IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 6, December 2026, pp. 5925~5935
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i6.pp5925-5935

Dynamic oversampling and generative techniques for enhanced credit card fraud detection

Mohamed Akouhar[1], Abdallah Abarda[2], Mohamed El Fatini[1], Mohamed Ouhssini[3], El Hafed Agherrabi[4]
[1]Laboratory of Analysis, Geometry and Applications, Ibn Tofail University, Kenitra, Morocco
[2]Laboratory of Mathematical Modeling and Economic Calculations, Hassan First University, Settat, Morocco
[3]C3S Laboratory, ESTC, Hassan II University, Casablanca, Morocco
[4]IR2M Laboratory, Hassan First University, Settat, Morocco

Article Info

Article history:

Received Apr 12, 2025
Revised Aug 9, 2026
Accepted Aug 31, 2026

Keywords:

Credit card fraud
Deep learning
Dynamic oversampling
Fraud detection
Generative techniques

ABSTRACT

The increasing sophistication and scale of credit card fraud necessitate advanced detection methods beyond traditional rule-based systems. Machine learning (ML) offers a promising approach but faces challenges such as class imbalance and overlap, leading to inaccurate predictions. This paper proposes a novel fraud detection system addressing these limitations through a dynamic oversampling rate (0.1 to 1.0) combined with advanced oversampling techniques, including the synthetic minority over-sampling technique (SMOTE), convolutional neural network generative adversarial networks (CNN-GANs), and deep neural network generative adversarial networks (DNN-GANs). This adaptive approach optimizes the balance between minority and majority classes while mitigating overfitting. Critically, model evaluation is performed exclusively on non-oversampled, real-world data, ensuring a realistic assessment of performance in practical scenarios. Experimental evaluation on the PaySim and European credit card datasets demonstrates the effectiveness of the proposed framework. On both datasets, the best results were obtained by combining CNN with CNN-GAN. On PaySim, this model achieved 99.99% accuracy and 99.99% area under the curve (AUC). On the European dataset, it attained 99.75% accuracy, 99.60% precision, 99.85% recall, 99.72% F1-score, 99.90% AUC, and 99.10% Matthews correlation coefficient (MCC), confirming robust performance across highly imbalanced financial transaction fraud detection tasks. The main contributions of this work include a flexible oversampling rate, advanced synthetic data generation, a robust testing methodology, and improved model robustness and adaptability for real-world fraud detection.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Mohamed Akouhar
Laboratory of Analysis, Geometry and Applications, University Ibn Tofail
Kenitra, Morocco
Email: m.akouhar@gmail.com

1. INTRODUCTION

Credit card fraud is a significant and escalating challenge in today's digital financial landscape, driven by technological advancements and the expansion of e-commerce [1], [2]. While digital payment systems and online marketplaces offer unprecedented convenience, they also expose vulnerabilities in online platforms and payment infrastructures that cybercriminals exploit [3], [4]. Traditional rule-based detection systems, relying on expert-defined rules and manual feature engineering, have become increasingly inadequate due to their static nature and inability to adapt swiftly to the dynamic strategies employed by

fraudsters [5], [6]. These systems are cumbersome to maintain and fail to keep pace with the volume and velocity of online transactions and the sophisticated tactics of modern fraud [7].

Machine learning (ML) presents a promising alternative by automatically detecting fraudulent patterns and adapting to evolving tactics [8], [9]. ML models can analyze vast amounts of transactional data to identify anomalies indicative of fraud, enhancing detection capabilities beyond traditional methods [10]. However, ML-based fraud detection faces significant limitations.

One primary issue is class imbalance legitimate transactions vastly outnumber fraudulent ones, which biases models toward the majority class [11]. This imbalance leads to unacceptable numbers of false negatives, resulting in substantial financial losses. Traditional techniques to address this, such as oversampling the minority class or undersampling the majority class, introduce new complexities. Oversampling can lead to overfitting by replicating minority class samples or generating synthetic data that doesn't generalize well, while undersampling risks underfitting by discarding valuable information from the majority class [10].

Another challenge is class overlap, where legitimate transactions share characteristics with fraudulent ones, complicating accurate detection and increasing false positives and negatives [12], [13]. Sophisticated fraud tactics, including cross-border and multi-party schemes, exacerbate this problem by blurring the lines between legitimate and illegitimate transactions [6], [14]. Moreover, the dynamic nature of fraud necessitates models that can continuously learn and adapt to evolving patterns. Static models, even with advanced ML techniques, quickly become outdated as new fraud strategies emerge, highlighting the need for continuous learning mechanisms to maintain effectiveness over time [15]–[18]. Despite substantial progress in resampling and generative adversarial network (GAN)-based augmentation, existing studies generally rely on fixed oversampling ratios and rarely combine adaptive rate selection with evaluation on an untouched, non-oversampled test set, leaving a clear gap in achieving both robust class balancing and realistic generalization.

To address these limitations, this study proposes a system that employs a dynamic oversampling rate, adjustable between 0.1 and 1.0. This flexibility allows for fine-tuning the degree of oversampling to find an optimal balance that mitigates class imbalance without inducing significant overfitting. This study utilizes advanced oversampling methods, including the synthetic minority over-sampling technique (SMOTE), convolutional neural network generative adversarial networks (CNN-GANs), and deep neural network generative adversarial networks (DNN-GANs). These techniques aim to generate synthetic minority samples that are more representative of the underlying data distribution, thereby improving the model's ability to generalize.

For evaluation, only real-world, non-oversampled data are used to assess model performance. This approach ensures that the proposed assessment reflects the model's true performance in practical scenarios and avoids the biases introduced by synthetic testing data. Testing on generated data is generally discouraged as it does not provide a reliable assessment of the model's performance on actual, unseen data, and can exacerbate overfitting due to synthetic data not accurately capturing the complexity of real-world distributions.

By combining a dynamic oversampling rate with robust testing on real data, the proposed approach effectively balances the dataset while minimizing the risks of overfitting and underfitting. This methodology enhances the model's robustness and reliability through continuous adaptation to the intrinsic variability of fraud patterns. It also provides a stronger generalization capability, allowing the system to maintain stable performance across different operational environments. Such properties make the approach particularly suitable for real-world fraud detection scenarios where data distributions are highly skewed and evolve over time. The main contributions of this paper are: i) dynamic oversampling rate: introducing a flexible oversampling rate (0.1 to 1.0) to optimize class balance; ii) advanced oversampling techniques: utilizing SMOTE, CNN-GANs, and DNN-GANs for generating representative synthetic samples; iii) robust testing methodology: evaluating models using non-oversampled, real-world data to ensure practical performance; iv) mitigation of overfitting and underfitting: addressing overfitting and underfitting issues inherent in traditional methods; v) enhanced model adaptability: improving model robustness and adaptability for real-world fraud detection scenarios.

2. LITERATURE REVIEW

Credit card fraud detection has been extensively studied, with a significant focus on addressing the inherent class imbalance in transactional datasets. Researchers have proposed various techniques to mitigate this imbalance and enhance the accuracy and reliability of fraud detection systems. This ongoing research effort underscores the importance of developing robust models capable of distinguishing rare fraudulent transactions from the overwhelming majority of legitimate ones.

One approach involves leveraging advanced neural network architectures. For instance, a study introduced a novel system using long short-term memory-recurrent neural networks (LSTM-RNN) integrated with attention mechanisms, achieving higher accuracy and reduced false positives compared to traditional classifiers like naive Bayes, support vector machine (SVM), and artificial neural network (ANN) [19]. However, the reliance on anonymized data limited insights into feature importance and real-world applicability.

Deep learning models using federated learning have also been explored. A model named jellyfish Namib beetle optimization-spinal network (JNBO-SpinalNet) combined hybrid optimization algorithms to detect fraud, addressing class imbalance through bootstrapping oversampling. While achieving strong accuracy, it faced limitations due to high false-positive rates and potential improvements needed in feature selection and handling imbalanced data [20]. Similarly, combining federated learning with graph neural networks (GNN) enhanced fraud detection across financial institutions by addressing data imbalance with oversampling techniques like the spiral oversampling balanced technique (SOBT) [21]. Yet, complexities in cross-institutional collaboration and computational demands posed challenges.

Hybrid sampling methods have been proposed to tackle dataset imbalance. A method using Mahalanobis distance-based SMOTE-edited nearest neighbors (ENN) hybrid sampling with random forest improved detection rates, although it introduced computational complexities and limited comparisons with non-ensemble classifiers [22]. Feature selection optimization using genetic algorithms (GA), coupled with oversampling techniques like SMOTE, was shown to enhance the performance of various ML classifiers, but testing on a single dataset raised concerns about generalizability [23].

Enhancements to existing algorithms have also been considered. An adaptive hybrid-weighted self-paced learning method improved the adaptive boosting (Adaboost) algorithm's classifier diversity, addressing class imbalance, though it faced potential underfitting in early iterations and sensitivity to parameter selection [24]. Additionally, an improved variational autoencoder generative adversarial network (VAEGAN) was utilized as an oversampling method, outperforming other techniques like GAN, VAE, and SMOTE in improving precision and F1-scores. Nevertheless, increased model complexity and variable recall and AUC scores were noted [25].

GNNs and autoencoders have been employed for real-time fraud detection using imbalanced datasets, showing promising results but still facing scalability limitations and difficulties in handling extreme imbalance levels [26]. Hybrid approaches combining conditional tabular GANs (CTGAN) and temporal convolutional networks (TCN) have attempted to address these challenges by generating synthetic samples and refining them through the neighborhood cleaning rule (NCL) to improve classification performance. Although these methods demonstrate measurable gains in precision and recall, they also introduce risks of overfitting and raise concerns about their ability to generalize effectively across diverse fraud detection datasets [27]. Threshold optimization techniques and undersampling were investigated to improve detection rates without oversampling, emphasizing the true positive rate over the true negative rate. However, these methods highlighted the trade-off between precision and dataset balance, with performance degradation observed when using random undersampling for precision [28]. Finally, deep learning approaches integrating various resampling techniques like SMOTE achieved high accuracy and eliminated false negatives, yet raised concerns about overfitting due to synthetic data generation and the need for generalization tests across different datasets [29].

Overall, while numerous strategies have been proposed to address class imbalance in credit card fraud detection, challenges remain in terms of computational complexity, overfitting, and generalizability. These limitations highlight the need for more adaptive and efficient methods capable of handling evolving data distributions without compromising model stability. Future research should therefore focus on developing robust models that effectively balance datasets while maintaining scalability and applicability to real-world scenarios. Table 1 synthesizes the landscape of current research against the requirements for an ideal fraud detection system.

Table1. Comparison of fraud detection techniques and research gaps

Approach category	Key techniques	Primary advantage	Significant research gap/limitation
Deep sequential	LSTM-RNN, TCN [19], [27]	Captures temporal patterns; low false positives.	Lack of feature interpretability; high overfitting risk.
Distributed/graph	Federated learning, GNN [20], [21], [26]	Privacy-preserving; maps complex relationships.	High computational demand; scalability issues in real-time.
Generative sampling	VAEGAN, CTGAN, SMOTE [22], [25], [29]	Directly addresses class imbalance.	Synthetic noise can lead to overfitting; variable recall scores.
Heuristic optimization	GA, self-paced learning [23], [24]	Optimizes feature sets and weightings.	High sensitivity to parameters; limited generalizability.
Proposed approach	CNN + CNNGAN	Superior feature extraction and high-fidelity synthetic data generation.	Requires significant training data for the GAN to stabilize.

3. PROPOSED APPROACH

This study proposes an approach to address the challenges of highly imbalanced data and high false negative rates, as identified in existing literature. Figure 1 and Algorithm 1 present the proposed approach, which involves a comprehensive comparison of various oversampling techniques, including SMOTE, CNNGAN, and DNNGAN. A key aspect of this approach is the careful adjustment of the oversampling rate to prevent overfitting, which can occur when the data is balanced at a 50% to 50% ratio. To enhance classification performance, these oversampling methods are integrated with advanced deep learning models such as CNN, LSTM, and GRU. The Hyperband technique is employed for hyperparameter tuning, optimizing these models effectively by exploring a wide range of hyperparameter configurations efficiently.

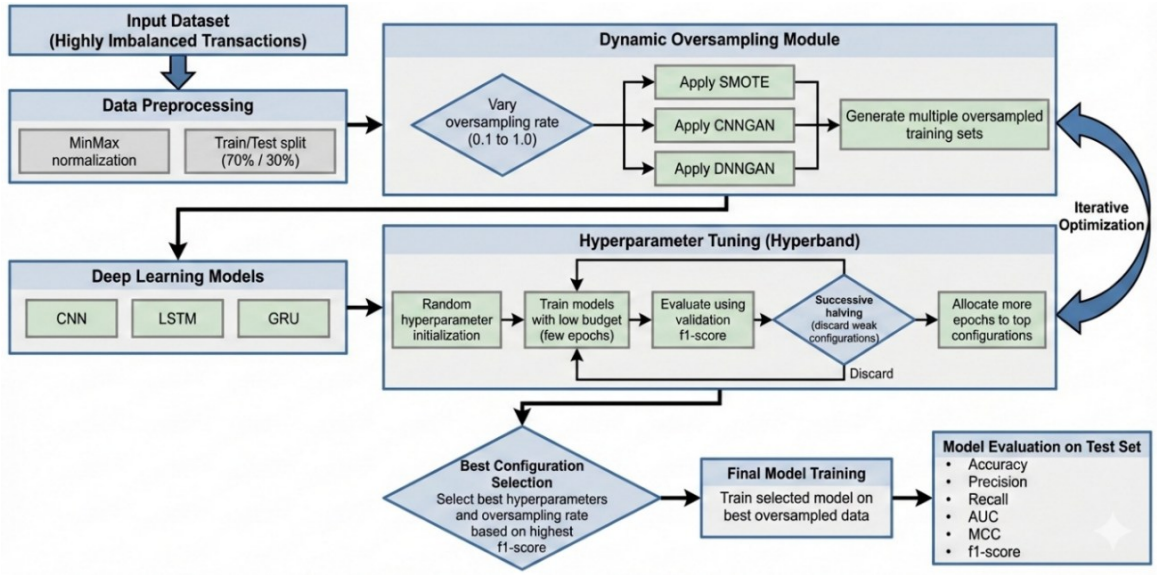

Figure 1. Systematic methodology for imbalanced data handling and hyperparameter selection in fraud detection systems

Algorithm 1. Data processing and model training

```
Preprocessing:
Normalize data using MinMax scaling.
Split data: 70% training, 30% testing.
Oversampling:
Methods: SMOTE, CNNGAN, and DNNGAN.
Rates: 0.1 to 1.0.
for each method and rate do
    Apply to training data.
end for
Model training:
Models: CNN, LSTM, GRU.
Hyperparameters: filters, units, dropout, learning rate.
Use hyperband for tuning.
for each model, method, and rate do
    Tune and train with hyperband.
end for
Evaluation:
Evaluate on testing set.
Calculate F1-score.
Select best:
Identify best rate and hyperparameters for highest F1-score.
Output:
Report best configurations and metrics.
```

The proposed approach is structured as follows: i) data preprocessing: normalize the dataset using MinMax scaling and split it into 70% training and 30% testing sets; ii) oversampling and tuning: apply oversampling methods (SMOTE, CNNGAN, and DNNGAN) to the training set at rates from 0.1 to 1.0 to evaluate their impact; iii) model training with hyperparameter tuning: use hyperband to tune hyperparameters for CNN, LSTM, and GRU models on the oversampled data, selecting the best configuration based on validation performance; iv) model evaluation: evaluate each model's performance on the test set using the F1-score; v) selection of best combination: identify the best oversampling rate and hyperparameters for each model-method combination based on the highest F1-score; vi) output results: report the best oversampling rate, hyperparameters, and performance metrics (accuracy, precision, recall, AUC, MCC, and F1-score) for each combination. This approach aims to enhance classification performance on imbalanced datasets by integrating advanced oversampling techniques with deep learning models and efficient hyperparameter tuning.

3.1. Datasets and preprocessing

3.1.1. Datasets used

In this study, two distinct datasets were used to evaluate the effectiveness of the proposed credit card fraud detection methods. The first dataset [30] is a synthetic dataset generated by PaySim, a simulator designed to mimic mobile money transactions. PaySim uses aggregated data from a private dataset to create a synthetic version that reflects typical transaction behaviors while injecting fraudulent activities for testing purposes. This dataset includes various transaction types such as CASH-IN, CASH-OUT, DEBIT, PAYMENT, and TRANSFER, and spans a simulated period of 30 days.

The second dataset [31] comprises real credit card transactions made by European cardholders in September 2013. This dataset is highly imbalanced, with only 492 fraudulent transactions out of 284,807 total transactions, representing a fraud rate of 0.172%. The features in this dataset are primarily the result of a principal component analysis (PCA) transformation, except 'Time' and 'Amount'. The 'Class' feature indicates whether a transaction is fraudulent.

3.1.2. Preprocessing

For preprocessing, normalization using the min-max scaling method was applied to ensure that all features were on a similar scale, which is crucial for many ML algorithms. The min-max normalization is defined by (1).

$$X_{new} = \frac{X - X_{max}}{X_{min} - X_{max}} \quad (1)$$

$X_{\min}$ and X_{max} are the minimum and maximum values of the feature, respectively. After normalization, each dataset was split into training and testing sets using a 70%-30% split, ensuring that the stratification strategy was applied. This stratification maintains the original distribution of the classes in both the training and testing sets, which is particularly important given the class imbalance in the credit card transactions dataset.

3.2. Oversampling and tuning

To address class imbalance in credit card fraud detection, several oversampling techniques are employed:

- SMOTE: generates synthetic samples of fraudulent transactions by interpolating between existing instances, balancing the dataset.
- CNNGAN: uses GANs with a CNN architecture to create realistic synthetic fraud samples, capturing complex transaction patterns.
- DNNGAN: utilizes a deep neural network to generate synthetic fraud samples, adaptable to various data types.

The oversampling rates are defined from 0.1 to 1.0 to explore different levels of data augmentation. For each method, these rates are applied to the training set, generating synthetic fraudulent transactions to balance the dataset. This approach aims to improve the model's ability to detect fraud by enhancing its sensitivity and specificity.

3.3. Model training with hyperparameter tuning

In this phase, the aim is to improve the performance of the models in detecting credit card fraud by carefully selecting and tuning their hyperparameters. Different hyperparameter configurations are evaluated to identify those that provide the best balance between detecting fraudulent transactions and minimizing false alarms. This process helps optimize the models' ability to distinguish between legitimate and fraudulent transactions.

3.3.1. Define models

This study focuses on training three types of models capable of learning complex patterns in data: CNN, LSTM, and GRU. CNN captures local patterns and hierarchical features, which can help identify relationships among transaction attributes. LSTM learns long-term dependencies in sequential data, making it useful for analyzing transaction histories. GRU offers a simpler architecture than LSTM, often requiring less training time while still capturing sequential dependencies.

3.3.2. Define hyperparameter search space

For each model, key hyperparameters that significantly affect performance are identified: i) for CNNs, the number of filters, kernel size, dropout rate, and learning rate; ii) for LSTMs, the number of units, dropout rate, and learning rate; and iii) for GRUs, the number of units, dropout rate, and learning rate are also tuned. Hyperband, a resource-efficient algorithm, is employed to explore the hyperparameter space. This process involves specifying parameters such as the maximum number of iterations and resource allocation to ensure a comprehensive search.

3.3.3. Model training process

For each model, the effects of different oversampling methods and rates on performance are investigated. Hyperband is employed to identify the optimal hyperparameters by training the models on data augmented using various oversampling techniques. Each configuration is evaluated using a validation set, allowing us to identify the best-performing setup based on validation results. This comprehensive approach ensures that each model is finely tuned to handle the nuances of the oversampled data, ultimately improving its ability to detect fraudulent transactions effectively.

3.4. Model evaluation

In this stage, the effectiveness of each trained model in detecting credit card fraud is assessed by evaluating their performance on a separate testing set. Once the models have been trained with the optimal hyperparameters, their performance is evaluated: i) each model is tested on a dedicated testing set to ensure that the evaluation reflects real-world performance and is not biased by the training data; and ii) for each combination of model, oversampling method, and rate, the F1-score is calculated. The F1-score is a crucial metric in fraud detection as it balances precision and recall, providing a comprehensive measure of a model's accuracy in identifying fraudulent transactions while minimizing false positives and negatives. This evaluation process helps determine how well each model can generalize to new, unseen data and highlights the most effective strategies for detecting fraud.

3.5. Selection and reporting of best model-method-rate combinations

In this final stage, the most effective configurations for detecting credit card fraud are identified and reported. The performance of the optimized models is compared to determine which configurations best balance fraud detection and false alarm reduction. The results are then summarized, and the configurations that offer the strongest overall performance are highlighted.

3.5.1. Determine best combination

For each combination of model and oversampling method, the results are analyzed to identify the oversampling rate and hyperparameter configuration that achieve the highest F1-score. The optimal setup for each model-method pair is selected based on the F1-score, which captures the balance between precision and recall. The best F1-score is then recorded alongside the corresponding oversampling rate and hyperparameters, providing a clear summary of the most effective configurations.

3.5.2. Output results

The results for each model-method combination are compiled and presented, providing a comprehensive overview of the best configurations: the optimal oversampling rate and hyperparameters are reported along with key performance metrics, such as accuracy, precision, recall, area under the curve (AUC), Matthews correlation coefficient (MCC), and F1-score. This detailed reporting not only highlights the most effective strategies for fraud detection but also provides valuable insights into model performance across various metrics, facilitating informed decision-making for future implementations.

4. RESULTS AND DISCUSSION

This section presents and discusses the experimental results by emphasizing key performance trends and comparative insights across models, oversampling strategies, and datasets. To improve clarity and

interpretability, detailed numerical results are summarized in Tables 2 and 3 to provide a consolidated view of the metrics. The discussion therefore focuses on the most salient findings that highlight how different techniques influence model behavior under varying data imbalance conditions.

On the PaySim dataset as shown in Table 2, the CNN model achieved its best performance with SMOTE at a relatively low oversampling rate (0.10), reaching an F1-score of 0.9885. This indicates that moderate rebalancing is sufficient for CNNs on synthetic transaction data. However, CNNGAN at the same oversampling rate achieved comparable performance with higher precision, highlighting its ability to generate higher-quality minority samples. In contrast, aggressive oversampling with DNNGAN (e.g., 0.90) resulted in degraded performance, suggesting that excessive synthetic data can negatively affect generalization. Recurrent models exhibited a different behavior. Both LSTM and GRU achieved their highest F1-scores when paired with CNNGAN at intermediate oversampling rates (0.60–0.65). This trend suggests that temporal models benefit from richer and more diverse synthetic patterns, but only up to a certain threshold. A comparative bar chart of F1-scores across oversampling rates further illustrates this non-linear relationship, showing performance peaks at moderate levels rather than at full class balance.

Table 2. The detailed results of different models on the PaySim dataset

Model type	Oversampling method	Oversampling rate	Accuracy	Precision	Recall	F1-score	AUC score	MCC
CNN	SMOTE	0.10	0.9998	0.9921	**0.9850**	**0.9885**	0.9998	0.9803
	CNNGAN	0.10	**0.9999**	**0.9992**	0.9750	**0.9869**	**0.9999**	**0.9807**
	DNNGAN	0.90	0.9995	0.9853	0.9600	0.9724	0.9990	0.9509
LSTM	SMOTE	0.15	0.9996	0.9505	0.9900	0.9700	0.9995	0.9604
	CNNGAN	0.60	0.9997	0.9980	0.9503	0.9734	0.9996	0.9708
	DNNGAN	0.55	0.9996	0.9855	0.9501	0.9674	0.9992	0.9509
GRU	SMOTE	0.10	0.9997	0.9857	0.9505	0.9676	0.9993	0.9510
	DNNGAN	0.65	0.9998	0.9985	0.9504	0.9739	0.9997	0.9711
	CNNGAN	0.10	0.9997	0.9859	0.9503	0.9678	0.9994	0.9512

The European dataset, as shown in Table 3, which reflects real-world credit card transactions, exhibits higher overall performance across all models. This dataset further confirms the superiority of CNNGAN as an oversampling strategy. The CNN combined with CNNGAN at a low oversampling rate (0.15) achieved the highest F1-score of 0.9972, along with near-perfect recall. This result is particularly important in fraud detection, where minimizing false negatives is critical. LSTM models demonstrated robust and stable performance across all oversampling techniques, with SMOTE at a 0.50 rate yielding an F1-score of 0.9954. However, GAN-based approaches provided marginal but consistent improvements. The GRU model showed similar behavior, achieving its best performance with CNNGAN at a 0.30 rate. Radar plots comparing precision, recall, F1-score, and MCC across models visually confirm that CNNGAN-based configurations provide a more balanced trade-off between detection accuracy and robustness.

Table 3. The detailed results of different models on the European dataset

Model type	Oversampling method	Oversampling rate	Accuracy	Precision	Recall	F1-score	AUC score	MCC
CNN	SMOTE	0.10	0.9930	0.9950	0.9935	0.9942	0.9970	0.9880
	CNNGAN	0.15	0.9975	**0.9960**	0.9985	**0.9972**	0.9990	**0.9910**
	DNNGAN	0.65	0.9950	0.9930	0.9955	0.9942	0.9980	0.9850
LSTM	SMOTE	0.50	**0.9980**	0.9920	**0.9988**	0.9954	0.9990	0.9900
	CNNGAN	0.60	0.9965	0.9925	0.9975	0.9950	0.9985	0.9880
	DNNGAN	0.10	0.9940	0.9915	0.9950	0.9932	0.9975	0.9855
GRU	SMOTE	0.35	0.9920	0.9890	0.9930	0.9910	0.9950	0.9825
	DNNGAN	0.15	0.9950	0.9935	0.9960	0.9947	0.9990	0.9875
	CNNGAN	0.30	0.9960	0.9950	0.9970	0.9960	**0.9995**	0.9900

Across both the PaySim and European datasets, the results clearly demonstrate that the choice of oversampling technique and its associated rate has a substantial impact on model effectiveness, particularly in terms of F1-score and MCC. Overall, GAN-based oversampling methods consistently outperform the classical SMOTE approach, especially when combined with convolutional architectures.

When comparing both datasets, three main observations emerge, as illustrated in Figure 2, i) CNNGAN consistently delivers the most reliable performance gains, particularly for CNN architectures, ii) optimal oversampling rates are dataset- and model-dependent, with lower rates favored for CNNs and moderate rates for recurrent models, and iii) over-aggressive oversampling degrades performance, reinforcing the need for dynamic and controlled rebalancing strategies.

Hyperparameter tuning using the hyperband strategy resulted in an optimized CNNGAN architecture. The generator employs lightweight Conv1D layers to capture local feature dependencies, while the discriminator uses deeper convolutional layers to improve adversarial learning stability. The CNN classifier integrates dropout regularization to mitigate overfitting. This configuration achieves an effective balance between model complexity and generalization, as reflected in the consistently high F1-scores and MCC values across datasets.

Table 4 compares various models for credit card fraud detection, with CNNGAN combined with a CNN classifier emerging as the best due to its high recall and balanced metrics. Here's a concise analysis including our method:

- Tang and Liang [21] achieve 99.83% accuracy and 100% recall, effectively identifying all fraudulent transactions.
- Rojas et al. [30] reported perfect scores across all metrics, which may suggest overfitting.
- This study delivers strong performance with 99.75% accuracy, 99.85% recall, and a high MCC of 99.00%, indicating a well-balanced model that effectively detects fraud while minimizing false positives and negatives.
- Other studies show varied results, with some models achieving high precision but lower recall, highlighting trade-offs between detecting fraud and minimizing false alarms.

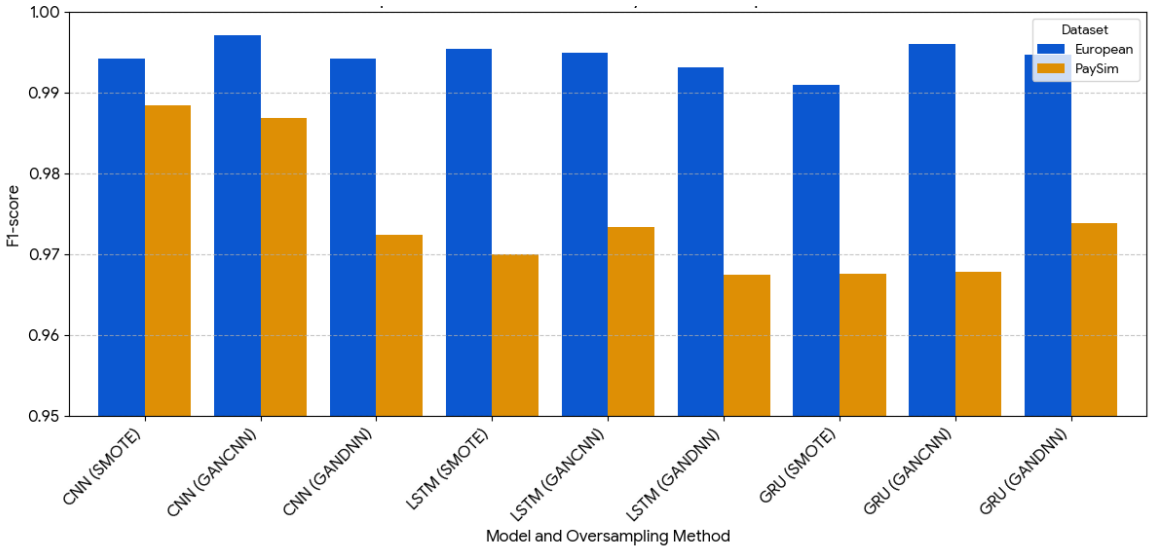

Figure 2. Comparison of model F1-Scores by oversampling method across PaySim and European datasets

Table 4. Comparison with recent papers

Ref.	Year	Dataset	Evaluation metrics used	Best results
Roseline *et al.* [19]	2024	Credit card European dataset	Accuracy, recall, precision, F1-score	Accuracy 99.83%, recall 100%, precision 99.67%, F1-score 99.83%
Reddy *et al.* [20]	2024		Accuracy	Accuracy 89.10%
Tang and Liang [21]	2022		Accuracy, precision, recall, F1-score	Accuracy 1, precision 1, recall 1, F1-score 1
Xie and Huang [22]	2024		Accuracy, recall, precision, F1-score, AUC	Accuracy 95.99%, recall 91.10%, precision 93.81%, AUC 94.03%
Ning *et al.* [24]	2022		Accuracy, recall, precision, F1-score	Accuracy: 99.98%, precision: 95.34% ,recall: 77.87%,F1-score: 82.85%
Ding *et al.* [25]	2023		Recall, precision, F1-score, AUC	Precision: 0.9404, recall: 0.8323, F1-score: 0.8805, AUC score: 0.98506
Alarfaj and Shahzadi [26]	2023		Recall, accuracy, F1-score, AUC	Accuracy: 0.9993, recall: 0.8299, F1-score: 0.8187, AUC score: 0.9147
Zhao and Guan [27]	2023		Precision, recall	Precision: 0.9560, recall: 0.8053
Leevy *et al.* [28]	2023		Accuracy, recall, precision, F1-score, AUC	Accuracy: 1.0, precision: 1.0, recall: 1.0, F1-score: 1.0, AUC score: 1.0
Ndama *et al.* [29]	2022		Accuracy, recall, precision, F1-score	Accuracy: 78.06% ,precision: 81.63% ,recall: 80.62%,F1-score: 81.36%
This Study	2025		Accuracy, recall, precision, F1-score, AUC, MCC	Accuracy: 99.75% ,precision: 99.20% ,recall: 99.85%, F1-score: 99.72%, AUC:99.90%, MCC: 99.00%

Overall, the CNNGAN with a CNN classifier and the proposed method both demonstrate excellent capabilities in fraud detection, with the proposed method providing a robust balance of accuracy and recall. To strengthen the interpretation of these findings, a component-wise analysis was conducted by comparing SMOTE, CNNGAN, and DNNGAN across multiple oversampling rates and classifier architectures, while the results were also benchmarked against recent studies in Table 4 to clarify the contribution and relative effectiveness of each element of the proposed framework.

5. CONCLUSION

This paper proposes a novel credit card fraud detection system that leverages dynamic oversampling and GANs to address the challenges of imbalanced datasets and overlapping transaction patterns. By adjusting the oversampling rate dynamically and using advanced techniques like SMOTE, CNN-GANs, and DNN-GANs, the system effectively balances the dataset and generates synthetic fraud samples representative of real-world data. Critically, the evaluation is performed on a held-out, non-oversampled dataset, ensuring a realistic assessment of the model's performance and mitigating the risk of overfitting. Experimental results on both the PaySim synthetic dataset and a real-world European credit card transaction dataset demonstrate that the proposed approach, particularly using CNN-GANs with a CNN classifier, achieves state-of-the-art performance in terms of F1-score, accuracy, and recall, outperforming several existing methods. The flexible oversampling rate and the use of advanced GANs for data augmentation contribute to the improved fraud detection capabilities. Future research could explore the application of this approach to other imbalanced datasets and investigate further refinements to the GAN architectures for enhanced synthetic data generation.

FUNDING INFORMATION

The authors declare that this research received no external funding.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Mohamed Akouhar	✓	✓		✓	✓	✓			✓	✓	✓			
Abdallah Abarda				✓		✓				✓		✓	✓	
Mohamed El Fatini				✓	✓					✓		✓		
Mohamed Ouhssini	✓	✓							✓	✓				
El Hafed Agherrabi				✓		✓				✓				

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY

The datasets used in this study are publicly available. Their sources are described in the article and cited in references [30] and [31].

REFERENCES

[1] European Central Bank, *Seventh Report on Card Fraud*. Frankfurt am Main, Germany: European Central Bank, Oct. 2021.
[2] Board of Governors of the Federal Reserve System, *The 2019 Federal Reserve Payments Study*. Washington, DC, USA, 2019.
[3] N. F. R.-Tubb, P. Krause, and W. Garn, "How artificial intelligence and machine learning research impacts payment card fraud detection: a survey and industry benchmark," *Engineering Applications of Artificial Intelligence*, vol. 76, pp. 130–157, Nov. 2018, doi: 10.1016/j.engappai.2018.07.008.
[4] T. Pourhabibi, K. L. Ong, B. H. Kam, and Y. L. Boo, "Fraud detection: a systematic literature review of graph-based anomaly detection approaches," *Decision Support Systems*, vol. 133, 2020, doi: 10.1016/j.dss.2020.113303.
[5] K. G. Dastidar, J. Jurgovsky, W. Siblini, and M. Granitzer, "NAG: neural feature aggregation framework for credit card fraud detection," *Knowledge and Information Systems*, vol. 64, no. 3, pp. 831–858, Mar. 2022, doi: 10.1007/s10115-022-01653-0.

[6] Z. Li, M. Huang, G. Liu, and C. Jiang, "A hybrid method with dynamic weighted entropy for handling the problem of class imbalance with overlap in credit card fraud detection," *Expert Systems with Applications*, vol. 175, Aug. 2021, doi: 10.1016/j.eswa.2021.114750.
[7] A. D. Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "Credit card fraud detection: a realistic modeling and a novel learning strategy," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 8, pp. 3784–3797, Aug. 2018, doi: 10.1109/TNNLS.2017.2736643.
[8] B. Yousefimehr and M. Ghatee, "A systematic survey and empirical comparison of hybrid methods for imbalanced fraud detection: combining resampling and machine learning," *AUT Journal of Mathematics and Computing*, vol. 7, no. 1, pp. 85–116, 2026, doi: 10.22060/ajmc.2025.24642.1446.
[9] B. A. N.G. and S. S., "Anomaly detection framework for Internet of things traffic using vector convolutional deep learning approach in fog environment," *Future Generation Computer Systems*, vol. 113, pp. 255–265, 2020, doi: 10.1016/j.future.2020.07.020.
[10] K. Fu, D. Cheng, Y. Tu, and L. Zhang, "Credit card fraud detection using convolutional neural networks," *International Conference on Neural Information Processing*, 2016, pp. 483–490, doi: 10.1007/978-3-319-46675-0_53.
[11] J. Jurgovsky *et al.*, "Sequence classification for credit-card fraud detection," *Expert Systems with Applications*, vol. 100, pp. 234–245, Jun. 2018, doi: 10.1016/j.eswa.2018.01.037.
[12] A. Roy, J. Sun, R. Mahoney, L. Alonzi, S. Adams, and P. Beling, "Deep learning detecting fraud in credit card transactions," in *2018 Systems and Information Engineering Design Symposium (SIEDS)*, 2018, pp. 129–134, doi: 10.1109/SIEDS.2018.8374722.
[13] D. Zhang, J. Yin, X. Zhu, and C. Zhang, "Network representation learning: a survey," *IEEE Transactions on Big Data*, vol. 6, no. 1, pp. 3–28, 2020, doi: 10.1109/TBDATA.2018.2850013.
[14] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*, Longeach, CA, USA, 2017, pp. 1025–1035, doi: 10.5555/3294771.3294869.
[15] A. Pumsirirat and L. Yan, "Credit card fraud detection using deep learning based on auto-encoder and restricted Boltzmann machine," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 1, 2018, doi: 10.14569/IJACSA.2018.090103.
[16] T. Suzumura *et al.*, "Towards federated graph learning for collaborative financial crimes detection," 2019, *arXiv: 1909.12946*.
[17] F. Carcillo, Y.-A. Le Borgne, O. Caelen, Y. Kessaci, F. Oblé, and G. Bontempi, "Combining unsupervised and supervised learning in credit card fraud detection," *Information Sciences*, vol. 557, pp. 317–331, May 2021, doi: 10.1016/j.ins.2019.05.042.
[18] K. Razzaq and M. Shah, "Next-generation machine learning in healthcare fraud detection: current trends, challenges, and future research directions," *Information*, vol. 16, no. 9, Aug. 2025, doi: 10.3390/info16090730.
[19] J. F. Roseline, G. Naidu, V. S. Pandi, S. A. Rajasree, and D. N. Mageswari, "Autonomous credit card fraud detection using machine learning approach☆," *Computers and Electrical Engineering*, vol. 102, 2022, doi: 10.1016/j.compeleceng.2022.108132.
[20] V. V. K. Reddy, R. V. K. Reddy, M. S. K. Munaga, B. Karnam, S. K. Maddila, and C. S. Kolli, "Deep learning-based credit card fraud detection in federated learning," *Expert Systems with Applications*, vol. 255, Dec. 2024, doi: 10.1016/j.eswa.2024.124493.
[21] Y. Tang and Y. Liang, "Credit card fraud detection based on federated graph learning," *Expert Systems with Applications*, vol. 256, Dec. 2024, doi: 10.1016/j.eswa.2024.124979.
[22] Z. Xie and X. Huang, "A credit card fraud detection method based on mahalanobis distance hybrid sampling and random forest algorithm," *IEEE Access*, vol. 12, pp. 162788–162798, 2024, doi: 10.1109/ACCESS.2024.3421316.
[23] E. Ileberi, Y. Sun, and Z. Wang, "A machine learning based credit card fraud detection using the GA algorithm for feature selection," *Journal of Big Data*, vol. 9, no. 1, Dec. 2022, doi: 10.1186/s40537-022-00573-8.
[24] W. Ning, S. Chen, S. Lei, and X. Liao, "AMWSPLAdaboost credit card fraud detection method based on enhanced base classifier diversity," *IEEE Access*, vol. 11, pp. 66488–66496, 2023, doi: 10.1109/ACCESS.2023.3290957.
[25] Y. Ding, W. Kang, J. Feng, B. Peng, and A. Yang, "Credit card fraud detection based on improved variational autoencoder generative adversarial network," *IEEE Access*, vol. 11, pp. 83680–83691, 2023, doi: 10.1109/ACCESS.2023.3302339.
[26] F. K. Alarfaj and S. Shahzadi, "Enhancing fraud detection in banking with deep learning: graph neural networks and autoencoders for real-time credit card fraud prevention," *IEEE Access*, vol. 13, pp. 20633–20646, 2025, doi: 10.1109/ACCESS.2024.3466288.
[27] X. Zhao and S. Guan, "CTCN: a novel credit card fraud detection method based on conditional tabular generative adversarial networks and temporal convolutional network," *PeerJ Computer Science*, vol. 9, Oct. 2023, doi: 10.7717/peerj-cs.1634.
[28] J. L. Leevy, J. Hancock, T. M. Khoshgoftaar, and A. A. Zadeh, "Investigating the effectiveness of one-class and binary classification for fraud detection," *Journal of Big Data*, vol. 10, Art. no. 157, Oct. 2023, doi: 10.1186/s40537-023-00825-1.
[29] O. Ndama, I. Bensassi, and E. M. En-Naimi, "Optimizing credit card fraud detection: a deep learning approach to imbalanced datasets," *International Journal of Electrical and Computer Engineering*, vol. 14, no. 4, pp. 4802–4814, Aug. 2024, doi: 10.11591/ijece.v14i4.pp4802-4814.
[30] E. A. L.-Rojas, A. Elmir, and S. Axelsson, "PaySim: a financial mobile money simulator for fraud detection," in *Proc. 28th European Modeling and Simulation Symposium (EMSS 2016)*, Larnaca, Cyprus, 2016, pp. 249–255.
[31] A. Dal Pozzolo, O. Caelen, Y.-A. Le Borgne, S. Waterschoot, and G. Bontempi, "Learned lessons in credit card fraud detection from a practitioner perspective," *Expert Systems with Applications*, vol. 41, no. 10, pp. 4915–4928, Aug. 2014, doi: 10.1016/j.eswa.2014.02.026.

BIOGRAPHIES OF AUTHORS



Mohamed Akouhar earned his master's degree in mathematics from Ibn Tofail University in Kenitra and his Ph.D. in 2026, specializing in statistics and artificial intelligence. His research interests include data science, machine learning, supervised and unsupervised learning, classification, neural networks, statistical learning, feature selection, and feature extraction. His publications explore deep learning applications in cybersecurity and machine learning techniques for fraud detection. He can be contacted at email: m.akouhar@gmail.com.



Abdallah Abarda earned his Ph.D. in Statistics and Data Analysis from Ibn Tofail University, Morocco, in 2017. He is a full professor at Hassan First University in Settat, Morocco, where he has been teaching since April 2018. His research expertise includes data analysis, R programming, data clustering, and multivariate statistics. He has significantly contributed to various fields through his publications, focusing on statistical methods for big data, machine learning applications in health, and techniques for cybersecurity. He can be contacted at email: abardabdallah@gmail.com.



Mohamed El Fatini is a prominent mathematician at Ibn Tofail University in Kenitra, Morocco, specializing in stochastic modeling and its applications in epidemiology. His influential research includes notable publications on epidemic models and stochastic control in healthcare, with over 800 citations. He is work focuses on developing mathematical frameworks for understanding disease spread and optimizing treatment strategies. He is actively involved in academic journals and conferences, contributing significantly to the field of mathematical epidemiology. He can be contacted at email: mohamed.elfatini@uit.ac.ma.



Mohamed Ouhssini received his master's degree in Information Technology Offshoring in 2017 from the Faculty of Science at Ibnou Zohr University, Agadir, Morocco. He earned his Ph.D. in 2024, specializing in computer security and artificial intelligence, and is currently a professor at Hassan II University of Casablanca, Morocco. His research interests include computer security, deep learning, machine learning, and cloud computing. He can be contacted at email: mohamed.ouhssini@gmail.com.



El Hafed Agherrabi received his master's degree in offshoring of Information Technologies in 2017 from the Faculty of Science, University Ibnou Zohr, Agadir, Morocco. He is currently pursuing a Ph.D. in Network Security at Hassan First University, Settat, Morocco. His research interests include computer security, deep learning, machine learning, and cloud computing. He can be contacted at email: hafed.agherrabi@gmail.com.