International Journal of Electrical and Computer Engineering (IJECE)
Vol. 16, No. 1, February 2026, pp. 311~320
ISSN: 2088-8708, DOI: 10.11591/ijece.v16i1.pp311-320

Credit card fraud data analysis using proposed sampling algorithm and deep ensemble learning

Aye Aye Khine, Zin Thu Thu Myint
Information and Communication Technology Research Centre, Yangon, Myanmar

Article Info

Article history:

Received Jun 3, 2025
Revised Sep 26, 2025
Accepted Nov 23, 2025

Keywords:

Convolutional neural network
long short-term memory
Credit card fraud detection
Ensemble model
Multilayer perceptron

ABSTRACT

Credit card fraud detection is challenging due to the severe imbalance between legitimate and fraudulent transactions, which hinders accurate fraud identification. To address this, we propose a deep learning-based ensemble model integrated with a proposed sampling algorithm based on random oversampling. Unlike traditional methods, the proposed sampling algorithm addresses the oversight of parameter selection and manages class imbalance without eliminating any legitimate samples. The ensemble framework combines the strengths of convolutional neural networks (CNN) for spatial feature extraction, long short-term memory (LSTM) networks for capturing sequential patterns, and multilayer perceptrons (MLP) for efficient classification. Three ensemble strategies—Weighted average, unweighted average, and unweighted majority voting—are employed to aggregate predictions. Experimental results show that all ensemble methods achieve perfect scores (1.00) in precision, recall, and F1-score for both fraud and non-fraud classes. This study demonstrates the effectiveness of ensemble model with optimized sampling approach for robust and accurate fraud detection.

Corresponding Author:

Aye Aye Khine
Information and Communication Technology Research Centre
Yangon, Myanmar
Email: ayekhine71@gmail.com

1. INTRODUCTION

Credit card fraud is an escalating threat in the financial sector, exacerbated by the rapid growth of e-commerce and digital payments. This not only causes substantial financial losses but also undermines consumer trust and poses serious challenges to financial institutions. Existing fraud detection systems predominantly rely on rule-based techniques, which, while simple and interpretable, are increasingly inadequate against the evolving strategies of sophisticated fraudsters. These systems suffer from high false positive rates and lack the adaptability required for real-time and dynamic fraud detection [1]–[3].

To overcome these limitations, the research community has increasingly turned to machine learning and deep learning techniques. Several state-of-the-art approaches have explored the use of individual models such as convolutional neural networks (CNNs), long short-term memory networks (LSTMs), and multilayer perceptrons (MLPs), each offering distinct advantages: CNNs are capable of learning spatial patterns within transaction data, LSTMs excel at modeling temporal sequences to detect irregularities over time, and MLPs are effective at learning complex non-linear relationships. However, prior studies have largely focused on standalone models or conventional ensemble approaches, often without adequately addressing a critical challenge in this domain-class imbalance, where fraudulent transactions represent a tiny fraction of the overall dataset [4], [5].

This study aims to advance the state of the art by proposing a hybrid ensemble framework combining CNN, LSTM, and MLP architectures integrated with a novel sampling algorithm specifically designed to balance the dataset and improve fraud detection performance. Unlike prior works, our approach leverages the complementary strengths of each model type while mitigating the adverse effects of skewed class distributions. Our model is compared against baseline single-architecture models as well as existing ensemble strategies to demonstrate its superiority in detecting both frequent and rare fraud patterns.

These contributions are demonstrated in the experimental sections that follow, offering financial institutions a scalable and intelligent solution to detect fraud more accurately and reliably. The contributions of this paper are as follows:

- The design of a robust deep learning-based ensemble that fuses CNN, LSTM, and MLP components.
- The integration of a proposed sampling method to effectively handle class imbalance.
- A comprehensive evaluation of the proposed system against standard benchmarks and prior methods.
- A detailed analysis showing significant improvements in accuracy, precision, and reduction of false positives.

The structure of the paper unfolds as follows: section 2 provides a comprehensive review of existing literature, encompassing credit fraud detection methodologies, imbalanced datasets, and sampling techniques. Section 3 details the methodology, outlining the proposed sampling algorithm, dataset characteristics, and the architectural details of the integrated deep learning models. Section 4 presents experimental results and analyses. Finally, section 5 concludes the paper by summarizing findings, discussing the system's implications, and suggesting avenues for future research.

2. LITERATURE REVIEWS

Studies [6]–[10] explored advanced machine learning and deep learning techniques for addressing class imbalance in credit card fraud detection, including federated learning with hybrid resampling [6], Transformer-based models [7], AE-Net with extreme gradient boosting [8], CNNs for pattern recognition [9], and AFLCS using Approx-SMOTE with CNN optimization [10]. Research in [11]–[15] highlighted the effectiveness of gradient boosting and random forests [11], optimized ANN models for data sparsity [12], CNN-LSTM hybrids with data augmentation [13], probability-based K-nearest neighbors (KNN) for efficient classification [14], and CNNs combined with deep autoencoders for improved fraud detection [15]. Further advancements [16]–[20] include a particle swarm optimization (PSO)-optimized stacking ensemble with high scalability [16], analysis of DL parameters with Random Forest achieving 99.5% accuracy [17], impact assessment of the "Time" feature across multiple models [18], superior graph neural network (GNN) performance in graph-based anomaly detection [19], and a comparative evaluation of supervised vs. deep learning methods using support vector machines (SVM), K-nearest neighbors (KNN), and artificial neural network (ANN) [20]. Studies [15] emphasized hybrid deep learning architectures such as autoencoder-deep neural networks models for real-time fraud detection [15], CNNs with fully connected layers for enhanced recall [21], and MLPs with dropout and augmentation to handle overfitting and imbalance [22]. Collectively, these works demonstrate the growing effectiveness of hybrid, ensemble, and optimization-based deep learning strategies in building robust and scalable fraud detection systems.

3. METHODOLOGY

The proposed system performs a comparative analysis of credit card fraud detection using CNN, LSTM, MLP models [23], and their ensemble, applied individually and collectively to evaluate their effectiveness in handling a highly imbalanced dataset. An enhanced sampling algorithm is introduced to balance the dataset, mitigating the bias from non-fraudulent transactions and exposing the models to a more representative class distribution. CNN captures spatial patterns, LSTM focuses on temporal dependencies in transaction sequences, and MLP serves as a baseline classifier. The ensemble model combines the strengths of these individual models, leveraging CNN's feature extraction, LSTM's sequential learning, and MLP's classification efficiency. Figure 1 illustrates the proposed system's flow diagram, emphasizing the importance of balanced data in improving model accuracy and reliability.

3.1. Data collection

The dataset used in this study [24] comprises 284,807 credit card transactions by European cardholders over two days in September 2013, of which only 492 (0.172%) are fraudulent, highlighting a severe class imbalance. It includes only numerical features transformed via principal component analysis (PCA), labeled V1 to V28, along with transaction amount and timestamp, allowing for the capture of complex behavioral patterns. The outcome variable, referred to as "Class," is binary—where a value of 0

indicates a genuine transaction and a value of 1 flag a transaction as fraudulent. This rich and multifaceted dataset structure enables detailed analysis and modeling for detecting and preventing fraudulent financial activities. Figure 2 describes the credit card fraud dataset.

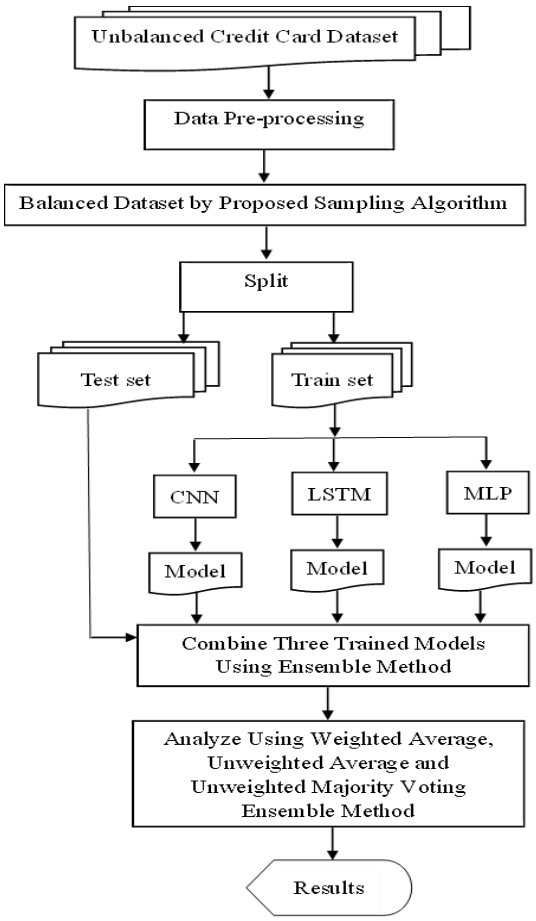

Figure 1. Flow diagram for proposed credit card fraud data analysis system

	Time	V1	V2	V3	V4	V5	.....	V21	V22	V23	V24	V25	V26	V27	V28	Amount	Class
0	0	-1.35981	-0.07278	2.536347	1.378155	-0.33832	.....	-0.01831	0.277838	-0.11047	0.066928	0.128539	-0.18911	0.133558	-0.02105	149.62	0
1	0	1.191857	0.266151	0.16648	0.448154	0.060018	.....	-0.22578	-0.63867	0.101288	-0.33985	0.16717	0.125895	-0.00898	0.014724	2.69	0
2	1	-1.35835	-1.34016	1.773209	0.37978	-0.5032	.....	0.247998	0.771679	0.909412	-0.68928	-0.32764	-0.1391	-0.05535	-0.05975	378.66	0
3	1	-0.96627	-0.18523	1.792993	-0.86329	-0.01031	.....	-0.1083	0.005274	-0.19032	-1.17558	0.647376	-0.22193	0.062723	0.061458	123.5	0
4	2	-1.15823	0.877737	1.548718	0.403034	-0.40719	.....	-0.00943	0.798278	-0.13746	0.141267	-0.20601	0.502292	0.219422	0.215153	69.99	0
5	2	-0.42597	0.960523	1.141109	-0.16825	0.420987	.....	-0.20825	-0.55982	-0.0264	-0.37143	-0.23279	0.105915	0.253844	0.08108	3.67	0
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
284803	172637	-2.31223	1.951992	-1.60985	3.997906	-0.52219	.....	0.517232	-0.03505	-0.46521	0.320198	0.044519	0.17784	0.261145	-0.14328	0	1
284804	172638	-3.04354	-3.15731	1.088463	2.288644	1.359805	.....	0.661696	0.435477	1.375966	-0.2938	0.279798	-0.14536	-0.25277	0.035764	529	1
284805	172638	-2.52189	1.720516	-0.89097	4.51669	0.103394	.....	-0.02768	1.038627	-0.59236	0.03839	1.155201	0.856059	-0.97022	-0.69805	302.1	0
284806	172639	-0.90596	1.710245	1.934267	4.190227	-0.38852	.....	0.311748	0.806358	-0.29809	-0.36394	0.198075	0.563902	-0.09658	-0.00336	100.2	0

Figure 2. Dataset for credit card fraud data

3.2. Data preprocessing

Data normalization is a crucial preprocessing step in machine learning that ensures all features are on a comparable scale, preventing features with larger ranges from dominating the learning process. In this

system, min-max normalization is applied to scale data within the range of 0 to 1, standardizing feature values for improved model performance and computational efficiency [25]. This consistent scaling not only accelerates model convergence during training but also helps maintain the integrity of feature relationships across different learning algorithms. The normalization formula used is:

$$v' = \frac{v - min_s}{max_s - min_s}(new\_max_s - new\_min_s) + new\_min_s \quad (1)$$

This normalization process adjusts each data point v to a new value v′ that fits within the normalized range. Following data normalization, the system applies a proposed sampling technique to address data imbalance. This sampling method ensures that the dataset used for model training is more representative, enhancing the performance and accuracy of the fraud detection models.

3.3. Proposed sampling algorithm

Class imbalance in the Credit Card dataset can lead to biased machine learning models that struggle to correctly classify minority class instances, resulting in high false-negative rates and reduced sensitivity. This system tackles class imbalance using the proposed sampling approach, avoiding the overfitting issues of oversampling and the data loss risks of undersampling. In Algorithm 1, the proposed data balancing algorithm is shown. The proposed system introduces a proposed sampling algorithm to balance the dataset by generating synthetic minority class instances, ensuring equal representation of normal and fraudulent transactions. By removing the label from the feature vector, the system prevents data leakage and maintains model integrity while applying the sampling technique effectively. Unlike traditional oversampling methods, this approach creates new, synthetic samples rather than duplicating existing ones, reducing overfitting and improving model performance.

Algorithm 1. Proposed data balancing algorithm

```
Step 1. Computes the quantity of normal and anomalous entries in the training dataset.
Step 2. If the count of the normal class is greater than the count of the abnormal class,
         from step 3 to step 9 is performed.
Step 3. Calculates the difference (num) between the counts of the two classes.
Step 4. For each iteration (from 0 to num),
Step 5. Extracts a feature vector from the abnormal class.
Step 6. Eliminates the last component of the feature vector that was extracted.
Step 7. Acquires a feature value from the neighboring anomalous data point.
Step 8. Integrates this feature into the updated feature vector.
Step 9. The updated feature vector is appended to the abnormal dataset, along with the
         label 1.
Step 10. When a class imbalance is detected, with the abnormal class being predominant,
         steps 11 to 17 are performed.
Step 11. Computes the numerical difference (num) between the class counts.
Step 12. For each iteration (from 0 to num),
Step 13. Extracts a feature vector from the normal class.
Step 14. The extracted feature vector is modified by removing its last entry.
Step 15. A feature value is obtained from a nearby normal instance.
Step 16. Integrates this feature into the updated feature vector.
Step 17. The updated feature vector and label 0 are added to the normal dataset.
```

A key step in balancing credit card transaction datasets involves removing the final element of the feature vector. This last element typically represents the label, such as "fraudulent" or "non-fraudulent," which indicates whether a transaction is fraudulent or not. By excluding this label from the feature vector, the proposed sampling techniques can be applied more effectively to address class imbalance. Removing the label from the feature vector is crucial for several reasons. Firstly, it prevents data leakage, ensuring that the model is trained on features alone without the influence of the target variable. This practice adheres to model input requirements, maintaining the integrity of the data and ensuring that the model learns from the actual features rather than being biased by the labels. Secondly, it ensures that synthetic samples are generated solely from genuine transaction characteristics rather than being influenced by class identifiers. Thirdly, it preserves a clear separation between features and labels, which is essential for reproducibility and correct implementation of the sampling algorithm. Finally, this approach supports fair model evaluation by preventing artificially inflated performance that could arise if label information leaked into the feature space. The proposed sampling algorithm is adaptable to both numerical and categorical data. It allows for adjustments based on specific needs, particularly by modifying the method used to generate synthetic samples. Unlike traditional oversampling methods that simply duplicate existing samples, this algorithm focuses on creating new, synthetic samples. This sampling algorithm helps in reducing the risk of overfitting by introducing variability into the dataset rather than merely increasing the number of identical instances.

3.4. Deep ensemble learning-based credit card detection

To improve fraud detection performance, the dataset is first balanced using a proposed sampling algorithm that increases the representation of fraudulent transactions without eliminating legitimate samples, effectively addressing class imbalance. The balanced data is then divided into training and testing sets to assess the performance of three individual deep learning models: CNN, LSTM, and MLP. These models are chosen for their respective strengths-CNN for capturing spatial patterns in transaction features, LSTM for modeling temporal dependencies, and MLP for handling complex nonlinear relationships. An ensemble model is then constructed by integrating the outputs of these three models to enhance overall classification accuracy. This ensemble is evaluated using three aggregation strategies: weighted average, where stronger models have greater influence on the final prediction; unweighted average, which treats all models equally; and unweighted majority voting, where the final class is determined by the majority of individual model predictions. This approach ensures robust and fair decision-making in fraud detection.

4. RESULTS AND DISCUSSION

The performance of CNN, LSTM, MLP, and their ensemble model is evaluated for credit card fraud detection, with the dataset split into 80% training and 20% testing portions. The ensemble model, which combines the strengths of each individual model, is assessed using metrics like precision, recall, and F1-score, with tests conducted on both the original and balanced datasets to evaluate the impact of the sampling strategy. Table 1 describes the parameter configurations of the system. Figure 3 describes the receiver operating characteristic (ROC) curve of CNN with and without balancing algorithm. Table 2 describes the results of CNN with and without balancing algorithms. Figure 4 describes the ROC Curve of LSTM with and without balancing algorithm. Table 3 describes the results of LSTM with and without balancing algorithm. Figure 5 describes the ROC Curve of MLP with and without balancing algorithm. Table 4 describes the results of MLP with and without balancing algorithm. Table 5 describes the analysis results of ensemble with weighted average, unweighted average and unweighted majority voting by proposed sampling algorithm.

Table 1. Parameter configurations

Parameters	CNN	LSTM	MLP
Number of	Filters :16	Units: 50	Units: 16
Activation function	sigmoid	sigmoid	sigmoid
Dropout rate	0.5	-	-
Optimizer	Adam	Adam	Adam
Epochs	10	10	10
Batch size	64	64	64
Kernel size	3	-	-
Pooling strategy	Max Pooling (pool size = 2)	-	-
Learning rate	0.001	0.001	0.001
Loss function	*binary_crossentropy*	*binary_crossentropy*	*binary_crossentropy*

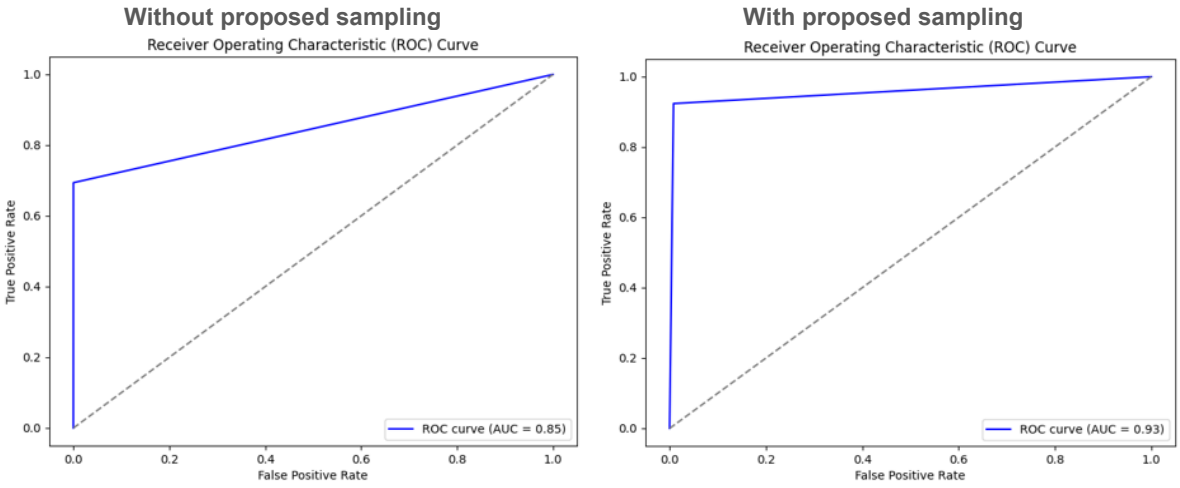

Figure 3. ROC curve for CNN

Table 2. Comparison between normal and fraud using CNN

CNN	Normal			Fraud		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Without Proposed Sampling	0.89	0.91	0.9	0.88	0.75	0.80
With Proposed Sampling	0.96	0.97	0.96	0.96	0.97	0.96

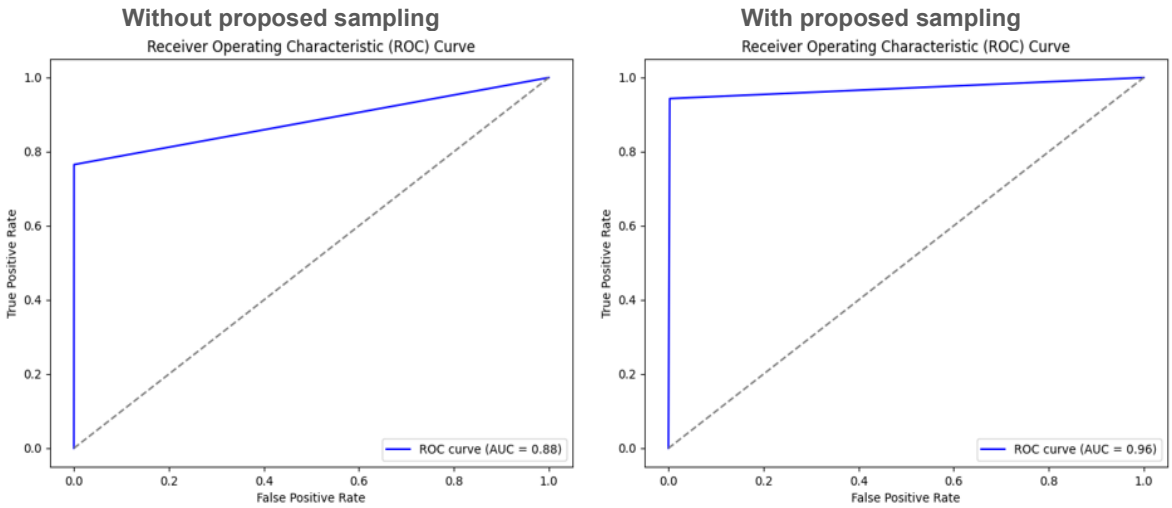

Figure 4. ROC curve for LSTM

Table 3. Comparison between normal and fraud using LSTM

LSTM	Normal			Fraud		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Without Proposed Sampling	0.91	0.74	0.77	0.88	0.73	0.80
With Proposed Sampling	0.98	0.90	0.94	0.98	0.91	0.94

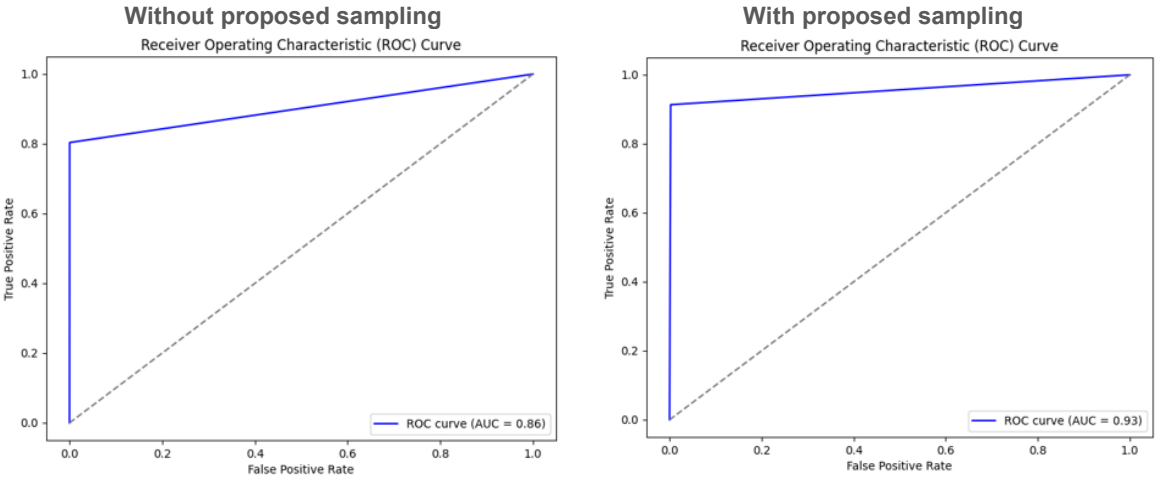

Figure 5. ROC curve for MLP

Table 4. Comparison between normal and fraud using MLP

MLP	Normal			Fraud		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Without Proposed Sampling	0.82	0.83	0.82	0.80	0.83	0.81
With Proposed Sampling	0.90	0.94	0.92	0.92	0.91	0.91

Table 5. Analysis results of ensemble model

Ensemble Model	Normal			Fraud		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Weighted Average	1	1	1	1	1	1
Unweighted Average	1	1	1	1	1	1
Unweighted Majority Voting	1	1	1	1	1	1

The results show that all three ensemble approaches—Weighted average, unweighted average, and unweighted majority voting—achieve perfect scores of 1.0 for both normal and fraudulent transactions, indicating high accuracy and no misclassifications. The ensemble models, despite their different methodologies, demonstrate robustness and reliability in accurately identifying transaction categories. The proposed sampling method significantly improves classification metrics for individual models, including CNN, LSTM, and MLP, enhancing precision, recall, and F1-scores for both normal and fraudulent transactions. Overall, the ensemble models, supported by the sampling method, deliver optimal performance in fraud detection, effectively addressing class imbalance and ensuring precise and reliable predictions.

Table 6 depicts the performance analysis of enhanced sampling with other sampling approaches for credit card dataset. Table 6 presents a comparative performance analysis of various sampling techniques applied to fraud detection models, highlighting their impact on precision, recall, and F1-Score for both normal and fraudulent transactions. The proposed enhanced sampling method significantly outperforms all others by achieving perfect scores (1.00) across all metrics for both classes, effectively addressing class imbalance without sacrificing legitimate transaction accuracy. In contrast, methods like undersampling and NearMiss suffer from extremely low precision for fraud detection (0.04 and 0.05, respectively), while SMOTE, ADASYN, and Tomek Links yield moderate improvements but still fall short compared to the enhanced sampling approach. The consistent 1.00/1.00/1.00 for the Normal class across multiple sampling methods is largely due to the extreme class imbalance in the ULB dataset, where the vast majority of samples belong to the Normal class, making it easy for models to classify them correctly. However, this does not reflect true model robustness, as the minority Fraud class remains challenging, and in some cases the imbalance combined with resampling may also introduce hidden leakage or bias.

Table 6. Performance analysis results with various sampling approaches

Models	Normal			Fraud		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Original Unbalanced	1	1	1	0.94	0.82	0.87
Oversampling	1	1	1	0.95	0.77	0.85
Undersampling	1	0.96	0.98	0.04	0.92	0.08
SMOTE	1	1	1	0.84	0.83	0.83
ADASYN	1	1	1	0.82	0.81	0.81
NearMiss	1	0.97	0.98	0.05	0.91	0.09
Tomek Links	1	1	1	0.93	0.82	0.87
Enhanced Sampling	1	1	1	1	1	1

Table 7 presents a comparative performance analysis between the proposed ensemble model and other state-of-the-art approaches using different datasets. While the CNN, ANN, and RNN models reported high precision, recall, and F1-Scores ranging from 0.91 to 0.99, the proposed ensemble model of CNN, LSTM, and MLP outperformed all with perfect scores of 1.0 across all metrics. This demonstrates the effectiveness of the ensemble approach in capturing complex patterns and improving fraud detection accuracy compared to individual deep learning models.

Table 7. Comparison with other cutting-edge techniques

Ref.	Model	Dataset	Precision	Recall	F1-Score
[7]	CNN	Commercial bank B2C online transaction data	0.91	0.94	0.93
[13]	ANN	Credit card fraud detection from Kaggle	0.98	0.97	0.97
[15]	RNN	Balanced shared 559856-record Kaggle repository	0.98	1	0.99
Proposed Ensemble model	Ensemble model of CNN, LSTM, and MLP	Credit card fault detection from Kaggle	1	1	1

The study demonstrates that integrating a deep ensemble learning model with an enhanced oversampling technique effectively addresses the critical issue of class imbalance in credit card fraud detection. Individual models (CNN, LSTM, and MLP) showed substantial improvements in F1-Score for fraud detection when trained with the proposed sampling method—rising from 0.80 to 0.96 for CNN, 0.80 to 0.94 for LSTM, and 0.81 to 0.91 for MLP. The ensemble strategies—weighted average, unweighted average, and majority voting—achieved perfect scores (1.00 precision, recall, and F1-Score) for both fraud and normal classes, outperforming individual models. Compared to traditional sampling techniques like SMOTE, ADASYN, and NearMiss, the proposed enhanced sampling algorithm demonstrated superior effectiveness by achieving perfect classification metrics. These findings confirm that the proposed approach offers a highly accurate, balanced, and reliable fraud detection system, surpassing existing state-of-the-art techniques.

5. CONCLUSION

With the use of the proposed sampling strategy, a comparative study of credit card fraud detection using CNN, LSTM, MLP, and their ensemble model on the credit card dataset provides crucial insights into how effectively these models handle unbalanced data. The proposed sampling approach significantly mitigates the issue of class imbalance, leading to measurable gains in F1-Score, precision, and recall across all models. Among the individual models, LSTM excelled in capturing temporal patterns within transaction data, making it particularly effective for fraud detection, while CNN demonstrated strong feature extraction capabilities, and MLP provided a straightforward yet effective approach. The ensemble model further amplified these strengths, achieving superior performance through the combined advantages of all three models. The results highlight substantial improvements, particularly in recall and F1-Score for fraud detection, when the proposed sampling method is applied. Without this method, the models struggled with imbalances, especially in recall for fraudulent cases. However, the proposed sampling technique enabled consistent and significant performance gains across all models, with the ensemble model achieving the best overall results due to its ability to leverage the complementary strengths of CNN, LSTM, and MLP. These findings underscore the value of the ensemble approach and the proposed sampling strategy in improving classification accuracy for both normal and fraudulent instances. Future research could focus on refining the sampling algorithm further and extending its application to other types of fraud and diverse datasets.

ACKNOWLEDGMENTS

We would like to express our sincere gratitude to all instructors, colleagues, and institutions whose guidance and support greatly contributed to the successful completion of this research. We are also thankful to our family and friends for their continuous encouragement throughout the study.

FUNDING INFORMATION

Authors state no funding involved.

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
Aye Aye Khine	✓	✓	✓	✓	✓	✓		✓	✓	✓				
Zin Thu Thu Myint		✓								✓	✓	✓	✓	

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

To ensure fair and objective decision-making, authors must declare any associations that pose a conflict of interest (financial, personal, or professional) in connection with manuscripts. Non-financial

competing interests include a declaration of political, personal, religious, ideological, academic, and intellectual competing interests. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. If there are no conflicts of interest, please include the following author's statement: Authors state no conflict of interest.

INFORMED CONSENT

Not Applicable.

ETHICAL APPROVAL

This study did not involve any experiments on humans or animals, and all procedures complied with institutional and international ethical guidelines. As the research was conducted using publicly available datasets, no additional ethical approval was required.

DATA AVAILABILITY

Publicly available datasets were analyzed in this study. This data can be found here: Kaggle, "Credit card fraud detection." 2025, [Online]. Available: *https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud*. [Accessed: Oct. 5, 2024].

REFERENCES

[1] M. Jabeen, S. Ramzan, A. Raza, N. L. Fitriyani, M. Syafrudin, and S. W. Lee, "Enhanced credit card fraud detection using deep hybrid CLST model," *Mathematics*, vol. 13, no. 12, p. 1950, 2025, doi: 10.3390/math13121950.

[2] F. Z. El Hlouli, J. Riffi, M. A. Mahraz, A. Yahyaouy, K. El Fazazy, and H. Tairi, "Credit card fraud detection: Addressing imbalanced datasets with a multi-phase approach," *SN Computer Science*, vol. 5, no. 1, p. 173, 2024, doi: 10.1007/s42979-023-02559-6.

[3] E. Ileberi and Y. Sun, "A hybrid deep learning ensemble model for credit card fraud detection," *IEEE Access*, vol. 12, pp. 175829–175838, 2024, doi: 10.1109/ACCESS.2024.3502542.

[4] I. D. Mienye and Y. Sun, "A machine learning method with hybrid feature selection for improved credit card fraud detection," *Applied Sciences (Switzerland)*, vol. 13, no. 12, p. 7254, 2023, doi: 10.3390/app13127254.

[5] M. Seera, C. P. Lim, A. Kumar, L. Dhamotharan, and K. H. Tan, "An intelligent payment card fraud detection system," *Annals of Operations Research*, vol. 334, no. 1–3, pp. 445–467, Mar. 2024, doi: 10.1007/s10479-021-04149-2.

[6] M. Abdul Salam, K. M. Fouad, D. L. Elbably, and S. M. Elsayed, "Federated learning model for credit card fraud detection with data balancing techniques," *Neural Computing and Applications*, vol. 36, no. 11, pp. 6231–6256, 2024, doi: 10.1007/s00521-023-09410-2.

[7] C. Yu, Y. Xu, J. Cao, Y. Zhang, Y. Jin, and M. Zhu, "Credit card fraud detection using advanced transformer model," in *Proceedings - 2024 IEEE International Conference on Metaverse Computing, Networking, and Applications, MetaCom 2024*, 2024, pp. 343–350, doi: 10.1109/MetaCom62920.2024.00064.

[8] N. Thalji, A. Raza, M. S. Islam, N. A. Samee, and M. M. Jamjoom, "AE-Net: Novel autoencoder-based deep features for SQL injection attack detection," *IEEE Access*, vol. 11, pp. 135507–135516, 2023, doi: 10.1109/ACCESS.2023.3337645.

[9] Y. Murugan, M. Vijayalakshmi, L. Selvaraj, and S. Balaraman, "Credit card fraud detection system using CNN," *Lecture Notes in Networks and Systems*, vol. 340 LNNS, no. 3, pp. 194–204, 2022, doi: 10.1007/978-3-030-94507-7_19.

[10] J. Wang, W. Liu, Y. Kou, D. Xiao, X. Wang, and X. Tang, "Approx-SMOTE federated learning credit card fraud detection system," in *Proceedings - International Computer Software and Applications Conference*, 2023, vol. 2023-June, pp. 1370–1375, doi: 10.1109/COMPSAC57700.2023.00208.

[11] I. Vejalla, S. P. Battula, K. Kalluri, and H. K. Kalluri, "Credit card fraud detection using machine learning techniques," in *2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing, PCEMS 2023*, 2023, pp. 1–6, doi: 10.1109/PCEMS58491.2023.10136040.

[12] P. Yadlapalli, P. Srivatsal, N. Polimera, and M. Srinivas, "Credit card fraud detection using machine learning algorithms and artificial neural network," in *2025 International Conference on Artificial Intelligence and Data Engineering, AIDE 2025 - Proceedings*, 2025, pp. 539–542, doi: 10.1109/AIDE64228.2025.10987537.

[13] P. K. Zorion, L. Sachan, R. Chhabra, V. Pandey, and D. H. Fatima, "Credit card financial fraud detection using deep learning," *SSRN Electronic Journal*, 2023, doi: 10.2139/ssrn.4629093.

[14] K. R.V., S. Ganesh, M. D, B. G.R., and M. Thiruvengadam, "Credit card fraud data analysis and prediction using machine learning algorithms," *Security and Privacy*, vol. 8, no. 3, p. e70043, 2025, doi: 10.1002/spy2.70043.

[15] E.-S. M. El-Kenawy, "Credit card fraud detection based on deep learning models," *Mesopotamian Journal of Computer Science*, vol. 2024, pp. 204–213, 2024, doi: 10.58496/MJCSC/2024/016.

[16] R. K. Gupta *et al.*, "Enhanced framework for credit card fraud detection using robust feature selection and a stacking ensemble model approach," *Results in Engineering*, vol. 26, p. 105084, 2025, doi: 10.1016/j.rineng.2025.105084.

[17] P. Sundaravadivel, R. A. Isaac, D. Elangovan, D. KrishnaRaj, V. V. L. Rahul, and R. Raja, "Optimizing credit card fraud detection with random forests and SMOTE," *Scientific Reports*, vol. 15, no. 1, p. 17851, 2025, doi: 10.1038/s41598-025-00873-y.

[18] A. P. Lopes, S. Parshionikar, A. Kale, N. Sharma, and A. A. Varghese, "A comparative study of deep learning techniques for credit card fraud detection," in *2021 International Conference on Advances in Computing, Communication, and Control (ICAC3)*, 2021, pp. 1–5, doi: 10.1109/ICAC353642.2021.9697205.

[19] 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.

[20] A. RB and S. K. KR, "Credit card fraud detection using artificial neural network," *Global Transitions Proceedings*, vol. 2, no. 1, pp. 35–41, 2021, doi: 10.1016/j.gltp.2021.01.006.
[21] K. Yamini, V. Anitha, S. Polepaka, R. Chauhan, Y. Varshney, and M. Singh, "An intelligent method for credit card fraud detection using improved CNN and extreme learning machine," in *2023 8th International Conference on Communication and Electronics Systems (ICCES)*, 2023, pp. 810–815, doi: 10.1109/ICCES57224.2023.10192774.
[22] S. Negi, S. K. Das, and R. Bodh, "Credit card fraud detection using deep and machine learning," in *2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, 2022, pp. 455–461, doi: 10.1109/ICAAIC53929.2022.9792941.
[23] J. Brownlee, "A gentle introduction to transfer learning for deep learning," *Machine Learning Mastery*, 2017.
[24] Machine Learning Group - ULB, "Credit card fraud detection," *Kaggle*, 2025. https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud (accessed Oct. 05, 2024).
[25] J. Han, M. Kamber, and J. Pei, *Data mining: Concepts and techniques*, 3rd Editio. Elsevier, 2011.

BIOGRAPHIES OF AUTHORS

Aye Aye Khine is an accomplished Associate Professor at IMCEITS at Information Communication Technology Research Center. She holds a M.I.Sc from the University of Computer Studies Yangon, where she specialized in advanced research within her field. Her areas of expertise include Java programming, Oracle DB, data science, and deep learning. She can be contacted her at email: ayekhine71@gmail.com.

Zin Thu Thu Myint is an accomplished Associate Professor at the Faculty of Information Science at the University of Computer Studies, Yangon. She holds a Ph.D. from the University of Technology (Yadanapon Cyber City), where she specialized in advanced research within her field. Her areas of expertise include semantic word processing, data science, and cybersecurity. She can be contacted at email: zinthuthumyint@ucsy.edu.mm.