JCTA Journal of Computing Theories and Applications ISSN:3024-9104

Research Article

Improving Credit Card Fraud Detection with Ensemble Deep Learning-Based Models: A Hybrid Approach Using SMOTE-ENN

Lossan Bonde [1,*] and Abdoul Karim Bichanga [2]

[1] Department of Applied Sciences, Adventist University of Africa; Nairobi, Kenya; e-mail: bondel@aua.ac.ke
[2] Ecole Supérieure d'Informatique, Université Nazi Boni; Bobo-Dioulasso, Burkina Faso;
e-mail: abdoulbichanga@gmail.com
* Corresponding Author : Lossan Bonde

Abstract: Advances in information and internet technologies have significantly transformed the business environment, including the financial sector. The COVID-19 pandemic has further accelerated this digital adoption, expanding the e-commerce industry and highlighting the necessity for secure online transactions. Credit Card Fraud Detection (CCFD) stands critical as the prevalence of fraudulent activities continues to rise with the increasing volume of online transactions. Traditional methods for detecting fraud, such as rule-based systems and basic machine learning models, tend to fail to keep pace with fraudsters' evolving tactics. This study proposes a novel ensemble deep learning-based approach that combines Convolutional Neural Networks (CNN), Gated Recurrent Units (GRU), and Multilayer Perceptron (MLP) with the Synthetic Minority Oversampling Technique and Edited Nearest Neighbors (SMOTE-ENN) to address class imbalance and improve detection accuracy. The methodology integrates CNN for feature extraction, GRU for sequential transaction analysis, and Multilayer Perceptron (MLP) as a meta-learner in a stacking framework. By leveraging SMOTE-ENN, the proposed approach enhances data balance and prevents overfitting. With synthetic data, the robustness and accuracy of the model have been improved, particularly in scenarios where fraudulent examples are scarce. The experiments conducted on real-world credit card transaction datasets have established that our approach outperforms existing methods, achieving higher metrics performance.

Keywords: Credit card frauds detection; Credit card transaction datasets; Deep learning-based ensemble models; Imbalanced datasets; Synthetic minority over-sampling technique with edited nearest neighbors.

Received: December, 29th 2024
Revised: February, 9th 2025
Accepted: February, 10th 2025
Published: February, 12th 2025

Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) licenses (https://creativecommons.org/licenses/by/4.0/)

1. Introduction

Advances in information and internet technologies have significantly transformed the business environment, including the financial sector. The COVID-19 pandemic has further accelerated this digital shift, expanding the e-commerce industry and highlighting the necessity for secure online transactions[1]. This rapid digitalization has created significant security challenges, particularly in credit card fraud. Credit cards are currently the primary method for online shopping and card-not-present transactions, making them particularly susceptible to fraud[2]. Globally, transaction fraud has risen significantly, causing substantial financial losses and undermining consumer confidence in digital payment systems[3]. Credit card fraud detection (CCFD) is crucial to financial security as fraud increases with the increasing volume of online transactions. The main difficulty comes from the always-changing nature of the operations performed by criminals, which classical methods, such as rule-based systems and simple machine learning, have not been able to handle successfully. While deep learning (DL) is extensively leveraged in fields like computer vision and natural language processing, its application in credit card fraud detection remains limited[4]. Traditional models have considerable handicaps in modeling complex and dynamic patterns of fraud.

DOI : 10.62411/jcta.12021

In comparison, DL models, like Convolutional Neural Networks (CNNs), perform excellently for automatic feature extraction, and Gated Recurrent Units (GRUs) are suitable for processing sequential data such as transaction history[5]. In this domain, a significant issue is a class imbalance, where the number of fraudulent transactions is many folds lower than that of legitimate transactions, distorting the model's performance indicators [6]. As a result, the serious imbalance causes a huge degradation in identifying the minority class (fraudulent transactions) due to the overwhelming predominance of the majority class (legitimate transactions) during the model training process.

In this research, the SMOTE-ENN approach is used. SMOTE-ENN is a compounding resampling method that merges the Synthetic Minority Over-sampling Technique with Edited Nearest Neighbors to deal with the issue of data trapped within the class imbalance. This approach has demonstrated greater effectiveness in handling imbalanced data than traditional oversampling and under-sampling techniques[7], [8]. We aim to enhance feature extraction and classification processes by integrating CNN and GRU models. Ensemble learning performs training of multiple base classifiers and combines their outputs to achieve higher performance than single classifiers[9]. Consequently, ensemble learning-based classifiers mostly outperform equivalent single classifiers[10], [11]. However, existing literature shows few applications of deep learning-based ensemble models for CCFD. To narrow this research gap while addressing the challenges facing CCFD, this research uses CNN and GRU networks as base learners within a stacking ensemble model. While most ensemble approaches for CCFD employ voting-based methods, this study uses the MLP neural network as the meta-learner in the stacking ensemble, following the approach suggested by Mienye[7]. The proposed approach leverages the advantages of sequential modeling and ensemble learning and enhances CCFD.

This study proposes a novel approach to CCFD by integrating CNN and GRU models, addressing the class imbalance problem, and enhancing feature extraction and classification processes. The key contributions of this research include:

- Improved Model: Develop a CCFD model using CNN for feature extraction, GRU for learning transaction sequences, and MLP as a meta-learner in an ensemble model.
- Class Imbalance: Address class imbalance with SMOTE-ENN resampling and advanced feature selection techniques.
- Enhanced Detection Rate: Combine ensemble and deep learning for robust fraud detection.

The remaining part of the paper is organized into five sections. Section 2 explores related works and identifies the research gaps this paper addresses. Section 3 describes the work's methodology, including the results' validation process. Section 4 presents the experiment and its setup, and then section 5 discusses the results. Finally, we conclude the paper in section 6 by summarizing the findings and contribution and suggesting areas of future research.

2. Related Work

The CCFD field has witnessed significant advancements in adopting machine learning (ML) and DL techniques. Early approaches primarily focused on classical ML algorithms. However, recent studies have shown that deep learning architectures, especially CNNs, Recurrent Neural Networks (RNNs), and ensemble methods, perform better in detecting fraudulent transactions.

Several traditional ML techniques have been applied to credit card fraud detection with varying degrees of success. For instance, Vardhani et al. proposed a Condensed nearest-neighbor algorithm to reduce computational complexity during fraud detection tasks [12]. This non-parametric method condenses the dataset to improve query times and memory usage, particularly useful in distributed data mining applications. The study demonstrated the algorithm's potential to reduce computational overhead without compromising detection accuracy. Additionally, Vardhani et al. [12] compared various ML algorithms for credit card fraud detection, including Condensed Nearest Neighbor. While Condensed Nearest Neighbor outperformed other algorithms in terms of accuracy, the authors pointed out challenges related to the speed of feature extraction and model testing, emphasizing the need for continuous improvement in detection models.

Using CNNs in fraud detection has proven effective in capturing complex transaction patterns. Nalayini et al. introduced a three-layered CNN model with a smart matrix algorithm[13]. This model utilized random under-sampling and normalization techniques to improve dataset preprocessing and training efficiency. The results showed that CNN outperformed ML algorithms like Naive Bayes and K-Nearest Neighbours (KNN), especially in handling large datasets and real-time applications. Fu et al. [14] investigated CNN-based models in their research, introducing a framework that transformed transaction data into a feature matrix to identify hidden patterns of fraudulent activity. Their model outperformed conventional fraud detection techniques when utilized on actual datasets. By expanding on these results, Zhang et al. [15] created a CNN model to process low-dimensional, non-derivative online transaction data. By restructuring raw transaction attributes into convolutional patterns, the model attained excellent precision and recall, further confirming CNNs as valuable instruments for detecting fraud.

Detecting fraud is difficult due to class imbalance between the extremely few fraudulent transactions and the bulk of legitimate transactions. This has stimulated interest in sampling methods, particularly the Synthetic Minority Oversampling Technique (SMOTE) and its variations. Sisodia et al. proved that the combination of SMOTE and Edited Nearest Neighbors (SMOTE-ENN) performs better than other resampling techniques in detecting fraud[15]. Esenogho et al.[10] combined SMOTE-ENN with Long Short-Term Memory (LSTM) networks and achieved 99.6% sensitivity and 99.8% specificity by oversampling with error-correcting sampling to prevent overfitting. Akazue et al. [16] proposed an ensemble method for feature selection combining recursive feature elimination, information gain, and Chi-squared methods using a random forest algorithm, thus achieving 99.6%-F1 scores and 100% precision. Finally, Mienye and Sun presented a stacked ensemble model using LSTM and GRU with 100% sensitivity and 99.7% specificity, advocating for treating class imbalance in fraud detection[7].

In a time when deep learning models are becoming more capable of articulating temporal dynamics and intra-transaction interactions, Li et al. [17] proposed a sandwich-structured model based on GRU, combining ensemble models, deep sequential learnings, and attention mechanisms in it, thereby identifying complex transaction patterns. A survey by Arora et al. [4] on machine learning algorithms used for fraud detection mentions that hybrid models and deep learning methods performed the best on larger datasets. Setiadi et al. [18] developed a bidirectional GRU and feature selection for phishing website detection. Çetin and Öztürk [19] explored and developed an ensemble learning model for IoT cybersecurity across multi-class and binary classification tasks. Dhahir et al. [20] combined the CBLOF with XGBoost to increase DDoS detection accuracy up to 99.99%.

Finally, Transformer's attention has been turned towards its newfound ability to engage sequential data efficiently. Iqbal and Amin [21] applied a Transformer-based model to the European Credit Card dataset, which yielded an incredible 100% accuracy, sensitivity, and specificity, with an AUC of 100%. Though Transformer models stand out for their consummate performance, they call for computational requirements, demonstrating the tradeoff between accuracy and the underlying efficiency issues. Pathirana et al.[22] explored a related approach in mental health applications, demonstrating the potential of reinforcement learning (RL) combined with multimodal emotion recognition for personalized interventions. Although the study was centered on mental health, its multimodal RL framework offers promising avenues for fraud detection research.

Credit card fraud detection is accelerating based on the integration of machine learning and deep learning techniques. CNNs, RNNs, and advanced techniques like SMOTE-ENN have improved model performance. However, a gap exists in applying deep learning-based ensemble models, particularly combining CNNs and GRUs, for real-time fraud detection. Due attention has not been given to hybrid models integrating these techniques and ensemble learning in the context of credit card fraud detection. It is worth mentioning that the integration of MLP networks as meta-learners has not yet been formulated. In future works, such deep learning models will be a priority for further developments in class imbalance handling and computational efficiency optimization to outpace increasingly sophisticated fraudulent activities [23].

3. Proposed Method

The suggested CCFD approach employs a strong stacking-based ensemble model that combines CNN, MLP, and GRU neural networks. The stacking architecture constitutes two levels, Level 0 and 1. At Level 0, the base classifiers generate predictions following training and assessment of data points out of the sample. The meta-classifier at level 1 is trained using a new dataset comprising these predictions and the real labels[24].

The MLP is the meta-learner at level 1 in this framework, whereas GRU and CNN are the base learners at level 0. The integration of GRU and CNN is motivated by their respective advantages: CNN efficiently captures spatial information, whereas GRU is excellent at handling sequential data (like transaction history). Their unique design adds variety to the group, essential for raising performance levels. The ensemble is more robust because multiple base models are more likely to make different kinds of mistakes. Figure 1 presents the suggested methodology.

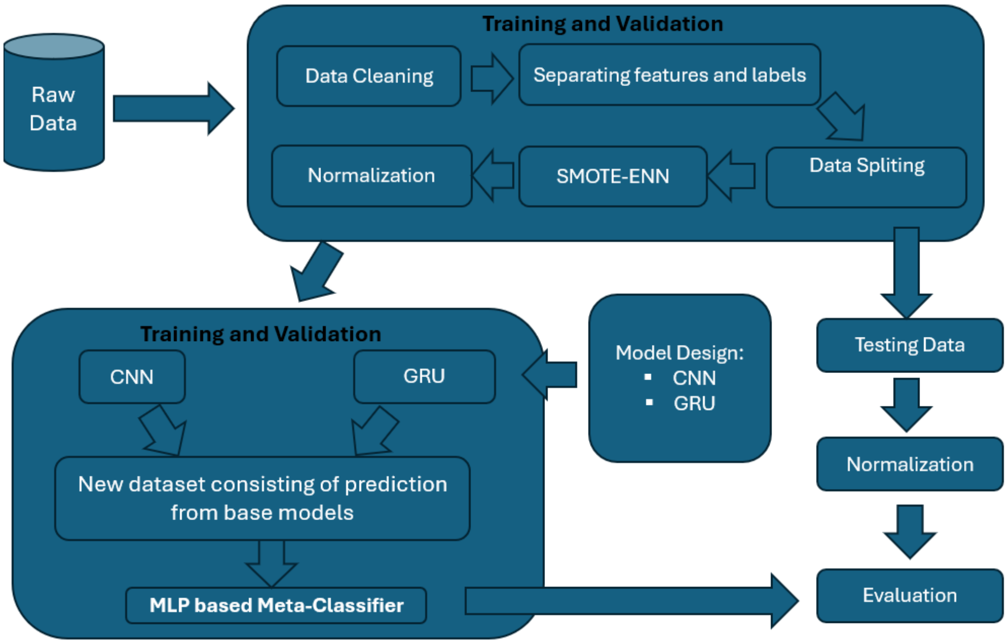

Figure 1. Proposed Deep Learning-based Ensemble Model

3.1. Dataset

This research applies the CCFD dataset [25], which comprises credit card transactions from European cardholders from two days in September 2013. It possesses a high degree of imbalance, where only 492 fraudulent transactions are part of 284,807. Owing to its largeness, real-world availability, and frequent use for assessing fraud detection algorithms, it is an apt candidate for benchmarking purposes. The diversity of the dataset also contributes to realistic fraud detection challenges.

The distributions before and after applying SMOTE-ENN are shown in Figure 2. Initially, there were 284,315 non-fraudulent transactions, and only 492 were fraudulent. After applying SMOTE-ENN, the distributions become more balanced, with 265,395 non-fraudulent and 275,740 fraudulent transactions, thus reducing bias and favoring better fraud detection with more balanced datasets.

The dataset is publicly available at https://www.kaggle.com/mlg-ulb/creditcardfraud from which one can reproduce and compare the research with other works. Preprocessing steps were performed, where attributes except for 'Time' and 'Amount' were transformed into numerical features (V1 to V28) for privacy. 'Amount' denotes the transaction value, 'Time' denotes the duration since the first transaction, and 'Class' determines whether a given transaction is fraudulent (1) or legitimate (0).

The extremely low number of only 0.172% of fraudulent samples leads to the dramatic class imbalance that affects model generalization. Oversampling or undersampling methods are often used to deal with this [26], [27]. Since oversampling could lead to overfitting and undersampling could lead to some useful data being lost, we will use SMOTE-ENN, a combination of both methods. SMOTE generates synthetic instances in the minority class[28], while ENN discards overlapping samples based on some neighborhood cleaning rule[29]. The SMOTE-ENN algorithm details can be found in[7].

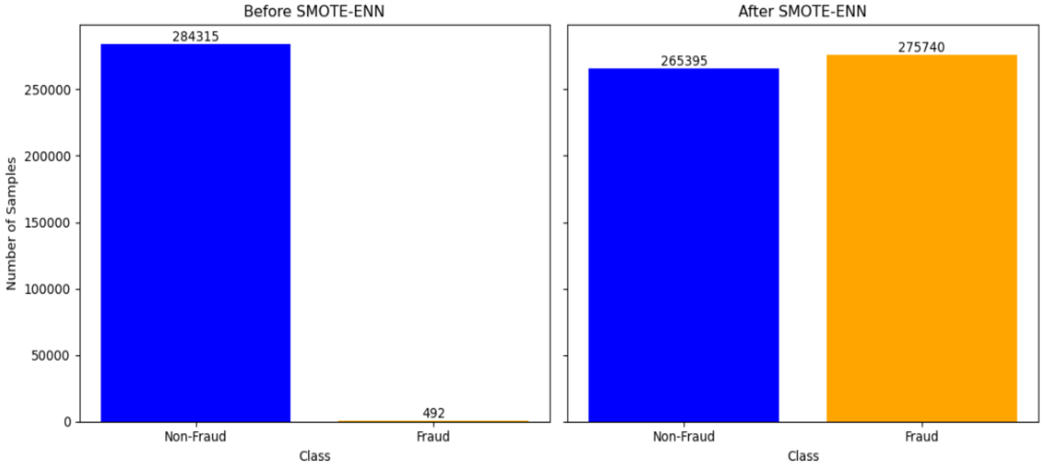

Figure 2. Dataset Distribution Before and After Applying SMOTE-ENN

3.2. Deep Learning Ensemble

A strong ensemble model based on stacking, comprising GRU, CNN, and MLP neural networks, is used in the suggested CCFD technique. The stacking architecture consists of two levels: Level 0 and Level 1. Base models such as GRU and CNN produce initial predictions at the first level. In contrast, Level 1 employs a meta-classifier known as MLP, which enhances these predictions by using a novel dataset created from the outputs of the base models alongside their corresponding true labels[24]. The GRU model excels at recognizing sequential patterns, while the CNN is adept at extracting spatial features. This combination reduces complementary errors within the ensemble, thereby enhancing robustness.

As depicted in Figure 1, this methodology follows three primary steps. Initially, both GRU and CNN are trained through cross-validation to generate out-of-sample forecasts. For each sample $(\hat{x}_i, y_i)$ represented as Equation (1).

$$\hat{x}_i = \{h_1(x_i), h_2(x_i)\} \tag{1}$$

Where h_1 and h_2 are the base models and represent their combined predictions.

Second, the MLP meta-classifier is trained on these predictions and their true labels, ensuring no overlap with the data used at Level 0 to prevent overfitting. For any test input x, the final ensemble prediction is calculated using Equation (2).

$$\hat{h}\big(h_1(x), h_2(x)\big) \tag{2}$$

Where $\hat{h}$ represents the meta-classifier.

Finally, the meta-classifier combines predictions from GRU and CNN to classify transactions as fraudulent (1) or legitimate (0).

Ultimately, this framework allows transactions to be classified as fraudulent (1) or legitimate (0). The stacking method boosts detection accuracy by harnessing GRU's ability to handle sequential information alongside CNN's expertise in analyzing spatial context with MLP to fine-tune all ultimate predictions.

To ensure this model effectively manages imbalanced datasets—adverse conditions often faced—it utilizes a preprocessing pipeline that incorporates SMOTE-ENN techniques for class balancing, data cleansing, and feature normalization processes[7], [30].

3.2.1. Preprocessing Techniques

The data processing included several vital procedures to set the dataset to train the model. First, missing values were checked, and no imputation was performed since none existed. The dataset was split into features (X) and the target variable, called "Class" (y), where "Class" indicates whether a transaction is legitimate or fraudulent. Due to considerable class imbalance, SMOTE-ENN was implemented, producing synthetic samples for the minority class: 265,395 non-fraudulent and 275,740 fraudulent transactions. Features were transformed into number representations to protect confidentiality, leaving the "Time" and "Amount" features untransformed. Features were normalized using StandardScaler, with mean = 0 and standard deviation = 1. Feature selection was done finally to remove redundant or irrelevant features to avoid overfitting and improve the model's generalization.

3.2.2. Class Balancing with SMOTE-ENN

To address the class imbalance, a preprocessing pipeline utilizes a hybrid method called SMOTE-ENN, a merger of SMOTE and ENN. SMOTE generates synthetic samples for the minority class by interpolation between existing samples. x_i and x_j in the process described in Equation (3)

$$x_{new} = x_i + \lambda(x_j - x_i) \tag{3}$$

Where x_i and x_j are two randomly selected samples from the minority class; λ is a random value drawn from a uniform distribution, $\lambda \in [0,1]$.

SMOTE increases the representation of the minority class by creating synthetic samples in feature space, thereby reducing overfitting with respect to random oversampling. ENN complements SMOTE by refining the dataset and addressing noisy or misclassified majority class samples. For each sample x_i in the majority class, ENN evaluates its classification using k k-nearest neighbors (commonly k=3). A majority class sample is removed if it is misclassified, as determined by the following condition, see Equation (4).

$$Class(x_i) \neq argmax_c \sum_{x_j \in N_k(x_i)} I(Class(x_j) = c) \tag{4}$$

Where $N_k(x_i)$ represents the set of k k-nearest neighbors of x_i; $I(\cdot)$ is the indicator function; $Class(x_j)$ is the class label of the neighbor x_j; and c is the candidate class label.

While traditional SMOTE focuses exclusively on oversampling the minority class, SMOTE-ENN introduces a two-step process incorporating data cleaning. This hybrid approach balances the dataset and improves its quality by reducing noise and eliminating overlaps between classes. Other variants of SMOTE, such as SMOTE-Tomek, also address noise but use different methodologies. For instance, Tomek links remove borderline samples between classes, whereas ENN specifically targets misclassified majority-class samples. This distinction makes SMOTE-ENN particularly effective in scenarios with significant noise and overlap.

3.2.3. Models Design

In the context of credit card fraud detection, the system is designed with a two-level architecture. At Level 0, two base models are employed: a CNN to extract spatial patterns from transaction data and a GRU to capture temporal dependencies within sequential transactions.

Table 1. CNN base model design.

No	Parameter	Value	Description
1	Input Shape	(X_train.shape[1], 1)	The input is reshaped into sequences of features with one channel.
2	Conv1D Filters	64	The number of convolutional layer filters.
3	Kernel Size	2	Size of the convolution kernel.
4	Activation Function	ReLU	Introduces non-linearity into the model.
5	Pooling Layer	MaxPooling1D (pool size=2)	Reduces the dimensionality of feature maps.
6	Dense Layer Neurons	128	Number of neurons in the fully connected layer.
7	Output Activation	Sigmoid	Outputs probabilities for binary classification.
8	Optimizer	Adam	Optimizer for weight updates during training.
9	Loss Function	Binary Cross-Entropy	Evaluates the difference between predictions and true labels.
10	Epochs	10	Number of complete training iterations.
11	Batch Size	64	Number of samples per batch

These models independently learn distinct features from the data. At Level 1, a Multi-layer Perceptron (MLP) acts as the meta-learner. It combines the predictions generated by CNN and GRU, leveraging their complementary strengths to optimize the final classification. The CNN is designed to extract spatial features from transaction data, capturing localized patterns indicative of fraud. Its architecture is summarized in Table 1. The output layer of CNN generates binary classification probabilities using a sigmoid activation function, while the convolutional and dense layers introduce non-linearity using the ReLU activation function. The model effectively captures spatial dependencies within the transaction features.

The GRU is designed to process sequential transaction data, capturing temporal patterns and dependencies. Its architecture is outlined in Table 2.

Table 2. GRU base model design.

No	Parameter	Value	Description
1	Input Shape	(X_train.shape[1], 1)	The input is reshaped for sequential processing.
2	GRU Units	32	Number of recurrent units in the GRU layer.
3	Dense Layer Neurons	64	Number of neurons in the dense layer.
4	Activation Function	ReLU	Adds non-linearity to the dense layer.
5	Output Activation	Sigmoid	Outputs probabilities for binary classification.
6	Optimizer	Adam	Optimizer for efficient gradient updates.
7	Loss Function	Binary Cross-Entropy	Evaluates the discrepancy between predictions and labels.
8	Epochs	10	Number of training cycles.
9	Batch Size	64	Number of samples processed per batch.

The GRU leverages its gating mechanism to efficiently learn temporal dependencies over transaction sequences, mitigating challenges associated with long-term dependencies in recurrent models.

The MLP acts as the meta-learner, combining CNN and GRU base model predictions. Its architecture is summarized in Table 3.

Table 3. MLP Meta Learner Design.

No	Parameter	Value	Description
1	Hidden Layer Sizes	(64, 32)	There are two hidden layers, each with 64 and 32 neurons.
2	Activation Function	ReLU	Used in the hidden layers to capture complex relationships.
3	Output Activation	Sigmoid	Produces final probabilities for binary classification.
4	Optimizer	Adam	Optimizer for minimizing the loss function.
5	Loss Function	Binary Cross-Entropy	Measures the error in classification tasks.
6	Epochs	100	Maximum number of training iterations.

By combining out-of-fold predictions from the CNN and GRU, the MLP learns to optimize the final classification, leveraging the unique strengths of the base models.

3.3. Validation of the Proposed Model

The model's capability to identify credit card fraud was determined by using multiple metrics: sensitivity, specificity, AUC-ROC, accuracy, precision and F1-Score. Sensitivity checks the proportion of frauds correctly classified as fraudulent transactions, while specificity checks on real transactions that are identified as real transactions. The precision ratio quantifies how many fraudulent transactions were accurately identified from the total number of transactions that were flagged as fraudulent. A high AUC-ROC denotes a model that effectively differentiates between genuine and fraudulent transactions while accuracy checks for the model's overall correctness. The F1-Score describes the harmonic mean of precision and recall, providing a balanced view of the model's performance, especially in imbalanced data sets.

4. Experiment and Results

4.1. Experimental Environment

The study proposes a deep learning ensemble technique that is coupled with data resampling to develop a trial for detecting fraud in the use of cards. The experimental findings are reported in two sections, one for the performance of the classifiers before and one for after data resampling. The suggested stacking ensemble employs a CNN and GRU neural network as the base learners, while an MLP neural network acts as the meta-learner. This technique is evaluated next to six other classifiers: Logistic Regression (LR), Random Forest (RF), MLP, CNN, GRU, and CNN with GRU combinations, respectively. The modeling for all classes has been performed in the Python environment with the scikit-learn library, while tests are based on a Windows 11-based computer with a Ryzen 5600H processor and 20 GB of RAM.

4.2. Results

The performance of various machine learning and deep learning models for credit card fraud detection was evaluated using a comprehensive set of metrics, including accuracy, specificity, sensitivity, precision, F1-Score, and AUC-ROC. The models assessed include LR, RF, ML, CNN, GRU, a hybrid CNN + GRU model, and an ensemble model combining CNN, GRU, and MLP. The evaluation was conducted before and after applying the SMOTE-ENN technique to address the class imbalance in the dataset. The results are presented in Table 4 and Table 5, and the performance of the ensemble model is further illustrated in Figure 3.

4.2.1. Model Performance Before SMOTE-ENN

Before applying SMOTE-ENN, the dataset exhibited significant class imbalance, with fraudulent transactions representing a small fraction of the total data. This imbalance adversely affected the models' performance, particularly regarding sensitivity. The results are summarized in Table 4.

Table 4. Performance of the classifiers before SMOTE-ENN.

Models	Accuracy	Specificity	Sensitivity	Precision	F1-Score	AUC-ROC
LR	0.923	0.981	0.684	0.845	0.789	0.833
RF	0.938	0.975	0.742	0.873	0.824	0.859
MLP	0.945	0.972	0.798	0.889	0.864	0.885
CNN	0.942	0.969	0.782	0.876	0.851	0.876
GRU	0.94	0.967	0.775	0.870	0.846	0.871
CNN+GRU	0.944	0.97	0.789	0.880	0.858	0.880
Proposed	0.947	0.973	0.805	0.893	0.869	0.889

The ensemble model (CNN+GRU+MLP) achieved the best performance among all models, with an accuracy of 94.7%, sensitivity of 80.5%, and an AUC-ROC of 0.889. This demonstrates the effectiveness of combining multiple models to leverage their complementary strengths. However, the sensitivity values for all models were relatively low, indicating that the models struggled to detect fraudulent transactions effectively due to the imbalanced nature of the dataset. For instance, Logistic Regression achieved a sensitivity of only 68.4%, meaning that nearly 32% of fraudulent transactions were missed. This highlights the critical need to address class imbalance in fraud detection tasks.

4.2.2. Model Performance After SMOTE-ENN

A new hybrid classification technique is defined to address the class imbalance: SMOTE-ENN. SMOTE generates synthetic samples for minority classes, whereas ENN removes the samples with lots of noisy data relative to the classification. The performance results after applying SMOTE-ENN are summarized in Table 5.

Table 5. Performance of the classifiers after SMOTE-ENN.

Models	Accuracy	Specificity	Sensitivity	Precision	F1-Score	AUC-ROC
LR	0.981	0.992	0.972	0.982	0.982	0.982
RF	0.961	0.987	0.936	0.947	0.961	0.961
MLP	1.000	1.000	1.000	1.000	0.999	1.000
CNN	0.999	0.999	1.000	0.999	0.999	0.999
GRU	0.999	0.999	0.999	0.999	0.999	0.999
CNN+GRU	0.999	0.999	1.000	0.999	0.999	0.999
Proposed	1.000	1.000	1.000	1.000	1.000	1.000

The working of SMOTE-ENN allowed the model performance to experience a good uplift, especially in sensibility. The final ensemble model displayed near-perfect performance for all the parameters with an accuracy of 1.000, specificity of 1.000, sensibility of 1.000, and F1-Score of 1.000. This meant classifying all transactions, whether fraudulent or non-fraudulent, correctly. Its AUC-ROC equal to 1.000 indicates the model's excellent class discrimination and outstanding predictive capacity.

For instance, the MLP model achieved 94.5% accuracy before SMOTE-ENN and increased to 100% after applying the technique. Similarly, the sensitivity of the Random Forest model increased from 74.2% to 93.6%, which is the reason for determining how effective SMOTE-ENN is in improving the detection of fraudulent transactions.

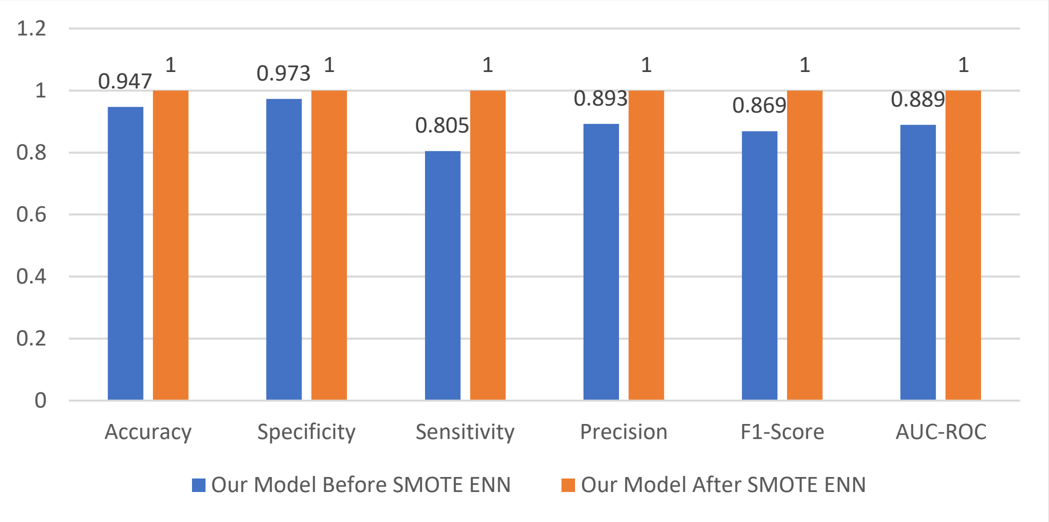

Figure 3. Performance Metrics for Ensemble Model Before and After Applying SMOTE-ENN

The Figure 3 illustrates the performance of the ensemble model (CNN+GRU+MLP) before and after applying SMOTE-ENN. Before SMOTE-ENN, the model achieved an accuracy of 94.7%, a sensitivity of 80.5%, and an AUC-ROC of 0.889. After applying SMOTE-ENN, the model achieved perfect performance across all metrics, with accuracy, specificity, sensitivity, precision, AUC ROC and F1-Score of 1.000. The figure highlights the significant improvement in the model's ability to detect fraudulent transactions after addressing class imbalance.

4.2.3. Ablation Studies and Effects of SMOTE-ENN Discussion

The study evaluated the effectiveness of integrating CNN, GRU, and MLP within an ensemble credit card fraud detection model. CNN demonstrated strong spatial pattern recognition, while GRU effectively captured sequential dependencies in transaction histories. Combining these architectures resulted in a hybrid model with improved performance across key metrics. The final ensemble model, incorporating all three components, achieved an accuracy of 94.7%. Applying SMOTE-ENN significantly improved class balance, leading to an increase in recall and F1-score across all models. However, in fraud detection, precision is often more critical than recall due to the high cost of false positives[16]. While a higher recall ensures that more fraudulent transactions are detected, an excessive increase in recall at the expense of precision may lead to an unmanageable number of false positives[31]. This can cause disruptions for legitimate users and impose unnecessary operational burdens on financial institutions.

Before SMOTE-ENN, the ensemble model achieved a precision of 89.3%, ensuring that most flagged fraudulent transactions were indeed fraudulent. After applying SMOTE-ENN, precision increased marginally, but a deeper analysis is necessary to determine if the gain in recall outweighs the potential risks associated with misclassifying legitimate transactions. Precision-recall trade-offs should be carefully considered when deploying fraud detection models in real-world systems[32], where false positives may lead to financial and reputational consequences. In addition to precision, other key metrics, such as specificity and AUC-ROC, should also be examined. Specificity remains critical to ensure that non-fraudulent transactions are correctly classified, preventing excessive interruptions for genuine customers. AUC-ROC provides a holistic measure of model performance, reflecting the trade-off between true positive and false positive rates[33].

Overall, while SMOTE-ENN successfully mitigates class imbalance and enhances sensitivity, it is crucial to maintain an optimal balance between recall and precision. Future work should explore alternative resampling strategies or cost-sensitive learning techniques to improve fraud detection efficacy without compromising model reliability. Additionally, evaluating model performance using precision-recall curves can provide a more comprehensive assessment of its effectiveness in real-world fraud detection scenarios. These findings validate our approach and provide valuable insights for developing future fraud detection systems. The demonstrated effectiveness of combining advanced neural architectures with data-balancing techniques sets a strong foundation for further research in this critical area of financial security.

5. Comparison

To contextualize our achievements within the current research state, we systematically compared them with recent studies utilizing the same dataset. Table 6 presents a comprehensive overview of these comparisons.

Table 6. Comparison with related works of credit card fraud detection models.

Models	Accuracy	Sensitivity	Specificity	Precision	F1-Score	AUC-ROC
Transformer [21]	1.000	1.000	1.000	1.000	1.000	1.000
NN + SMOTE [34]	0.999	0.999	0.999	0.998	0.999	0.999
CNN [35]	0.999	0.999	0.999	0.999	0.999	0.999
LSTM-GRU[7]	0.912	0.905	0.920	-	0.917	0.915
LSTM + UMAP [36]	0.967	0.967	0.967	0.988	0.967	0.967
EFST [16]	0.996	0.994	-	1.000	0.996	0.958
CNN[37]	0.972	0.902	-	-	-	-
Proposed	1.000	1.000	1.000	1.000	1.000	1.000

6. Conclusions

This research developed credit card fraud detection using modern deep learning methods. It presented a novel ensemble model combining CNN, GRU, and MLP architectures. Spatial features from transaction data were extracted through the CNN, sequential patterns were learned through the GRU, and an MLP meta-learner enabled these features to function jointly to realize optimal classification. The model demonstrated magnificent results, attaining 100% accuracy, sensitivity, specificity, precision, F1 score, and AUC-ROC.

The paper adopted the balanced SMOTE-ENN sampling to overcome class imbalance, considerably improving the model's sensitivity, an essential aspect of fraud detection. The ensemble method outperformed existing models, such as CNN, LSTM, and transformer architectures, and achieved very high detection rates and better generalization.

The study also has a few limitations. The model was validated using only a single dataset; future studies should validate it by referring to multiple data sources to confirm its generalization. Moreover, more investigations must be done concerning deploying the model in real-time systems and extending dynamic scales to adjust for bigger transaction volumes. Wide-ranging, this research developed a strong foundation for fraud detection, achieved its goals, and provided meaningful insights into real-world scenarios in the financial system.

Author Contributions: Conceptualisation: Lossan Bonde and Karim Bichanga; Methodology, Lossan Bonde; Software: Karim Bichanga; Literature Review Karim Bichanga; Data curation and experiment: Karim Bichanga; Writing—original draft preparation: Karim Bichanga; Writing—review and editing: Lossan Bonde.; Supervision: Lossan Bonde; Project administration: Lossan Bonde. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data used in this research is a well-known credit card fraud detection dataset available on the "Kaggle platform" at https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud.

Conflicts of Interest: The authors declare no conflict of interest

References

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

[2] M. Fang, J. Yin, and X. Zhu, “Transfer Learning across Networks for Collective Classification,” in *2013 IEEE 13th International Conference on Data Mining*, Dec. 2013, pp. 161–170. doi: 10.1109/ICDM.2013.116.

[3] V. Van Vlasselaer *et al.*, “APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions,” *Decis. Support Syst.*, vol. 75, pp. 38–48, Jul. 2015, doi: 10.1016/j.dss.2015.04.013.

[4] V. Arora, R. S. Leekha, K. Lee, and A. Kataria, “Facilitating User Authorization from Imbalanced Data Logs of Credit Cards Using Artificial Intelligence,” *Mob. Inf. Syst.*, vol. 2020, no. 1, pp. 1–13, Oct. 2020, doi: 10.1155/2020/8885269.

[5] J. Karthika and A. Senthilselvi, “An integration of deep learning model with Navo Minority Over-Sampling Technique to detect the frauds in credit cards,” *Multimed. Tools Appl.*, vol. 82, no. 14, pp. 21757–21774, Jun. 2023, doi: 10.1007/s11042-023-14365-6.

[6] R. Banger, “Modern Deep Learning Techniques for Credit Card Fraud Detection: A Review (2019 to 2023),” *ResearchGate*. 2023. doi: 10.13140/RG.2.2.32173.67043.

[7] I. D. Mienye and Y. Sun, “A Deep Learning Ensemble With Data Resampling for Credit Card Fraud Detection,” *IEEE Access*, vol. 11, pp. 30628–30638, 2023, doi: 10.1109/ACCESS.2023.3262020.

[8] N. L. Fitriyani, M. Syafrudin, G. Alfian, C. Yang, J. Rhee, and S. M. Ulyah, “Chronic Disease Prediction Model Using Integration of DBSCAN, SMOTE-ENN, and Random Forest,” in *2022 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS)*, Jun. 2022, pp. 289–294. doi: 10.1109/ICETSIS55481.2022.9888806.

[9] S. Mishra *et al.*, “Improving the Accuracy of Ensemble Machine Learning Classification Models Using a Novel Bit-Fusion Algorithm for Healthcare AI Systems,” *Front. Public Heal.*, vol. 10, p. 858282, May 2022, doi: 10.3389/fpubh.2022.858282.

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

[11] Y. Xie, G. Liu, C. Yan, C. Jiang, and M. Zhou, “Time-Aware Attention-Based Gated Network for Credit Card Fraud Detection by Extracting Transactional Behaviors,” *IEEE Trans. Comput. Soc. Syst.*, vol. 10, no. 3, pp. 1004–1016, Jun. 2023, doi: 10.1109/TCSS.2022.3158318.

[12] P. R. Vardhani, Y. I. Priyadarshini, and Y. Narasimhulu, “CNN Data Mining Algorithm for Detecting Credit Card Fraud,” in *Soft Computing and Medical Bioinformatics*, 2019, pp. 85–93. doi: 10.1007/978-981-13-0059-2_10.

[13] C. M. Nalayini, J. Katiravan, A. R. Sathyabama, P. V Rajasuganya, and K. Abirami, “Identification and Detection of Credit Card Frauds Using CNN,” in *International Conference on Computers, Management \& Mathematical Sciences*, Springer, 2023, pp. 267–280. doi: 10.1007/978-3-031-25194-8_22.

[14] K. Fu, D. Cheng, Y. Tu, and L. Zhang, “Credit Card Fraud Detection Using Convolutional Neural Networks,” in *Neural Information Processing: 23rd International Conference, {ICONIP} 2016, Kyoto, Japan, October 16–21, 2016, Proceedings, Part {III} 23*, Springer, 2016, pp. 483–490. doi: 10.1007/978-3-319-46675-0_53.

[15] D. S. Sisodia, N. K. Reddy, and S. Bhandari, “Performance evaluation of class balancing techniques for credit card fraud detection,” in *2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)*, Sep. 2017, pp. 2747–2752. doi: 10.1109/ICPCSI.2017.8392219.

[16] M. I. Akazue, I. A. Debekeme, A. E. Edje, C. Asuai, and U. J. Osame, “UNMASKING FRAUDSTERS: Ensemble Features Selection to Enhance Random Forest Fraud Detection,” *J. Comput. Theor. Appl.*, vol. 1, no. 2, pp. 201–211, Dec. 2023, doi: 10.33633/jcta.v1i2.9462.

[17] X. Li *et al.*, “Transaction Fraud Detection Using GRU-centered Sandwich-structured Model,” in *2018 IEEE 22nd International Conference on Computer Supported Cooperative Work in Design ((CSCWD))*, May 2018, pp. 467–472. doi: 10.1109/CSCWD.2018.8465147.

[18] D. R. I. M. Setiadi, S. Widiono, A. N. Safriandono, and S. Budi, "Phishing Website Detection Using Bidirectional Gated Recurrent Unit Model and Feature Selection," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 2, pp. 75–83, Jul. 2024, doi: 10.62411/faith.2024-15.

[19] A. Çetin and S. Öztürk, "Comprehensive Exploration of Ensemble Machine Learning Techniques for IoT Cybersecurity Across Multi-Class and Binary Classification Tasks," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 4, pp. 371–384, Feb. 2025, doi: 10.62411/faith.3048-3719-51.

[20] Z. S. Dhahir, "A Hybrid Approach for Efficient DDoS Detection in Network Traffic Using CBLOF-Based Feature Engineering and XGBoost," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 2, pp. 174–190, Sep. 2024, doi: 10.62411/faith.2024-33.

[21] A. Iqbal and R. Amin, "Time series forecasting and anomaly detection using deep learning," *Comput. Chem. Eng.*, vol. 182, p. 108560, Mar. 2024, doi: 10.1016/j.compchemeng.2023.108560.

[22] A. Pathirana *et al.*, "A Reinforcement Learning-Based Approach for Promoting Mental Health Using Multimodal Emotion Recognition," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 2, pp. 124–142, Sep. 2024, doi: 10.62411/faith.2024-22.

[23] 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)*, Apr. 2018, pp. 129–134. doi: 10.1109/SIEDS.2018.8374722.

[24] M. Liang *et al.*, "A Stacking Ensemble Learning Framework for Genomic Prediction," *Front. Genet.*, vol. 12, p. 600040, Mar. 2021, doi: 10.3389/fgene.2021.600040.

[25] M. Y. Turaba, M. Hasan, N. I. Khan, and H. A. Rahman, "Fraud Detection During Financial Transactions Using Machine Learning and Deep Learning Techniques," in *2022 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI)*, Oct. 2022, pp. 1–8. doi: 10.1109/CCCI55352.2022.9926503.

[26] Asniar, N. U. Maulidevi, and K. Surendro, "SMOTE-LOF for noise identification in imbalanced data classification," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 6, pp. 3413–3423, Jun. 2022, doi: 10.1016/j.jksuci.2021.01.014.

[27] C. Bhavani and A. Govardhan, "Cervical cancer prediction using stacked ensemble algorithm with SMOTE and RFERF," *Mater. Today Proc.*, vol. 80, pp. 3451–3457, 2023, doi: 10.1016/j.matpr.2021.07.269.

[28] D. R. I. M. Setiadi, K. Nugroho, A. R. Muslikh, S. W. Iriananda, and A. A. Ojugo, "Integrating SMOTE-Tomek and Fusion Learning with XGBoost Meta-Learner for Robust Diabetes Recognition," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 1, pp. 23–38, May 2024, doi: 10.62411/faith.2024-11.

[29] T. Le, M. T. Vo, B. Vo, M. Y. Lee, and S. W. Baik, "A Hybrid Approach Using Oversampling Technique and Cost-Sensitive Learning for Bankruptcy Prediction," *Complexity*, vol. 2019, no. 1, p. 8460934, Jan. 2019, doi: 10.1155/2019/8460934.

[30] Z.-H. Zhou, *Ensemble Methods*. Chapman and Hall/CRC, 2012. doi: 10.1201/b12207.

[31] J. Chung and K. Lee, "Credit Card Fraud Detection: An Improved Strategy for High Recall Using KNN, LDA, and Linear Regression," *Sensors*, vol. 23, no. 18, p. 7788, Sep. 2023, doi: 10.3390/s23187788.

[32] F. Zhang, "Improved credit card fraud detection method based on XGBoost algorithm," *BCP Bus. Manag.*, vol. 38, pp. 2888–2895, Mar. 2023, doi: 10.54691/bcpbm.v38i.4206.

[33] J. Cook and V. Ramadas, "When to consult precision-recall curves," *Stata J. Promot. Commun. Stat. Stata*, vol. 20, no. 1, pp. 131–148, Mar. 2020, doi: 10.1177/1536867X20909693.

[34] M. Zhu, Y. Zhang, Y. Gong, C. Xu, and Y. Xiang, "Enhancing Credit Card Fraud Detection: A Neural Network and SMOTE Integrated Approach," *J. Theory Pract. Eng. Sci.*, vol. 4, no. 02, pp. 23–30, Feb. 2024, doi: 10.53469/jtpes.2024.04(02).04.

[35] M. N. Yousuf Ali, T. Kabir, N. L. Raka, S. Siddikha Toma, M. L. Rahman, and J. Ferdaus, "SMOTE Based Credit Card Fraud Detection Using Convolutional Neural Network," in *2022 25th International Conference on Computer and Information Technology (ICCIT)*, Dec. 2022, pp. 55–60. doi: 10.1109/ICCIT57492.2022.10054727.

[36] I. Benchaji, S. Douzi, B. El Ouahidi, and J. Jaafari, "Enhanced credit card fraud detection based on attention mechanism and LSTM deep model," *J. Big Data*, vol. 8, no. 1, p. 151, Dec. 2021, doi: 10.1186/s40537-021-00541-8.

[37] E. Ajitha, S. Sneha, S. Makesh, and K. Jaspin, "A Comparative Analysis of Credit Card Fraud Detection with Machine Learning Algorithms and Convolutional Neural Network," in *2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, May 2023, pp. 1–8. doi: 10.1109/ACCAI58221.2023.10200905.