Article

Robust Fraud Detection with Ensemble Learning: A Case Study on the IEEE-CIS Dataset

Fatemeh Moradi [*] , Mehran Tarif [*] , MohammadHossein Homaei [*]

Posted Date: 21 July 2025

doi: 10.20944/preprints202507.1711.v1

Keywords: Credit Card Fraud Detection; Ensemble learning; Imbalanced Classification; IEEE-CIS Dataset; Machine Learning; Financial Security

Article

Robust Fraud Detection with Ensemble Learning: A Case Study on the IEEE-CIS Dataset

Fatemeh Moradi [1,*], Mehran Tarif [2,*] and Mohammadhossein Homaei [3,*]

[1] Faculty of Engineering, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran
[2] Department of Computer Science, University of Verona, Verona, Italy
[3] Media Engineering Group, University of Extremadura, Cáceres, Spain
[*] Correspondence: moradi.fatemeh2001@gmail.com (F.M.); mehran.tarifhokmabadi@univr.it (M.T.); homaei@ieee.org (M.H.)

Abstract

The rapid growth of digital financial transactions has led to a corresponding increase in credit card fraud, necessitating the development of sophisticated detection systems. This paper presents a comprehensive analysis of advanced ensemble learning techniques for imbalanced fraud detection using the IEEE-CIS dataset. We address the critical challenges of extreme class imbalance, concept drift, and real-time detection requirements through systematic evaluation of ensemble methods including Random Forest, XGBoost, LightGBM, and novel stacking approaches. Our methodology incorporates advanced data balancing techniques (SMOTE, ADASYN, Borderline-SMOTE) and feature engineering strategies optimized for the IEEE-CIS dataset containing 590,540 transactions with 3.5% fraud rate. Experimental results demonstrate that our proposed ensemble stacking approach achieves superior performance with 91.8% AUC-ROC, 0.891 AUC-PR, and significant improvements in fraud detection rates while maintaining low false positive rates. The study provides empirical evidence for the effectiveness of ensemble methods in handling severely imbalanced financial fraud datasets and offers practical insights for real-world implementation.

Keywords: credit card fraud detection; ensemble learning; imbalanced classification; IEEE-CIS dataset; machine learning; financial security

1. Introduction

Financial fraud has emerged as one of the most pressing challenges in the digital economy, with credit card fraud representing a significant threat to both financial institutions and consumers worldwide. The exponential growth of digital payment systems has created unprecedented opportunities for fraudulent activities, resulting in substantial financial losses and undermining consumer confidence in electronic transactions. According to recent industry reports, fraud losses continue to escalate dramatically, with card-not-present transactions being particularly vulnerable [1–3]. Machine learning has emerged as the predominant approach for addressing these challenges, with ensemble methods showing particular promise due to their ability to combine multiple base learners and create superior models [4,5].

Fraud detection faces several technical challenges that make traditional rule-based systems inadequate. First, extreme class imbalance in fraud datasets poses significant challenges, as fraudulent transactions typically represent less than 1% of all transactions, leading to models that achieve high overall accuracy while failing to detect actual fraud [6]. Second, the dynamic nature of fraudulent patterns, known as concept drift, requires detection systems to continuously adapt to evolving fraud tactics [7]. Third, real-time processing requirements demand algorithms that can make accurate decisions within milliseconds while processing thousands of transactions per second, making traditional batch processing approaches insufficient.

Despite significant advances, several gaps remain in the systematic evaluation of ensemble methods for imbalanced fraud detection. Many existing studies focus on individual algorithms or limited ensemble combinations, lacking comprehensive analysis of diverse ensemble strategies [8,9]. Additionally, most research employs older datasets with limited feature complexity, reducing applicability to modern scenarios. Few studies have leveraged comprehensive datasets like the IEEE-CIS Fraud Detection dataset, which contains 590,540 card-not-present transactions with a 3.5% fraud rate and 431 anonymized features capturing diverse transaction characteristics [10]. This motivates systematic evaluation of advanced ensemble learning techniques using realistic, large-scale datasets.

This paper addresses these limitations through comprehensive analysis of advanced ensemble learning techniques for imbalanced fraud detection using the IEEE-CIS dataset. Our key contributions include: (1) systematic evaluation of diverse ensemble methods; (2) comparison of multiple data balancing techniques; (3) proposal of a novel ensemble stacking architecture; (4) comprehensive feature engineering and selection analysis; and (5) detailed performance analysis using appropriate imbalanced classification metrics.

The remainder of this paper is organized as follows: Section 2 reviews the related work in fraud detection and ensemble learning. Section 3 presents the proposed methodology, including dataset preprocessing, feature engineering, and ensemble model design. Section 4 details the experimental setup and evaluation metrics used in the study. Section 5 provides comprehensive results and performance analysis. Section 6 discusses the implications, practical considerations, and limitations of our approach. Finally, Section 7 concludes the paper with key findings and outlines directions for future research.

2. Related Work

2.1. Ensemble Learning for Fraud Detection

Ensemble learning has emerged as one of the most effective approaches for fraud detection, offering superior performance compared to individual algorithms by combining multiple learners to create more robust and accurate models. Recent research has demonstrated the significant advantages of ensemble methods in financial fraud detection applications.

Khalid et al. [1] presented a comprehensive ensemble approach integrating Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Random Forest (RF), Bagging, and Boosting classifiers. Their methodology addressed dataset imbalance through under-sampling and Synthetic Minority Oversampling Technique (SMOTE), achieving significant improvements in fraud detection accuracy. This work highlighted the potential of ensemble methods to address multiple challenges simultaneously, including data imbalance, generalization, and real-time processing requirements.

Agarwal et al. [11] demonstrated the effectiveness of ensemble learning approaches specifically for fraud detection in financial transactions. Their study showed that Random Forest ensemble methods are particularly effective at detecting fraudulent financial transactions, using large datasets and advanced feature engineering to improve fraud detection by identifying subtle connections and patterns in transaction data.

The stacking ensemble approach has shown particular promise in recent research. Almalki and Masud [12] proposed a fraud detection framework combining stacking ensemble methods with explainable AI techniques. Their approach integrated XGBoost, LightGBM, and CatBoost using stacking ensemble methodology while incorporating model interpretability features. The study achieved 99% accuracy and 0.99 AUC-ROC on financial datasets, demonstrating the potential of combining high performance with model transparency.

2.2. Class Imbalance Handling in Fraud Detection

The extreme class imbalance inherent in fraud datasets represents one of the most significant challenges in fraud detection research. Fraudulent transactions typically represent less than 1% of all transactions, leading to models that may achieve high overall accuracy while failing to detect actual fraud cases.

The Synthetic Minority Oversampling Technique (SMOTE) has become the de facto standard for addressing class imbalance in fraud detection [13]. SMOTE generates synthetic minority instances by interpolating between existing minority samples and their k-nearest neighbors. Recent research has explored various SMOTE variants optimized for specific fraud detection scenarios.

Elreedy et al. [14] provided a comprehensive theoretical analysis of SMOTE for imbalanced learning, demonstrating its effectiveness across various machine learning applications. Their work established the theoretical foundations for understanding when and why SMOTE is effective in fraud detection scenarios.

Salehi and Khedmati [15] proposed a Cluster-based SMOTE Both-sampling (CSBBoost) ensemble algorithm for classifying imbalanced data. Their approach combines over-sampling, under-sampling, and different ensemble algorithms, including Extreme Gradient Boosting, to address class imbalance more effectively than traditional approaches.

2.3. IEEE-CIS Dataset Studies

The IEEE-CIS Fraud Detection dataset has become a standard benchmark for evaluating fraud detection algorithms, providing researchers with a realistic and comprehensive dataset for algorithm development and comparison. The dataset contains 590,540 card transactions, 20,663 of which are fraudulent (3.5%), with 431 features including both numerical and categorical variables [16].

Recent studies utilizing the IEEE-CIS dataset have demonstrated the effectiveness of various advanced techniques. Zhao et al. [17] presented a hybrid machine learning model for enhancing transaction fraud detection using the IEEE-CIS dataset, achieving significant improvements over baseline approaches through comprehensive feature engineering and model optimization.

The dataset's realistic characteristics, including temporal patterns, missing values, and anonymized features, make it an ideal testbed for evaluating the practical applicability of fraud detection algorithms in real-world scenarios.

2.4. Deep Learning and Advanced Approaches

Recent advances in deep learning have shown promising results for fraud detection applications. Mienye and Jere [6] conducted a comprehensive review of deep learning approaches for credit card fraud detection, analyzing 57 studies published between 2019-2024. Their survey revealed a clear trend toward deep learning approaches, with Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Transformer models becoming increasingly prevalent.

The integration of deep learning with ensemble methods has shown particular promise. Recent research has demonstrated that combining traditional ensemble approaches with deep learning components can achieve superior performance while maintaining computational efficiency for real-time fraud detection applications.

Tables 1 and 2 summarize key studies on ensemble learning methods for fraud detection and their dataset usage. Table 1 outlines the ensemble strategies, base classifiers, and core contributions across recent works. Notably, prior research explored hybrid ensembles and stacking with explainable AI, but did not leverage large-scale benchmarks like IEEE-CIS.

Table 1. Comparison of Ensemble Learning Approaches for Fraud Detection.

Year/Ref	Ensemble Method	Base Classifiers	Dataset	Key Contribution
2024 [1]	Bagging + Boosting + Individual	SVM, KNN, RF, Bagging, Boosting	Custom	Imbalance-aware ensemble integration across multiple classifiers
2024 [11]	Random Forest	Decision Trees	Financial Transactions	Advanced feature engineering for fraud signal enhancement
2024 [12]	Stacking Ensemble	XGBoost, LightGBM, CatBoost	Financial Dataset	Integration of explainable AI with high-performing ensemble

In Table 2, we further compare the datasets and performance metrics used in related studies. While most works employ proprietary or generic financial datasets, only Zhao et al. and the IEEE-CIS benchmark utilize the comprehensive IEEE-CIS dataset. However, their ensemble configurations and feature engineering pipelines remain limited compared to our proposed stacking ensemble, which integrates advanced sampling and feature processing strategies tailored to this dataset.

Table 2. Dataset Characteristics and Performance Metrics.

Study	Dataset	Size	Fraud Rate	Features	Best Performance
[1]	Custom	Not Specified	Imbalanced	Not Specified	Improved accuracy through ensemble and sampling strategies
[11]	Financial Transactions	Large Dataset	Not Specified	Engineered Features	Detection enhanced by feature construction
[12]	Financial Dataset	Not Specified	Not Specified	Not Specified	99% Accuracy, 0.99 AUC-ROC with explainable stacking
[17]	IEEE-CIS	590,540	3.5%	431	Significant improvement via hybrid learning
IEEE-CIS Benchmark	IEEE-CIS	590,540	3.5% (20,663)	431	Baseline reference for model validation

3. Methodology

This section presents our comprehensive methodology for developing and evaluating advanced ensemble learning techniques for imbalanced fraud detection using the IEEE-CIS dataset.

3.1. Dataset Description and Characteristics

The IEEE-CIS Fraud Detection dataset serves as the foundation for our experimental evaluation. This dataset represents one of the most comprehensive and realistic fraud detection benchmarks available, containing real-world e-commerce transaction data with sophisticated feature engineering.

The dataset comprises 590,540 card-not-present transactions collected over a six-month period, with 20,663 fraudulent transactions representing a 3.5% fraud rate. This imbalance ratio, while more balanced than typical real-world scenarios (often <1%), still presents significant challenges for machine learning algorithms and provides a realistic testbed for imbalanced learning techniques.

The original dataset contains 431 features across two files: transaction data (394 features) and identity data (37 features), joined by TransactionID. Features are categorized into several groups: temporal features (TransactionDT), transaction amount (TransactionAmt), payment card features (card1-card6), address features (addr1-addr2), distance features (dist1-dist2), and numerous anonymized categorical and numerical features (C1-C14, D1-D15, M1-M9, V1-V339) representing proprietary Vesta Corporation feature engineering.

3.2. Data Preprocessing and Feature Engineering

Our preprocessing pipeline addresses the challenges of missing values, feature selection, and data quality issues inherent in real-world financial datasets. The methodology follows a systematic approach designed to preserve signal while reducing noise and computational complexity.

3.2.1. Feature Preprocessing Pipeline

Table 3 details our systematic feature preprocessing pipeline that reduces the original 431 features to a manageable baseline set.

Table 3. Feature Preprocessing Pipeline

Preprocessing Step	Features Remaining	Features Removed
Original IEEE-CIS Dataset	431	-
Remove Features >95% Missing	298	133
Remove Zero-Variance Features	276	22
Remove Highly Correlated (>0.98)	203	73
Remove Low Information Gain (<0.001)	167	36
Baseline Feature Set	**167**	**264 total**

3.2.2. Missing Value Analysis and Treatment

The IEEE-CIS dataset exhibits substantial missing value patterns, with some features missing in over 90% of transactions. We implement a multi-stage missing value treatment strategy:

1. **Feature Removal**: Features with >95% missing values are removed to prevent sparse representations that may mislead ensemble learners.
2. **Strategic Imputation**: For categorical features with <95% missing values, we create explicit "missing" categories to capture the informational content of missingness patterns.
3. **Numerical Imputation**: Numerical features employ median imputation within fraud/legitimate groups separately to preserve class-specific distributions.
4. **Missingness Indicators**: Binary indicators are created for features with >20% missing values to capture missingness patterns as potential fraud signals.

3.2.3. Feature Engineering Strategy

Our feature engineering approach combines domain knowledge with automated feature generation to enhance the predictive power of the baseline 167-feature dataset:

Temporal Features (15 new features): The TransactionDT feature is decomposed into interpretable temporal components including hour of day, day of week, and day of month. Additionally, we create velocity features measuring transaction frequency within sliding time windows.

Amount-based Features (12 new features): Transaction amounts undergo log transformation to handle skewness, and we create amount percentile rankings within user groups and time periods.

Aggregation Features (28 new features): User-level aggregations compute transaction statistics over various time windows. Card-level and email-level aggregations capture usage patterns that may indicate account takeover or card sharing.

Interaction Features (25 new features): High-order interactions between categorical features are created using target encoding and frequency encoding.

The complete feature engineering pipeline creates 80 additional features, resulting in 247 total features for model training.

3.3. Handling Class Imbalance

The 3.5% fraud rate in the IEEE-CIS dataset necessitates sophisticated approaches to address class imbalance. We implement a comprehensive strategy combining multiple resampling techniques with cost-sensitive learning approaches.

3.3.1. Synthetic Oversampling Techniques

We evaluate three primary oversampling approaches:

SMOTE: Generates synthetic minority instances by interpolating between existing minority samples and their k-nearest neighbors [13].

Borderline-SMOTE: Focuses synthetic sample generation on borderline minority instances that are most likely to be misclassified.

ADASYN: Adaptively generates synthetic samples with density distribution, creating more synthetic instances for minority samples that are harder to learn.

3.4. Ensemble Model Architecture

Our ensemble architecture employs a three-tier approach designed to maximize diversity while maintaining computational efficiency. The architecture balances different algorithmic paradigms to capture diverse patterns in fraudulent behavior.

3.4.1. Base Learner Selection and Configuration

We select base learners representing different algorithmic families to ensure ensemble diversity:

Tree-based Learners: Random Forest, XGBoost, LightGBM, and CatBoost provide robust performance with different optimization strategies.

Linear Models: Logistic Regression with L1/L2 regularization provides interpretable linear decision boundaries.

Distance-based Models: K-Nearest Neighbors captures local patterns and non-linear relationships.

Neural Networks: Multi-layer Perceptron provides non-linear transformation capabilities.

3.4.2. Meta-Learning Strategy

Our meta-learning approach combines base learner predictions through a secondary learning phase using stacking with cross-validation. Base learners are trained on K-1 folds and predict on the remaining fold, creating out-of-fold predictions that serve as features for the meta-learner.

4. Experimental Setup

4.1. Implementation Environment and Tools

Our experimental framework leverages modern machine learning libraries including scikit-learn 1.2.0, XGBoost 1.7.3, LightGBM 3.3.5, and CatBoost 1.2 for gradient boosting implementations. Feature engineering utilizes pandas 1.5.3 and NumPy 1.24.2 for data manipulation.

All experiments are conducted on high-performance computing infrastructure with proper version control and containerized environments to ensure reproducibility.

4.2. Evaluation Metrics and Performance Assessment

Given the imbalanced nature of fraud detection, our evaluation employs multiple complementary metrics specifically designed for imbalanced classification problems.

4.2.1. Primary Evaluation Metrics

Area Under the ROC Curve (AUC-ROC): Serves as our primary metric for overall model discrimination capability.

Area Under the Precision-Recall Curve (AUC-PR): Provides complementary evaluation for imbalanced datasets [18].

F1-Score: Computes the harmonic mean of precision and recall, providing a balanced measure.

Balanced Accuracy: Calculated as the average of sensitivity and specificity.

G-Mean: Computes the geometric mean of sensitivity and specificity.

4.3. Cross-Validation and Model Selection

Our validation strategy addresses the unique challenges of time-series fraud data while ensuring robust model selection. We implement nested cross-validation with time-series aware folding to prevent temporal leakage.

5. Results and Analysis

This section presents comprehensive experimental results demonstrating the effectiveness of our proposed ensemble learning approach for imbalanced fraud detection.

5.1. Overall Performance Comparison

Table 4 presents the primary performance comparison across all evaluated methods.

Table 4. Overall Performance Comparison on IEEE-CIS Test Set

Method	AUC-ROC	AUC-PR	F1-Score	Balanced Acc.	G-Mean
Proposed Stacking	**0.918**	**0.891**	**0.856**	**0.923**	**0.918**
XGBoost	0.887	0.834	0.798	0.887	0.882
LightGBM	0.882	0.828	0.791	0.881	0.876
Random Forest	0.869	0.802	0.765	0.864	0.859
Weighted Voting	0.901	0.847	0.812	0.898	0.893
Simple Voting	0.878	0.823	0.784	0.876	0.871
CatBoost	0.873	0.821	0.775	0.869	0.864
Logistic Regression	0.829	0.743	0.701	0.821	0.815

Our proposed stacking ensemble achieves superior performance across all key metrics, demonstrating significant improvements over individual algorithms and simpler ensemble approaches. The stacking ensemble achieves an AUC-ROC of 0.918 and AUC-PR of 0.891, representing a 3.5% improvement in AUC-ROC and 6.8% improvement in AUC-PR over the best individual algorithm (XGBoost: 0.887 AUC-ROC, 0.834 AUC-PR).

5.2. Individual Algorithm Performance Analysis

Table 5 details the performance of individual base learners.

Table 5. Individual Algorithm Performance with Optimized Hyperparameters

Algorithm	AUC-ROC	AUC-PR	Training Time	Inference Time
XGBoost	0.887	0.834	18.3 min	45 ms
LightGBM	0.882	0.828	12.7 min	38 ms
CatBoost	0.873	0.821	24.1 min	52 ms
Random Forest	0.869	0.802	8.9 min	28 ms
Neural Network	0.841	0.786	15.6 min	35 ms
K-NN	0.826	0.771	2.1 min	125 ms
Logistic Regression	0.829	0.743	1.8 min	12 ms

Tree-based models dominate individual performance, with XGBoost achieving the highest AUC-ROC (0.887) and AUC-PR (0.834) followed closely by LightGBM (0.882 AUC-ROC, 0.828 AUC-PR). The computational analysis reveals important trade-offs between accuracy and efficiency.

5.3. Class Imbalance Handling Effectiveness

Table 6 demonstrates the critical importance of addressing class imbalance in fraud detection systems.

Table 6. Impact of Class Imbalance Handling Techniques

Technique	AUC-ROC	AUC-PR	FPR@95%R	Cost Reduction
SMOTE + Stacking	**0.918**	**0.891**	**0.94%**	**52.3%**
Borderline-SMOTE + Stacking	0.912	0.884	1.02%	48.7%
ADASYN + Stacking	0.908	0.876	1.15%	45.1%
SMOTE + Tomek + Stacking	0.915	0.888	0.98%	50.8%
No Sampling + Stacking	0.863	0.812	1.68%	31.2%
SMOTE + XGBoost	0.887	0.834	1.25%	42.6%
No Sampling + XGBoost	0.821	0.758	2.14%	24.8%

SMOTE combined with stacking achieves the best overall performance, with AUC-ROC improving from 0.863 (no sampling) to 0.918 and AUC-PR improving from 0.812 to 0.891, representing 6.4% and 9.7% relative improvements respectively. The cost-sensitive analysis reveals substantial operational benefits, with SMOTE + Stacking achieving 52.3% cost reduction compared to baseline approaches.

5.4. Feature Engineering Impact Assessment

Table 7 presents ablation study results quantifying the contribution of different feature engineering components.

Table 7. Feature Engineering Component Ablation Study

Feature Set	AUC-ROC	AUC-PR	F1-Score	Features	Δ AUC-PR
Complete Pipeline	0.918	0.891	0.856	247	-
- Interaction Features (25)	0.905	0.873	0.834	222	-0.018
- Aggregation Features (28)	0.892	0.856	0.812	219	-0.035
- Temporal Features (15)	0.883	0.847	0.801	232	-0.044
- Amount Engineering (12)	0.897	0.863	0.825	235	-0.028
Baseline Features Only	0.851	0.789	0.743	167	-0.102

The complete feature engineering pipeline provides a 10.2 percentage point improvement in AUC-PR over baseline features (0.891 vs 0.789), representing a 12.9% relative improvement and validating the systematic approach to feature development. Temporal features contribute the most significant improvement (4.4 percentage points), followed by aggregation features (3.5 percentage points).

5.5. Ensemble Architecture Analysis

Table 8 provides detailed comparison of different ensemble strategies.

Table 8. Ensemble Method Comparison with Statistical Significance

Ensemble Method	AUC-ROC	AUC-PR	Training Time	p-value
Stacking (Proposed)	**0.918**	**0.891**	**45.7 min**	**-**
Weighted Voting	0.901	0.847	32.4 min	< 0.001
Blending	0.895	0.842	38.9 min	< 0.001
Simple Voting	0.878	0.823	31.8 min	< 0.001
Bagging (RF)	0.869	0.808	26.3 min	< 0.001
AdaBoost	0.841	0.785	41.2 min	< 0.001

The stacking ensemble achieves statistically significant improvements over all alternative ensemble methods (p < 0.001 for all comparisons). The performance improvements indicate practically significant gains beyond statistical significance.

6. Discussion

Our findings demonstrate the effectiveness of systematic ensemble learning for addressing the complex challenges of imbalanced fraud detection. The superior performance of our stacking ensemble validates the hypothesis that combining diverse learning algorithms through meta-learning can significantly improve fraud detection accuracy.

The systematic feature preprocessing pipeline successfully reduced the original 431 features to a manageable 167-feature baseline set while preserving essential predictive information. The subsequent feature engineering process added 80 carefully designed features, resulting in substantial performance improvements that validate the importance of domain-informed feature development.

The success of SMOTE in addressing class imbalance, combined with the ensemble approach, provides practical guidance for implementing high-performance fraud detection systems. The substantial cost reduction achieved demonstrates clear operational benefits for financial institutions.

The comprehensive feature engineering analysis reveals the continued importance of domain expertise in machine learning applications, with systematic feature development providing significant performance improvements over baseline approaches.

7. Conclusions

This research presents a comprehensive analysis of advanced ensemble learning techniques for imbalanced fraud detection, demonstrating that our proposed stacking ensemble achieves superior performance compared to individual algorithms and simple ensemble methods. The key contributions include systematic evaluation of diverse ensemble methods, comprehensive analysis of class imbalance handling techniques, and a detailed feature preprocessing pipeline that reduces 431 original features to 167 baseline features through principled elimination, followed by targeted engineering to create 247 final features. This systematic methodology, combined with rigorous experimental design and computational analysis confirming practical feasibility for real-time deployment, represents substantial practical value for real-world fraud detection systems and provides a replicable framework for future fraud detection research.

Future research directions should focus on developing online learning adaptations to handle concept drift in evolving fraud patterns, integrating explainable AI techniques to enhance model interpretability and regulatory compliance, exploring transfer learning approaches to enable knowledge sharing across different fraud domains, and implementing federated learning frameworks for collaborative fraud detection while preserving data privacy. These research avenues will address critical challenges in deploying fraud detection systems at scale, ensuring they remain effective, transparent, and adaptable to emerging fraud tactics while maintaining the high-performance standards established in this work.

References

1. Khalid, A.R., Owoh, N., Uthmani, O., Ashawa, M., Osamor, J., and Adejoh, J. Enhancing credit card fraud detection: an ensemble machine learning approach. *Big Data and Cognitive Computing*, 8(1):6, 2024. MDPI.
2. Homaei, M.H., Caro Lindo, A., Sancho Núñez, J.C., Mogollón Gutiérrez, O., and Alonso Díaz, J. The Role of Artificial Intelligence in Digital Twin's Cybersecurity. In *Proceedings of the XVII Reunión Española sobre Criptología y Seguridad de la Información (RECSI 2022)*, 2022. Editorial Universidad de Cantabria.
3. Homaei, M., Mogollón-Gutiérrez, O., Sancho, J.C., Ávila, M., and Caro, A. A review of digital twins and their application in cybersecurity based on artificial intelligence. *Artificial Intelligence Review*, 57(8), 2024. Springer Science and Business Media LLC.
4. Gandhar, A., Gupta, K., Pandey, A.K., and Raj, D. Fraud detection using machine learning and deep learning. *SN Computer Science*, 5(5):453, 2024. Springer.
5. Moradi, F., Tarif Hokmabadi, M., and Homaei, M. A Systematic Review of Machine Learning in Credit Card Fraud Detection. *Preprints*, 2025.
6. Mienye, I.D. and Jere, N. Deep learning for credit card fraud detection: A review of algorithms, challenges, and solutions. *IEEE Access*, 2024. IEEE.
7. Chen, Y., Zhao, C., Xu, Y., and Nie, C. Year-over-Year Developments in Financial Fraud Detection via Deep Learning: A Systematic Literature Review. *arXiv preprint arXiv:2502.00201*, 2025.
8. Fernández, A., García, S., Galar, M., Prati, R.C., Krawczyk, B., and Herrera, F. *Learning from imbalanced data sets*, volume 10, 2018. Springer.
9. Talukder, M.A., Khalid, M., and Uddin, M.A. An integrated multistage ensemble machine learning model for fraudulent transaction detection. *Journal of Big Data*, 11(1), 2024. Springer Science and Business Media LLC.
10. Vesta Corporation. IEEE-CIS Fraud Detection Dataset. Kaggle Competition, 2019. https://www.kaggle.com/c/ieee-fraud-detection. Dataset provided for IEEE Computational Intelligence Society fraud detection competition.
11. Suganya, S.S., Nishanth, S., and Mohanadevi, D. Ensemble Learning Approaches for Fraud Detection in Financial Transactions. In *2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS)*, pages 805–810, 2023. IEEE.

12. Almalki, F. and Masud, M. Financial Fraud Detection Using Explainable AI and Stacking Ensemble Methods. *arXiv preprint arXiv:2505.10050*, 2025.
13. Chawla, N.V., Bowyer, K.W., Hall, L.O., and Kegelmeyer, W.P. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.
14. Elreedy, D., Atiya, A.F., and Kamalov, F. A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning. *Machine Learning*, 113(7):4903–4923, 2024. Springer.
15. Salehi, A.R. and Khedmati, M. A cluster-based SMOTE both-sampling (CSBBoost) ensemble algorithm for classifying imbalanced data. *Scientific Reports*, 14(1):5152, 2024. Nature Publishing Group UK London.
16. Papers with Code. IEEE CIS Fraud Detection Dataset. Online Repository, 2024. https://paperswithcode.com/dataset/ieee-cis-fraud-detection-1. Community-maintained dataset documentation.
17. Zhao, X., Zhang, Q., and Zhang, C. Enhancing Transaction Fraud Detection with a Hybrid Machine Learning Model. In *2024 IEEE 4th International Conference on Electronic Technology, Communication and Information (ICETCI)*, pages 427–432, 2024. IEEE.
18. Saito, T. and Rehmsmeier, M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, 10(3):e0118432, 2015. Public Library of Science San Francisco, CA USA.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.