Federated Learning-Based Credit Card Fraud Detection: Performance Analysis with Sampling Methods and Deep Learning Algorithms

Nahid Ferdous Aurna
Division of Information Science
Nara Institute of Science and Technology
Nara, Japan
aurna.nahid_ferdous.ap2@is.naist.jp

Md Delwar Hossain
Division of Information Science
Nara Institute of Science and Technology
Nara, Japan
delwar.hossain@is.naist.jp

Yuzo Taenaka
Division of Information Science
Nara Institute of Science and Technology
Nara, Japan
yuzo@is.naist.jp

Youki Kadobayashi
Division of Information Science
Nara Institute of Science and Technology
Nara, Japan
youki-k@is.naist.jp

2023 IEEE International Conference on Cyber Security and Resilience (CSR) | 979-8-3503-1170-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/CSR57506.2023.10224978

***Abstract*—The exponential technological advancement is turning everyone towards an easy and efficient way of financial transactions. Consequently, the use of credit cards is rising substantively, creating a more incredible opportunity for fraudsters which is an alarming concern nowadays since a fraudster may use several tools, techniques and tactics to make a fraudulent transaction. As a countermeasure, an effective fraud detection mechanism and highly sensitive data privacy preservation are imperative to detect fraudulent transactions. This paper proposes a Federated Learning (FL)-based fraud detection system since its key feature preserves the privacy of highly sensitive data, wherein the model could be trained without sharing the credit card data in the cloud. We contemplate three Deep Learning (DL) models: Convolutional Neural Network (CNN), Multi-Layer Perceptron (MLP), and Long Short-Term Memory (LSTM) regarding the FL approach. Subsequently, to overcome the data imbalance issue, four distinct sampling techniques are explored to inspect the impact on the traditional centralized and FL approaches. Finally, we further investigate and compare FL-based detection systems with diversified state-of-the-art models. Our experimental results demonstrate that the proposed method is superior compared with state-of-the-art methods and achieves high detection rate of 99.51%, 98.77% and 98.20% respectively for CNN, MLP and LSTM models.**

***Index Terms*—Credit card fraud, fraud detection system, federated learning, CNN, MLP, LSTM**

I. Introduction

With the help of digital transformation, traditional banking systems have significantly advanced and also have a tremendous impact on modern banking systems as well as financial sectors. In consequence, people are turning towards digital cashless system and preferring credit card for their day-to-day transaction. However, the easy access to several tools, techniques, and tactics of making fraudulent credit card transactions has become a significant concern about the acceptance of digital transformation. Financial sector has become a potential playground for fraudsters because of the possibility of obtaining a considerable amount of money within a very short period of time [1]. For example, the European central bank has reported losing billions of euros yearly through credit card fraud [2].

An attacker may effortlessly compromise the credit card data over the dark web, public Wi-Fi, etc. For instance, 1.2 million credit card details were revealed through a dark web marketplace in October 2022 [3]. The researchers devised several methodologies over the decade to detect/prevent fraud transactions; however, still, it's inefficient to be accepted by the financial systems since it produces vast amounts of false positives, also considering the evolution of fraud pattern [4]–[7]. The utmost data privacy and integrity concern reside in, in fact, the traditional Machine Learning (ML) approach, sensitive private data need to be shared into the cloud to train the ML models. Moreover, number of fraud transaction is significantly low compared to the number of legitimate transaction which results in data imbalance issue and makes it more challenging for the researchers to attain a satisfactory detection rate.

DL algorithms have been successfully adopted in numerous real-time applications due to their capability of recognizing complex features or behavioral patterns of data. Generally, DL algorithms require a massive amount of data to properly train the models. However, since privacy is a concern, it's challenging to share highly sensitive credit card transaction data by the financial organization. Here, FL techniques could be considered as a preeminent solution as users don't need to share their private data in a central server. In contrast with the traditional approach, in federated learning, models are trained locally without data sharing and it facilitates sharing of a base model weight among the local clients. Additionally, diversified

sampling techniques, i.e., Random Oversampling, Synthetic Minority Oversampling Technique, Random Undersampling, Near Miss Undersampling could be the effective ways to improve the detection rate and reduce false positives considering imbalance datasets.

In this research, we intend to develop a robust credit card fraud detection system, which is challenging due to the lack of real-world labeled datasets [8], skewed distribution of data [9], changes in fraud patterns, and so on. Subsequently, We contemplate addressing the aforementioned issues by employing federated learning and deep learning models. Thus, we try to adopt federated learning to combat data privacy issues and apply three different DL models: CNN, MLP, and LSTM as the base model. Before incorporating the FL approach, several experimentation are conducted to attain the most optimal architectures of the models to overcome the high false positive rate. Subsequently, through an experimental analysis, we endeavor to combat the data imbalance issue using four different sampling techniques. Our main contributions are outlined below:

- Developing an effective fraud detection system with high detection rate, also ensuring the privacy of highly sensitive credit card data using FL technology.
- Investigating the performance with three DL models: MLP, CNN, and LSTM, considering traditional centralized as well as FL-based approach.
- Experimenting with the proposed models on distinct sampling methods and benchmarking them against diversified state-of-the-art methods.

The remaining part of the paper is presented as follows: section II describes some recent existing researches related to our topic. In section III, the total methodology and experimental process are explained. Section IV thoroughly explains the results and analysis considering various aspects. Discussion and future direction of this work is presented in section V. Finally, the conclusion of the work is added in section VI.

II. Related Work

Due to financial significance, credit card fraud detection has always been an important area for the researchers [10]–[12]. As the patterns of attacks and ways are evolving, there is always the scope of improvement of existing researches. This section provides discussion on different existing methods and the scope of improvements of these works.

Asha et al. [13] proposed a deep learning-based method using neural network to detect the fraudulent transactions. Out of all models Artificial Neural Network (ANN) performed the best with 99% accuracy, however, the detection rate was quite low, that is about 76%. Chen et al. [14] proposed a deep neural network for credit card fraud detection and attained 99% accuracy, but for this type of problems detection rate is very much important which is not considered. Esenogho et al. [15] presented an ensemble-based approach using LSTM model as the base classifier by employing Adapting Boosting algorithm which attained a sensitivity of 99%. They did not explain the base LSTM model or its structure in detail so its difficult to reimplement this for future researchers. Fiore et al. [16] tried to address the data imbalance problem of the highly imbalance credit card dataset by applying Generative Adversarial Network (GAN) to generate synthetic data. An improved sensitivity was achieved at the cost of high false positive rate that degraded the overall performance. Varmedja et al. [17] showed several algorithms that can be used for detection fraud transaction. Among all the models, RF proved to be the best, with accuracy of 95.5%. Though the accuracy was quite good, the detection rate of fraud transaction was not satisfactory (81.63%).

There are few researches regarding credit card fraud detection with FL approach. Yang et al. [18] have used federated average approach where a CNN is used as the shared global model. It is shown that with this approach they achieved 10% better AUC score. Zheng et al. [19] used a federated meta learning based technique where ResNet-34 works as the feature extractor and CNN works for final classification. Their approach seemed to be better than other classical approaches but still there are scopes of further analysis on the FL base model. Suvarna et al. [20] used Encoder and Restricted Boltzmann Machine over the classical FL approach. Though the accuracy is being reduced, they could ensure data privacy through federated learning. However, the model architecture could have been added for better investigation of the method.

Data privacy is a major concern for any kind of banking transaction which has not been addressed in most of the researches. Also a thorough investigation of applied models, improving detection rate, experimentation on data sampling techniques etc. are important factors of this research. All these issues have been tried to address in our work.

III. Methodology

The aim of this work is to build and analyze a robust and privacy-preserving model for the detection of credit card fraud. We can summarize our study in five phases as shown in Fig. 1. The credit card data is first preprocessed, and three different models are built based on MLP, LSTM, and CNN. Afterward, to ensure the privacy of the data, the federated learning approach is incorporated with those conventional models. Furthermore, model validation is accomplished through various experiments. Finally, our proposed models are compared with recent state-of-the-art models.

A. Dataset Analysis and preprocessing

The dataset used in this paper contains credit card transactions made by the European card holders, publicly avaibale in [21]. These transactions occurred in September 2013 for two days. There are a total 284,807 transaction samples out of which only 492 transactions are fraud. The fraud transactions are only 0.172% of the whole dataset that indicates the dataset is highly imbalanced. There are total 31 features namely 'Time', 'V1' to 'V28', 'Amount' and 'Class'. Features 'V1' to 'V28' are Principal Component Analysis (PCA) transformed features to anonymize sensitive and confidential information, besides, all the features are already converted to numeric form.

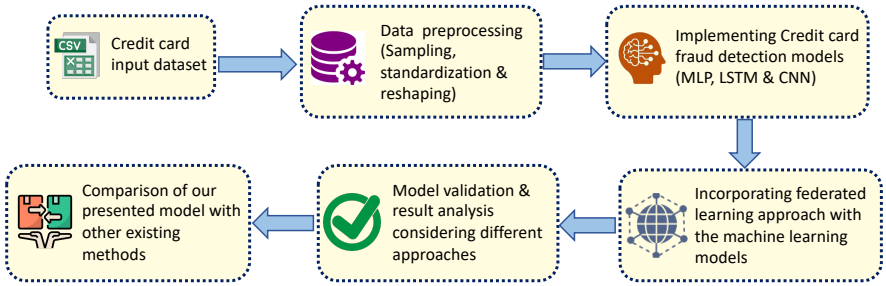

Fig. 1: The summarized workflow of the proposed method

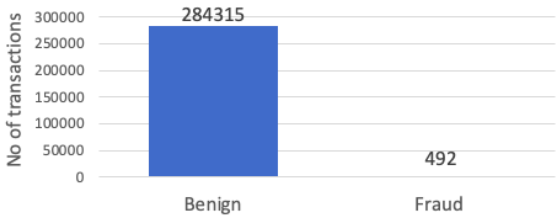

Fig. 2: Distribution of data among two classes (Benign and Fraud)

TABLE I: Hyper-parameter values used in proposed models

Hyper-parameter	CNN	MLP	LSTM
Input activation function	ReLU	ReLU	ReLU
Output activation function	Sigmoid	Sigmoid	Sigmoid
Optimizer	Adam	Adam	Adam
Initial learning rate	0.0001	0.0001	0.001
Learning rate decay	0.2	0.3	0.2
Dropout rate	0.2, 0.4	0.5	0.5
Communication round	50	50	50
No of federated clients	3	3	2
Train test ratio	80%-20%	80%-20%	80%-20%

As this dataset is used in other similar researches already, it's easier to analyze and compare with existing models and benchmark.

In this work, data preporcessing is mainly done in three steps: resampling, standardization and reshaping. The dataset used here is highly imbalanced as it is observed from Fig. 2. As the minority class belongs to the non fraudulent data (benign), there is high possibility of getting low detection rate if we use the raw data without any kind of sampling technique. Therefore, four different sampling methods are used, namely: Ranodom Oversampling (RO), Synthetic Minority Oversampling Technique (SMOTE), Random Undersampling (RU), and Near Miss Undersampling (NMU). After sampling the dataset, its standardized by using *StandardScaler*. As the features of the credit card dataset is not of same type and the distribution ranges are totally different from each other, standardization makes it more feasible for the ML models to train. Furthermore, the data is reshaped by adding one extra dimension to make it compatible with CNN and LSTM model.

B. Applied Algorithms

Three different algorithms are applied for our experiments: CNN, MLP and LSTM. These models are built after several trial and error experimenting with different hyper-parameters. Initially *keras tuner* and random search have been used for searching optimal architectures. Finally, the most optimal ones are chosen to implement the models. The hyper-parameter values used for the models are shown in TABLE I.

1) CNN for credit card fraud detection: The CNN model used in this work consists of total 10 layers as depicted in Fig. 3(a). It includes 2 convolution layers, 3 dropout layers, 2 batchnormalization layers, 1 flatten layer and 2 dense layers where number of filters used in convolution layer 1 and 2 are 32 and 64 respectively. According to different experiments and observation, this architecture seemed feasible enough for this particular problem.

2) MLP for credit card fraud detection: The MLP model used for our experiment consists of total 5 layers. As illustrated in Fig. 3(b), this model has 3 dense layer and 2 dropout layers. The unit of first 2 dense layer is 65 and dropout rate is 0.5, those are selected after several experiments and observations.

3) LSTM for credit card fraud detection: The LSTM model that we used consists of total 5 layers including 1 LSTM layer (50 units), 2 dropout layers of rate 0.5, and 2 dense layers (the first one having 65 units). The architecture is depicted in Fig. 3(c). All the optimal hyper-parameters are chosen according to the experimental analysis.

C. Incorporating Federated Learning

One of our most important target of this work is to analyze all of the models' performance after incorporating them with federated learning approach. As the conventional centralized approach cannot provide data privacy, federated learning has been indulged along with the conventional approach.

Federated learning is the decentralized form of conventional machine learning. As depicted in Fig.4, the central sever sends the initial model weight to the clients and the model is trained locally on each of the clients. Afterwards, the updated weights are sent back to the central server and it aggregates all the model weights using *FedAvg* [22]. In this way the global model gets updated at each communication round. We have conducted our experiments with all the three models using this

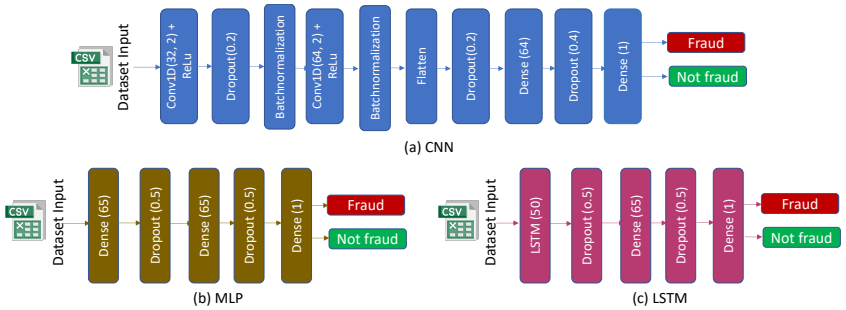

Fig. 3: Architecture of the proposed models (a) CNN model (b) MLP model (c) LSTM model

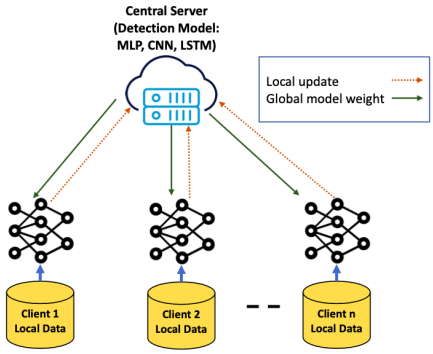

Fig. 4: Typical Federated Learning architecture

FedAvg approach. As per our experiments, 50 communication round seemed effective to attain a higher detection rate for each of the models. It has been observed that for CNN and MLP, 3 clients performed better whereas for LSTM, 2 clients performed better. As per experimental analysis the learning rate of 0.0001 is chosen for CNN and MLP, however the default learning rate (0.001) is used for LSTM.

IV. Result Analysis

The result analysis of the whole work has been done considering a few aspects. Firstly, the classical deep learning approach has been considered and then federated learning approach has been taken into account. We also present the impact after applying different sampling techniques of data. Finally, a comparison with diversified existing model is done to evaluate the proposed approach.

A. Experimental Result: Deep Learning Approach

The performance of the applied three models (LSTM, MLP, CNN) is depicted in Fig. 5. Here, the centralized approach is considered without using federated learning technique. Fig. 5(a) demonstrates the performance on original dataset. Its clearly visible that the accuracy score is very high but recall (detection rate) and other metrics do not seem to have a satisfactory outcome. From 5(b), we observe the performance after applying ROS on the dataset. It seems to have better performance regarding all the metrics. After applying SMOTE on the dataset, CNN and LSTM achieved very satisfactory result but LSTM seems to have a degradation in the performance which is depicted in 5(c). Performance after applying RUS and NMU on the dataset is demonstrated in 5(d) and 5(e) respectively. It is very evident that the models' performance deteriorates a bit here compared to the oversampling techniques. Still the overall performance of CNN and MLP was quite satisfactory in all of the cases.

B. Experimental Result: Federated Learning Approach

Federated learning is incorporated with all of the individual models and the performance of the models is demonstrated in 6. From 6(a), it is very much obvious that on the original dataset, the models do not perform well regarding all the metrics except accuracy. The reason behind the high accuracy is that the original dataset is extremely imbalanced and over 99% data belongs to the benign class and that's why it is quite possible for the model to detect all the benign samples, whereas other metrics including detection rate is not that much convincing. Fig. 6(b) and 6(c) depicts the performance of FL approach after applying ROS and SMOTE respectively. CNN and MLP both had a satisfactory performance using the oversampling techniques. Fig. 6(d) and 6(e) represent the performance of FL approach after applying RUS and NMU respectively. Visibly the performance of the models degraded after applying undersampling. This is happening because the undersampling methods might discard some of the samples having potential or significant feature sets. However, the overall performance of the proposed CNN model after applying FL approach outperformed the remaining models regarding all evaluation metrics.

C. Experimental Result: Considering Distinct Sampling Techniques

Total four distinct sampling methods are applied to analyse the models' performance namely ROS, SMOTE, RUS, and NMU. We can observe the difference in the *Pearson Correlation* matrix from Fig. 7 after applying sampling methods to the data. It is clearly visible that after sampling the data, the correlation among the features increased. TABLE II shows the accuracy comparison of the models based on the sampling techniques. It is visible that the accuracy is highest without any sampling technique. The oversampling techniques(SMOTE and RO) also attained quite good accuracy. However, the undersampling tecnique had a negative impact on accuracy

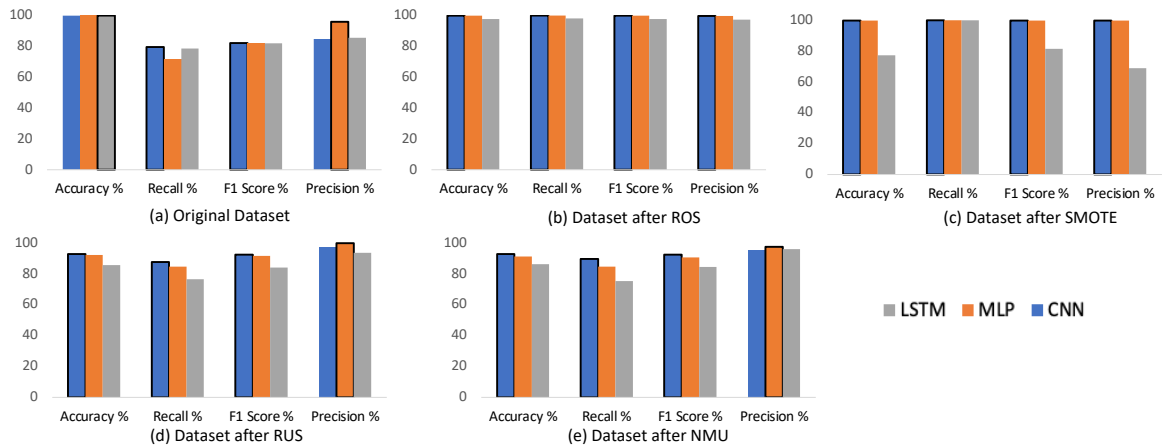

Fig. 5: Performance comparison of the proposed model while using traditional centralized DL approach considering diversified sampling techniques

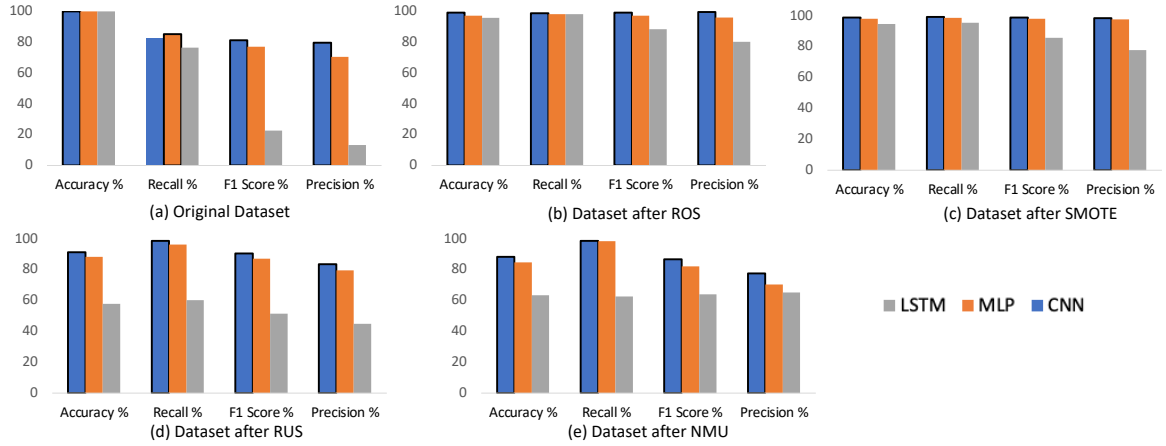

Fig. 6: Performance comparison of the proposed model while using FL approach considering diversified sampling techniques

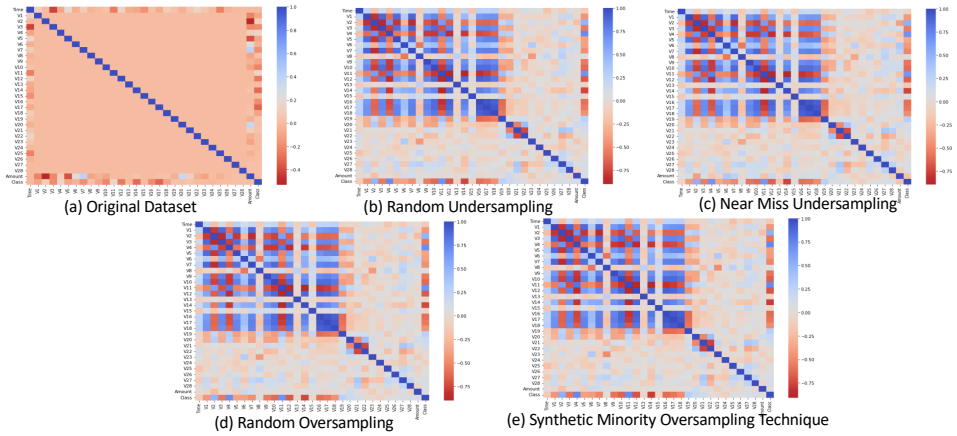

Fig. 7: Pearson correlation matrix of all the features: comparison among original dataset and datasets after four distinct sampling methods

specially considering the LSTM model. Its because of losing some significant information due to undersampling of the data.

TABLE III depicts the impact of different sampling methods on recall or detection rate. Recall is the most important metric to be considered in this type of problem because the detection rate actually denotes how well the model performs in detecting a fraud or anomaly. We can clearly observe that CNN and MLP had the best recall value of 99.51% and 98.77% respectively when SMOTE is applied whereas LSTM had the best recall value of 98.20% with RO technique. LSTM didn't perform well considering original data and undersampling techniques.

TABLE IV presents the precision value of the models considering the sampling methods. In this case CNN and LSTM performed best with RO technique whereas MLP performed best with SMOTE technique. CNN, MLP, and LSTM had the best precision value of 99.61%, 97.78%, and 80.23% respectively.

The impact of sampling methods on F1 score is shown in TABLE V which is also a very significant metric to be considered to evaluate the overall performance of a model. In this regard, CNN and LSTM performed best with RO technique attaining F1 score of 99.19% and 88.31% respectively. MLP had the best F1 score of 98.28% with SMOTE technique.

TABLE II: Experimentation on sampling techniques- **Accuracy(%)**

Sampling technique	CNN	MLP	LSTM
No sampling	**99.94**	**99.92**	**99.84**
Random oversampling	99.18	97.10	95.78
SMOTE	99.11	98.28	94.82
Random undersampling	91.37	88.33	57.87
Nearmiss undersampling	88.33	84.77	63.45

TABLE III: Experimentation on sampling techniques- **Recall(%)**

Sampling technique	CNN	MLP	LSTM
No sampling	82.98	85.18	76.47
Random oversampling	98.77	98.14	**98.20**
SMOTE	**99.51**	**98.77**	95.67
Random undersampling	98.79	96.30	60.27
Nearmiss undersampling	98.70	98.57	62.75

TABLE IV: Experimentation on sampling techniques- **Precision(%)**

Sampling technique	CNN	MLP	LSTM
No sampling	79.59	70.41	13.27
Random oversampling	**99.61**	96.02	**80.23**
SMOTE	98.71	**97.78**	77.97
Random undersampling	83.67	79.59	44.89
Nearmiss undersampling	77.55	70.41	65.31

D. Comparison with Diversified Existing Models

Some recent existing work on the same dataset has been considered and the performance comparison of the methods with the proposed models is represented in TABLE VI. Very few researches have been conducted in this particular problem using federated learning. Most of the works are focused on applying the conventional centralized ML approach. From this table, it is observable that we definitely did not attain the highest accuracy compared to other models. For example, Baaga et al. [23] had the best accuracy among all. However, if we observe the recall value of all the models, its quite evident that our models outperformed the other existing models. Even considering Precison and F1 score, proposed CNN and and MLP with FL method performed way better than the existing models. It is obvious that our proposed approach attained a balanced outcome considering all the evaluation metrics.

TABLE V: Experimentation on sampling techniques- **F1 score(%)**

Sampling technique	CNN	MLP	LSTM
No sampling	81.25	77.09	22.61
Random oversampling	**99.19**	97.07	**88.31**
SMOTE	99.11	**98.28**	85.92
Random undersampling	90.61	87.15	51.46
Nearmiss undersampling	86.86	82.14	64.00

V. Discussion and Future Direction

The intention of this investigation is to implement a robust and privacy-preserving credit card fraud detection system. The total experimental analysis can be contemplated in three aspects: combatting extreme data imbalance issue, maintaining data privacy of local clients, and attaining a good detection rate. Correspondingly, we try to address each of the issues with our thorough analysis. Four different sampling methods (SMOTE, ROS, NMU, RUS) are applied in order to mitigate the data imbalance. Three different DL models are taken into consideration throughout this work, namely; CNN, MLP, and LSTM. The most optimal architectures with suitable hyper-parameters are chosen through several experimentations and trials. The FL-based approach is adopted while considering all three optimal DL models as the global shared FL model. Subsequently, this approach turns out to be an effective way of fraud detection with a high detection rate. The data privacy issue and as well as the data imbalance problem are also resolved consequently. We emphasize the detection rate (recall) of the model and according to our experimental results, a balance of all the evaluation metrics is achieved unlike many of the state-of-the-art models.

This experimentation is confined to a single dataset that is used as a base for the implementation of a credit card fraud detection system. However, the real-world dataset would have more diversified features, and adding multiple datasets in different FL clients can mimic the actual scenario seamlessly. Furthermore, significant research scope is available in FL area in terms of communication overhead, model aggregation, client selection etc. We may consider to work on the improvement of aforementioned areas.

VI. Conclusion

Credit card frauds are affecting the financial sector and the stakeholders with an undeniable implication. This paper presents a thorough analysis in implementing a robust and privacy preserving credit card fraud detection model. To accomplish this, federated learning is incorporated with

TABLE VI: Performance comparison with state-of-the-art methods

Ref.	Year	Method used	Accuracy (%)	Recall (%)	Precision (%)	F1 Score (%)
Alfaiz et al. [10]	2021	AllKNN-CatBoost	99.96	95.91	80.28	87.40
Asha et al. [13]	2021	ANN	99.92	76.19	81.15	–
Bagga et al. [23]	2020	Pipelining	**99.99**	86.00	92.00	92.50
Fiore et al. [16]	2017	GAN	99.96	70.23	95.83	81.06
Forough et al. [5]	2020	Ensemble LSTM	–	74.38	90.77	81.16
Trivedi et al. [12]	2020	Random Forest	94.99	95.12	95.98	95.11
Varmedja et al. [17]	2019	Random Forest	99.96	81.63	96.38	–
Suvarna et al. [20]	2020	FL with RBF	94.00	–	–	–
Yang et al. [18]	2019	FL with CNN	–	–	–	95.34
Hema et al. [24]	2021	Random Forest	99.95	79.20	91.95	85.10
Proposed model	–	**FL with CNN**	99.11	**99.51**	**98.71**	**99.11**
		FL with MLP	98.28	98.77	97.78	98.28
		FL with LSTM	95.78	98.20	80.23	88.31

CNN, MLP and LSTM models with the aim of mitigating the problem of data privacy in this area. Furthermore, to overcome the data skewness due to excessive imbalance of two data classes, four different sampling methods are also considered. According to the experimental results, FL with CNN attained the highest detection rate of 99.51% with SMOTE approach. MLP and LSTM with FL also achieved satisfactory detection rate and outperformed the existing models. In this paper, The inspection of FL with the three DL models demonstrates a feasible way of employing a substantial credit card fraud detection system.

ACKNOWLEDGMENT

Part of this study was funded by the ICSCoE Core Human Resources Development Program and MEXT Scholarship, Japan.

REFERENCES

[1] M. Zareapoor, P. Shamsolmoali, *et al.*, "Application of credit card fraud detection: Based on bagging ensemble classifier," *Procedia computer science*, vol. 48, no. 2015, pp. 679–685, 2015.

[2] A. C. Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, "Feature engineering strategies for credit card fraud detection," *Expert Systems with Applications*, vol. 51, pp. 134–142, 2016.

[3] H. Ravichandran, "14 Ways Scammers Can Steal Your Credit Card Numbers in 2023." https://www.aura.com/learn/how-do-people-steal-credit-card-numbers, 2023. [Online; accessed 26-February-2023].

[4] 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)*, pp. 2747–2752, IEEE, 2017.

[5] J. Forough and S. Momtazi, "Ensemble of deep sequential models for credit card fraud detection," *Applied Soft Computing*, vol. 99, p. 106883, 2021.

[6] V. N. Dornadula and S. Geetha, "Credit card fraud detection using machine learning algorithms," *Procedia computer science*, vol. 165, pp. 631–641, 2019.

[7] 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, pp. 1–17, 2022.

[8] S. Jha, M. Guillen, and J. C. Westland, "Employing transaction aggregation strategy to detect credit card fraud," *Expert systems with applications*, vol. 39, no. 16, pp. 12650–12657, 2012.

[9] A. C. Bahnsen, A. Stojanovic, D. Aouada, and B. Ottersten, "Improving credit card fraud detection with calibrated probabilities," in *Proceedings of the 2014 SIAM international conference on data mining*, pp. 677–685, SIAM, 2014.

[10] N. S. Alfaiz and S. M. Fati, "Enhanced credit card fraud detection model using machine learning," *Electronics*, vol. 11, no. 4, p. 662, 2022.

[11] A. Thennakoon, C. Bhagyani, S. Premadasa, S. Mihiranga, and N. Kuruwitaarachchi, "Real-time credit card fraud detection using machine learning," in *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pp. 488–493, IEEE, 2019.

[12] N. K. Trivedi, S. Simaiya, U. K. Lilhore, and S. K. Sharma, "An efficient credit card fraud detection model based on machine learning methods," *International Journal of Advanced Science and Technology*, vol. 29, no. 5, pp. 3414–3424, 2020.

[13] R. Asha and S. K. KR, "Credit card fraud detection using artificial neural network," *Global Transitions Proceedings*, vol. 2, no. 1, pp. 35–41, 2021.

[14] J. I.-Z. Chen and K.-L. Lai, "Deep convolution neural network model for credit-card fraud detection and alert," *Journal of Artificial Intelligence*, vol. 3, no. 02, pp. 101–112, 2021.

[15] 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.

[16] U. Fiore, A. De Santis, F. Perla, P. Zanetti, and F. Palmieri, "Using generative adversarial networks for improving classification effectiveness in credit card fraud detection," *Information Sciences*, vol. 479, pp. 448–455, 2019.

[17] D. Varmedja, M. Karanovic, S. Sladojevic, M. Arsenovic, and A. Anderla, "Credit card fraud detection-machine learning methods," in *2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH)*, pp. 1–5, IEEE, 2019.

[18] W. Yang, Y. Zhang, K. Ye, L. Li, and C.-Z. Xu, "Ffd: A federated learning based method for credit card fraud detection," in *Big Data–BigData 2019: 8th International Congress, Held as Part of the Services Conference Federation, SCF 2019, San Diego, CA, USA, June 25–30, 2019, Proceedings 8*, pp. 18–32, Springer, 2019.

[19] W. Zheng, L. Yan, C. Gou, and F.-Y. Wang, "Federated meta-learning for fraudulent credit card detection," in *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pp. 4654–4660, 2021.

[20] R. Suvarna and A. M. Kowshalya, "Credit card fraud detection using federated learning techniques," *International Journal of Scientific Research in Science, Engineering and Technology*, vol. 7, pp. 356–367, 2020.

[21] M. L. G. ULB, "Credit Card Fraud Detection." https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud. [Online; accessed 20-December-2022].

[22] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*, pp. 1273–1282, PMLR, 2017.

[23] S. Bagga, A. Goyal, N. Gupta, and A. Goyal, "Credit card fraud detection using pipeling and ensemble learning," *Procedia Computer Science*, vol. 173, pp. 104–112, 2020.

[24] A. Hema, "Machine learning methods for discovering credit card fraud. irjcs:: International research journal of computer science, volume viii, 01-06," 2020.