Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "Credit card fraud detection: A realistic modeling and a novel learning strategy," IEEE Trans. Neural Netw. Learn. Syst., vol. 29, no. 8, pp. 3784–3797, Aug. 2018.

[24]. Machine Learning Group, "Credit Card Fraud Detection Dataset," Université Libre de Bruxelles, Brussels, Belgium, 2013. [Online]. Available: https://www.kaggle.com/mlg-ulb/creditcardfraud

[25]. A Dal Pozzolo, O. Caelen, R. A. Johnson, and G. Bontempi, "Calibrating probability with undersampling for unbalanced classification," in Proc. IEEE Symp. Comput. Intell. Data Mining, Cape Town, South Africa, Dec. 2015, pp. 159–166.

[26]. IEEE Computational Intelligence Society, "IEEE-CIS Fraud Detection," Kaggle Competition, 2019. [Online]. Available: https://www.kaggle.com/c/ieee-fraud-detection

[27]. R. J. A. Little and D. B. Rubin, Statistical Analysis with Missing Data, 3rd ed. Hoboken, NJ, USA: Wiley, 2019.

[28]. S. García, J. Luengo, J. A. Sáez, V. López, and F. Herrera, "A survey of discretization techniques: Taxonomy and empirical analysis in supervised learning," IEEE Trans. Knowl. Data Eng., vol. 25, no. 4, pp. 734–750, Apr. 2013.

[29]. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining, San Francisco, CA, USA, Aug. 2016, pp. 785–794.

[30]. J. Hancock and T. Khoshgoftaar, "CatBoost for big data: An interdisciplinary review," J. Big Data, vol. 7, no. 1, Art. no. 94, Dec. 2020.

[31]. C. Guo and F. Berkhahn, "Entity embeddings of categorical variables," arXiv:1604.06737, Apr. 2016.

[32]. Y. Jurgovsky, M. Granitzer, K. Ziegler, S. Calabretto, P.-E. Portier, L. He-Guelton, and O. Caelen, "Sequence classification for credit-card fraud detection," Expert Syst. Appl., vol. 100, pp. 234–245, Jun. 2018.

[33]. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in Proc. Adv. Neural Inf. Process. Syst., Montreal, QC, Canada, Dec. 2014, pp. 3104–3112.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, February 2026
Impact Factor: 8.2

[34]. T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. New York, NY, USA: Springer, 2009.
[35]. R. Quinlan, C4.5: Programs for Machine Learning. San Mateo, CA, USA: Morgan Kaufmann, 1993.
[36]. G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning with Applications in R. New York, NY, USA: Springer, 2013.
[37]. M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems?" J. Mach. Learn. Res., vol. 15, no. 1, pp. 3133–3181, Jan. 2014.
[38]. A Liaw and M. Wiener, "Classification and regression by randomForest," R News, vol. 2, no. 3, pp. 18–22, Dec. 2002.
[39]. M. García, C. A. Torres, and J. R. Hernández, "Ensemble methods and emerging paradigms in credit card fraud detection: A comparative study," Preprints, 2025, doi: 10.20944/preprints202509.1909.v1.
[40]. S. S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: A comparative study," Decis. Support Syst., vol. 50, no. 3, pp. 602–613, Feb. 2011.
[41]. C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," IEEE Trans. Neural Netw., vol. 13, no. 2, pp. 415–425, Mar. 2002.
[42]. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining, San Francisco, CA, USA, Aug. 2016, pp. 785–794.
[43]. G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in Proc. Adv. Neural Inf. Process. Syst., Long Beach, CA, USA, Dec. 2017, pp. 3146–3154.
[44]. N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," Amer. Statist., vol. 46, no. 3, pp. 175–185, Aug. 1992.
[45]. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, "When is 'nearest neighbor' meaningful?" in Proc. 7th Int. Conf. Database Theory, Jerusalem, Israel, Jan. 1999, pp. 217–235.
[46]. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.
[47]. T. Nguyen, H. Tahir, M. Abdelrazek, P. Vasa, and J. Grundy, "Deep learning methods for credit card fraud detection," arXiv:2012.03754, Dec. 2020.
[48]. Roy, M. Chatterjee, A. Bandyopadhyay, and S. K. Kar, "A comparative study of machine learning and deep learning techniques for credit card fraud detection," in Proc. IEEE Int. Conf. Mach. Learn. Appl., Orlando, FL, USA, Dec. 2018, pp. 1032–1037.
[49]. S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
[50]. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," IEEE Trans. Neural Netw. Learn. Syst., vol. 28, no. 10, pp. 2222–2232, Oct. 2017.
[51]. Y. Jurgovsky, M. Granitzer, K. Ziegler, S. Calabretto, P.-E. Portier, L. He-Guelton, and O. Caelen, "Sequence classification for credit-card fraud detection," Expert Syst. Appl., vol. 100, pp. 234–245, Jun. 2018.
[52]. R. Fu, Z. Zhang, and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in Proc. 31st Youth Acad. Annu. Conf. Chin. Assoc. Autom., Wuhan, China, Nov. 2016, pp. 324–328.
[53]. N. Hossain and M. Hassan, "Analyzing the classification accuracy of deep learning and machine learning for credit card fraud detection," Asian J. Converg. Technol., vol. 8, no. 3, pp. 31–37, Dec. 2022.
[54]. F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," Neural Comput., vol. 12, no. 10, pp. 2451–2471, Oct. 2000.
[55]. A Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proc. Adv. Neural Inf. Process. Syst., Lake Tahoe, NV, USA, Dec. 2012, pp. 1097–1105.
[56]. O'Shea and R. Nash, "An introduction to convolutional neural networks," arXiv:1511.08458, Nov. 2015.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, February 2026
Impact Factor: 8.2

[57]. X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in Proc. Adv. Neural Inf. Process. Syst., Montreal, QC, Canada, Dec. 2015, pp. 802–810.

[58]. J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Boston, MA, USA, Jun. 2015, pp. 2625–2634.

[59]. Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu, "LSTM network: A deep learning approach for short-term traffic forecast," IET Intell. Transp. Syst., vol. 11, no. 2, pp. 68–75, Mar. 2017.

[60]. Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2018). Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy. IEEE transactions on neural networks and learning systems, 29(8), 3784–3797. https://doi.org/10.1109/TNNLS.2017.2736643

[61]. Gama, João & Žliobaitė, Indrė & Bifet, Albert & Pechenizkiy, Mykola & Bouchachia, Hamid. (2014). A Survey on Concept Drift Adaptation. ACM Computing Surveys (CSUR). 46. 10.1145/2523813.

[62]. A Dal Pozzolo et al., "Calibrating probability with undersampling for unbalanced classification," in Proc. IEEE CIDM, 2015.

[63]. C. Bahnsen et al., "Cost-sensitive decision trees for fraud detection," Expert Syst. Appl., 2015.

[64]. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. ArXiv. https://doi.org/10.1613/jair.953

[65]. .Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA, 785–794. https://doi.org/10.1145/2939672.2939785

[66]. Y. Chen et al., "GA-based feature selection for fraud detection," J. Big Data, 2022.

[67]. J. Roy et al., "Deep learning detecting fraud," IEEE ICMLA, 2018.

[68]. E.-S. M. El-Kenawy et al., "DL-based credit card fraud detection," J. Big Data, 2024.

[69]. S. Xiang et al., "Graph neural networks for fraud detection," IEEE Access, 2023.

[70]. A Dal Pozzolo et al., "Adaptive ML for fraud detection," Pattern Recognit. Lett., 2017.

[71]. Al-Zoubi et al., "Metaheuristic feature selection," Expert Syst. Appl., 2021.

[72]. J. Assabil and I. Obagbuwa, "Comparative ML analysis," Int. J. Intell. Syst. Appl. Eng., 2024.

[73]. Zhu et al., "NN with SMOTE," arXiv, 2024.

[74]. Q. Sha et al., "Attention-based GNN for fraud detection," arXiv, 2025.

[75]. A Y. Hafez et al., "Systematic review of AI in fraud detection," J. Big Data, 2025.