Available online at www.qu.edu.iq/journalcm
JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS
ISSN:2521-3504(online) ISSN:2074-0204(print)

Enhanced Fraudulent Detection Using Isolation Forest and Multi-Cluster Deep Learning

Hayder K. Fatlawi*

Center of Information Technology Research and Development, University of Kufa, Najaf, Iraq. Email: hayder.fatlawi@uokufa.edu.iq

ARTICLE INFO

Article history:
Received: 30 /12/2024
Rrevised form: 25 /1/2025
Accepted : 2 /2/2025
Available online: 30 /3/2025

Keywords:

Anomaly Detection,

Deep Learning

Ensemble Machine Learning

ABSTRACT

The anomaly detection problem has received increasing research interest due to the negative effects of fraud on several essential systems. Since Iraq is currently moving towards activating electronic financial transactions in all government ministries and private trade, this follows an increase in the risks of financial fraud. This research aims to improve the ability to identify fraudulent financial operations based on a multistage classification model that utilizes several machine learning techniques. It focused on avoiding the outlier instances that can affect the performance of the learning process by utilizing Isolation Forest. The implementation of the proposed model indicates that the ensemble size has no significant impact on its performance while increasing the number of clusters has led to a decline in performance. The experimental results with real datasets produced an F1-score of 99.097 compared to 80.5 and 74.65 with typical DNN, K-NN, and confirmed its preference compared to many popular classifiers and recent research articles.

MSC..

https://doi.org/10.29304/jqcsm.2025.17.11964

1. Introduction

The anomaly detection problem has received increasing research interest due to the negative effects of fraud on various systems, such as computer networks and electronic payment systems. In financial transactions, monitoring and preventing fraud remains a real challenge. According to the US government report [1], more than 64,000 fraudulent transactions were reported in 2023. These financial fraud offences had a median loss of $116,545, and 10.5% of these amounts were greater than $550,000. On the other hand, Iraq is currently moving towards activating electronic financial transactions in all government ministries and private trade. An increase follows the increase in electronic payment operations in terms of the risks of financial fraud. This research aims to improve the ability to identify fraudulent financial operations based on a multistage classification model that utilizes machine learning techniques.

Machine learning is concerned with enhancing the ability of computer systems to recognize patterns. Machine learning techniques are classified based on the availability of the target value into supervised, unsupervised, and semi-supervised learning labeled. Supervised learning aims to classify the given data based on previously labeled target values [2], [3]. Classification techniques face several challenges in dealing with credit card data, including imbalanced class, verification latency, and concept drift [4]. Many methods [5–13] have been presented for handling

∗Corresponding author

Email addresses: *hayder.fatlawi@uokufa.edu.iq*

Communicated by ‘sub etitor’

these challenges, and most focus on improving the classifier's effectiveness by overcoming the imbalanced distribution of the class label. In this research, we utilize the multi-cluster ensemble method for enhancing deep learning classification in addition to many preprocessing steps.

Under-sampling approaches are vastly utilized in the classification process to handle imbalanced data distribution by reducing the ratio of the most frequent class (i.e., major class). On the other hand, outlier data samples can lead to a negative impact on a classifier's performance by introducing noise and distorting the learning process. These samples could generate skewed decision boundaries, overfitting, and reduced class imbalance handling. Therefore, in this work, an under-sampling approach was performed by eliminating all data samples if these points were recognized as outliers by Isolation Forest and they belonged to the major class (i.e., not fraud).

The rest of the article is arranged as follows. Section 2 describes the literature review, including many related works. Section 3 introduces the methodology that contains the proposed techniques. Section 4 demonstrates the results obtained from the proposed model's implementation, in addition to discussing the findings of these results. Finally, Section 5 summarizes the conclusion of this paper.

II. Literature Review

A customized Bayesian Network Classifier BNC was proposed by [14] for fraud detection in credit card transactions based on a Hyper-Heuristic Evolutionary Algorithm (HHEA). Their model utilized a hill-climbing search method for creating the BNC, in addition to the Heckerman-Geiger-Chickering (HGC) metric for evaluating the network structure. The results of their implementation showed promising performance compared to many other algorithms in terms of classification accuracy and economic efficiency metrics. Another hybrid method has been presented by [15] to overcome the imbalance class distribution. It starts with excluding some outliers from rare class instances and many majority instances, then applying a non-linear classifier. They proposed Dynamic Weighted Entropy to evaluate the classification quality outperforming many state-of-the-art classifiers.

[16] proposed a Recurrent Neural Network classifier based on Long Short-Term Memory provided by an attention mechanism. Implementing their model led to more efficient performance compared with the other three popular classifiers. In [17], the performance of several machine learning techniques has been evaluated for fraud detection using the European card benchmark dataset. Their evaluation included Decision Trees, Random Forests, Support Vector Machine SVM, Logistic Regression, XG Boost, and Extreme Learning. According to their results, the XGBoost classifier performed better in terms of many evaluation metrics.

A collaborative training framework was proposed in [18] by combining federated learning and graph neural networks. It includes mapping financial institutions' datasets to transaction graph representation using weighted feature similarity. The demonstration results showed that their model had the highest recall and Area under the curve AUC performance compared to the baseline model. The researchers in [19] presented an optimization-based technique for detecting fraudulent transactions. It starts with applying quantile normalization to the input data. The most valuable features were chosen using diverse distance measures. Using Bootstrapping, an augmentation process is applied to the resulting features subset, and then SpinalNet is used for fraud detection. The tuning SpinalNet classifier was enhanced by utilizing the JNBO model.

The study in [20] utilized fuzzy logistic regression to develop a real-time framework for addressing the nonstationary transaction changes, fraud attributes, and imbalanced class distribution. Their methodology indicates robust results, even on small data samples, in determining both fraudulent and non-fraudulent financial transactions. A detect framework is presented in [21] and includes both user and transaction attributes. It utilized a neural network classifier with an unsupervised clustering undersampling technique. Based on their results, the model demonstrated qualified performance compared to other machine learning classifiers.

Generally, fraud detection models tend to oversample the fraud class as it represents the minor class. The increase in fraud data instances could lead to an increase in the false positive rate. This raises the cost of manual review for normal transactions, which is incorrectly classified as fraud. Striking the balance between increasing the quality of fraud data classification and reducing false alerts remains a considerable challenge. Our work utilizes concepts and techniques from outlier detection, data augmentation, unsupervised learning, and supervised deep learning to achieve better fraud detection performance.

III. Proposed Techniques

In general, the data needs several pre-processing steps in order to apply the proposed classification model. These steps include outlier detection, important attribute selection, and extracting or generating new attributes. The first step in this stage is to detect outlier instances in both legal and fraud classes. Isolation Forest presented in [22] has been used to perform this task. It isolates the outlier data samples using many binary trees generated by choosing a random attribute and split point in each node. Isolation Forest depends on the average depth of all data points to detect the outliers that have a depth smaller than the average. All outlier data points with fraud class resulting from the isolation forest will be combined with the normal data (not outlier points) to be used in the next steps. This combination aims to maintain the existence of fraud class in both the outlier and non-outlier instances.

The second step in the preprocessing stage is to reduce dimensionality using Principal Component Analysis PCA. A statistical-based features transformation technique produces fewer principal components from the larger original feature set. PCA relied on the concept of eigenvector and eigenvalues; in our proposed model, the required ratio of information to be preserved is used as a user parameter instead of the number of required principal components. This process aims to reduce the computational complexity in the next stages of the proposed model, especially training deep learning classifiers. Since the problem of identifying fraudulent behavior depends on the precise classification of the rare class, it may require generating new data instances belonging to this class.

A Generative Adversarial Network GAN is used in this step to generate artificial samples based on the original ones using Generator and Discriminator. The Generator starts the adversarial training of GAN by generating data points based on random noise that is selected from one of the popular data distributions (such as uniform or Gaussian). The aim of the generation is to create data samples very similar to the original ones in which the Discriminator could not recognize between them correctly. The second part of GAN is the Discriminator, which receives both the original and the synthetic data samples. It aims to classify each data sample correctly as original or artificial. The final synthetic data points produced by GAN will imitate the original distribution of the minor class.

The major step in the proposed model is to create an ensemble model containing M base learners, where m represents the ensemble size and is defined by the user. In each base learner, K clusters are created using the K-means algorithm. Then, for each cluster, the Deep Neural Network DNN classifier is trained using data instances in that cluster. The model also included applying cross-validation to avoid any bias in data distribution during data splitting for training and testing tasks. All the steps of Multi-Cluster Ensemble Deep Learning (MCE-DL) are illustrated in Fig 1.

IV. Results And Discussion

Information extracted from the usage data of credit cards represents an essential source for fraud monitoring and prevention tasks. The analysis of this data faces various challenges, such as anomalies, verification latency, and imbalanced target class distribution [4]. In this implementation, a real benchmark dataset was used to evaluate the performance of the proposed model. This evaluation included comparing several popular classifiers and multiple values for ensemble size and number of clusters.

A. Experimental Setup

The European Cardholders Credit Card Dataset [23], [24] that is utilized in our evaluation includes transactions of two days in 2013, with only 492 data instances as fraud and 284,807 normal (not fraud) instances, as summarized in Table 1. Therefore, it can be considered highly unbalanced. All values in this dataset are numeric, and the attribute names are not declared for privacy reasons. Only the first attribute, named 'Time', refers to the time between a specific transaction and the first one in the dataset. This attribute has been eliminated in our implementation as it is not relevant. Another attribute named 'Amount' preserves its original name and contributes to the training process. This implementation relied mainly on Python programming language and its libraries, such as Pandas, Scikitlearn, and TensorFlow.

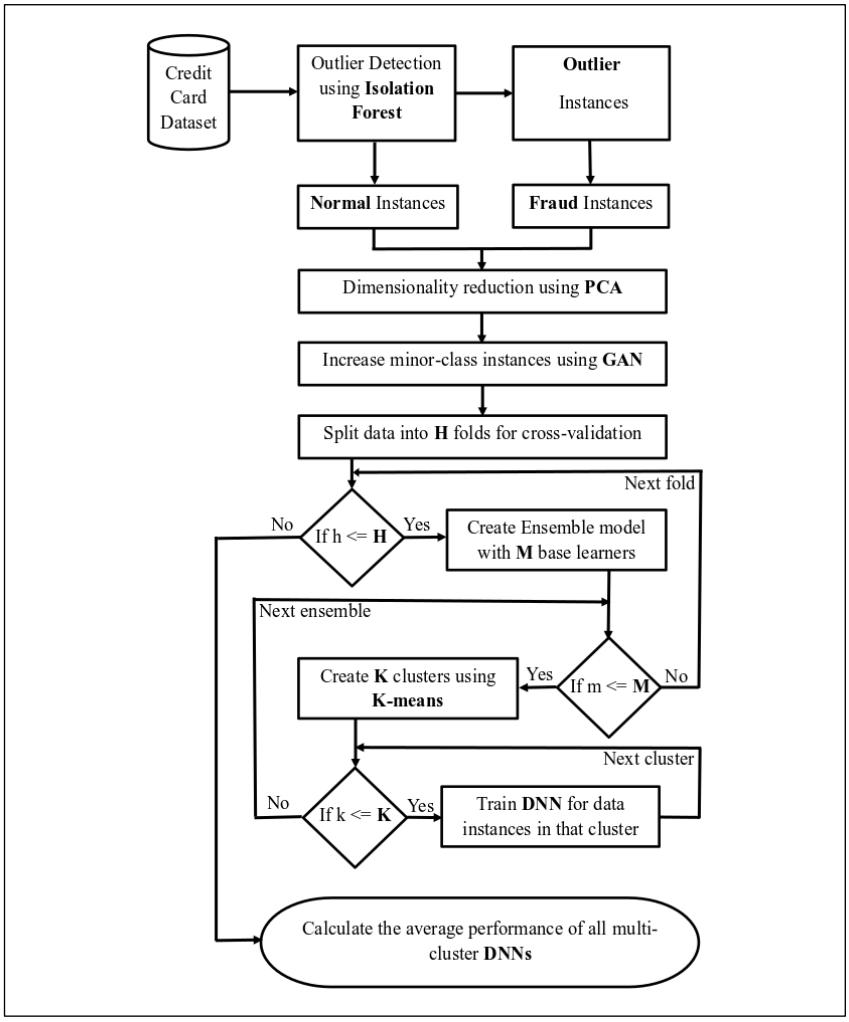

Fig. 1- Diagram of the proposed model MCE-DL

TABLE 1: Distribution of Credit Card Dataset

Outlier/Fraud	Not Fraud	Fraud
Normal	278978	132
Outlier	5337	360

B. Experimental Results

The proposed model's first step was utilizing Isolation Forest for anomaly detection. Table 1 shows the distribution of fraud transactions within the normal and outlier instances. The fraud data sample ratio represents only 0.00047 from the normal (not outlier) subset. This ratio increases significantly in outlier data to reach 0.0631. At the end of this step, outlier data that was not labeled as fraudulent transactions were excluded. Then, PCA dimensionality reduction is applied to the produced dataset. The ratio of mutual information was set to 80% to preserve as much variance as possible. The obtained number of principal components was 20, which represents the new attribute set.

The impact of principal components on the original attribute set has been illustrated in the heat map in Fig 2. In this figure, the first principal component positively impacted the attribute named 'Amount'; it also had a noticeable negative impact on the second property in the original data 'V2'. Despite the ambiguity caused by not revealing the original names of the attributes, the figure indicates an apparent impact of the first ten principal components pc1-

pc10 on the first ten (v1-v10) and the last ten (v20-Amount) original attributes. In contrast, the effect of the second ten components (pc11-pc20) on the middle ten original properties (v1-v19) increases.

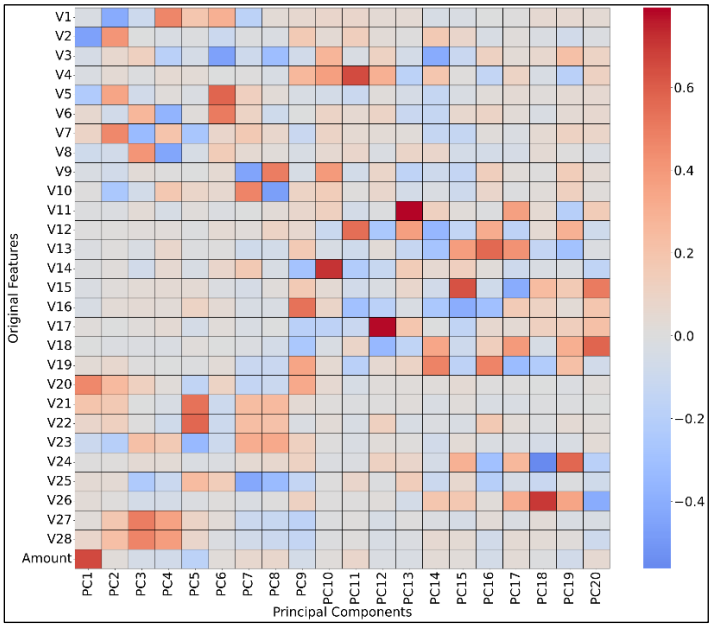

Fig. 2- A heat map of the impact of principal components on the original features

The preference of MCE-DL compared to four popular classifiers is illustrated in Fig 3. The comparison included applying K-Nearest Neighbor, Bagging classifier, AdaBoost, and DNN. Five metrics were used in this comparison: Recall, F1-score, Matthews's Correlation Coefficient MCC, Critical Success Index CSI, and Fowlkes-Mallows Index FMI. With regard to all those metrics, the proposed model performed noticeably better than the other four classifiers. The reason for this advantage is that the proposed method avoids the majority of anomalous elements (mainly from the non-rogue class) and uses several base classifiers based on data clustering. Despite the acceptable performance of the DNN deep learning classifier and the Bagging classifier according to F1-score, MCC, and FMI, the comparison produced low performance in terms of Recall and FMI in both classifiers. According to CSI, the K-Nearest Neighbors K-NN classifier had the lowest performance, reaching less than 60%.

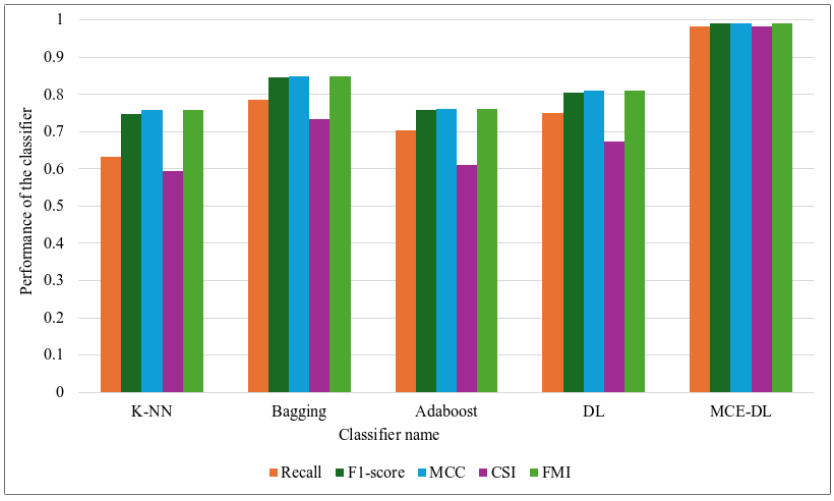

Fig. 3- A comparison of the performance of MCE-DL with four popular Classifiers

The main factor affecting the data clustering process in the K-means algorithm is K, which represents the number of expected clusters in the data and, thus, the number of random centers chosen at the beginning of the process. Fig 4 illustrates the tracking of MCE-DL performance with multiple values of K. In terms of Accuracy, the performance of the proposed method wasn't affected by the increase in K value. On the other hand, we can observe a decline in the MCE-DL performance when K increases; although the performance has improved with some values, it remains lower than it was with two clusters.

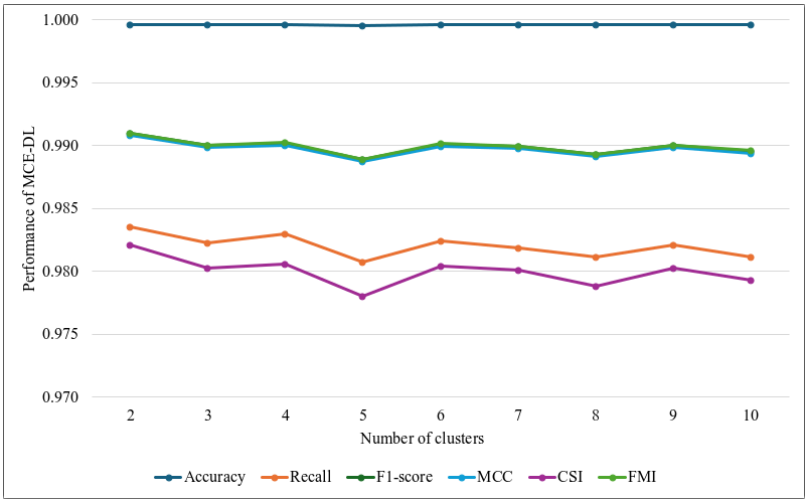

Fig. 4- Tracking the performance of MCE-DL with multiple values for the number of K-means clusters

Another essential parameter in MCE-DL is the number of base learners, which controls the size of the ensemble model. Fig 5 indicates that this parameter has no significant impact on the performance of MCE-DL. Regarding Recall and CSI, the performance has declined slightly after increasing the ensemble size, and the best result was with five base learners. On the other hand, the stability of the proposed model can be observed in the confusion matrix with the three folds cross-validation which has been summarized in Table 2.

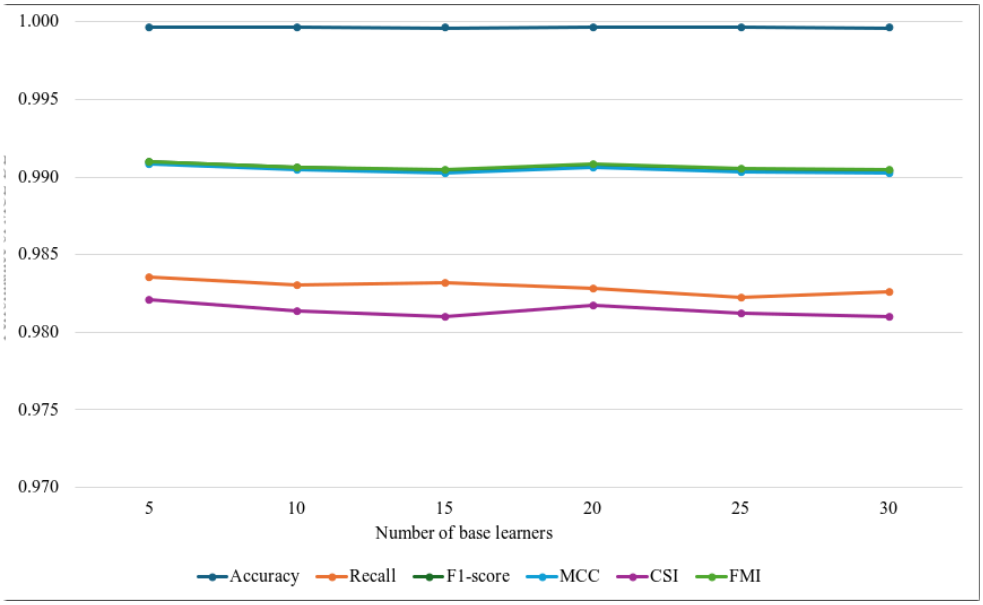

Fig. 5- Tracking the performance of MCE-DL with multiple values for the number of base learners

TABLE 2: Confusion matrix of the proposed model with three folds cross-validation

Actual values / Predicted values		Predicted values					
		Fold 1		Fold 2		Fold 3	
		Legal	Fraud	Legal	Fraud	Legal	Fraud
Actual values	Legal	93020	3	92996	3	92954	2
	Fraud	27	1747	31	1767	31	1809

Many researchers have proposed techniques to improve the classification of financial frauds and have used the same European Cardholders Dataset. Table 3 summarizes the performance comparison of our proposed method with the results of recent research papers. According to the available results, our process in MCE-DL has a clear advantage.

TABLE 3: Comparison of the performance of our proposed model with ten articles that used credit card dataset

Article	Accuracy	Recall	F1-score	Precision	MCC
[25]	-	75.53	85.29	97.95	-
[26]	99.46	77.7	33.24	-	-
[27]	90.36	-	-	-	-
[28]	97.16	97.82	95.98	-	-
[29]	99.95	-	85	-	-
[30]	-	78.9	-	92.74	-
[31]	99.941	-	-	-	82.3
[32]	96.64	93.62	-	-	-
[33]	-	94.8	-	-	-
[34]	98.9	91	-	97	-
MCE-DL	**99.966**	**98.356**	**99.097**	**99.997**	**99.083**

VI. Conclusions

Identifying fraudulent transactions in financial transactions contributes to reducing financial losses and enhancing users' confidence in electronic financial systems. In this research, a multi-stage classification model has been developed based on Multi-Cluster Deep Learning and Isolation Forest. Applying the proposed model to a real and big dataset produced better performance than four popular classifiers. Regarding the F1-score, MCE-DL reached 99.097 compared to 80.5 and 74.65 with typical DNN, K-NN. The experimental results also indicated a decline in MCE-DL performance when the number of clusters increases; also, it confirmed that the ensemble size parameter

has no significant impact on the performance of MCE-DL. The applied pre-processing steps contributed to improving the efficiency of the proposed model by pre-excluding part of the outlier data and generating additional data similar to the fraud data to reduce the effect of the unbalanced distribution of classes. The proposed model is specifically applied to a binary classification task and can be developed to handle multiple classes and regression tasks. Also, the model suffers from multiple parameters whose values are required to be specified by the user, and this can be overcome using optimization techniques.

References

[1] "Quick facts credit card and other financial instrument fraud," accessed: 2024-12-03. [Online]. Available: https://www.ussc.gov.
[2] J. Leskovec, A. Rajaraman, and J. D. Ullman, Mining of massive data sets. Cambridge university press, 2020.
[3] J. Han, J. Pei, and H. Tong, Data mining: concepts and techniques. Morgan kaufmann, 2022.
[4] S. S. Sulaiman, I. Nadher, and S. M. Hameed, "Credit card fraud detection challenges and solutions: A review." Iraqi Journal of Science, vol. 65, no. 4, 2024.
[5] H. Ahmad, B. Kasasbeh, B. Aldabaybah, and E. Rawashdeh, "Class balancing framework for credit card fraud detection based on clustering and similarity based selection (sbs)," International Journal of Information Technology, vol. 15, no. 1, pp. 325–333, 2023.
[6] T. A. Olowookere and O. S. Adewale, "A framework for detecting credit card fraud with cost-sensitive meta-learning ensemble approach," Scientific African, vol. 8, p. 00464, 2020.
[7] R. Van Belle, B. Baesens, and J. De Weerdt, "Catchm: A novel network-based credit card fraud detection method using node representation learning," Decision Support Systems, vol. 164, p. 113866, 2023.
[8] P. Juszczak, N. M. Adams, D. J. Hand, C. Whitrow, and D. J. Weston, "Off-the-peg and bespoke classifiers for fraud detection," Computational Statistics & Data Analysis, vol. 52, no. 9, pp. 4521–4532, 2008.
[9] I. Benchaji, S. Douzi, B. El Ouahidi, and J. Jaafari, "Enhanced credit card fraud detection based on attention mechanism and lstm deep model," Journal of Big Data, vol. 8, pp. 1–21, 2021.
[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. 16 400–16 407, 2022.
[11] E. F. Malik, K. W. Khaw, B. Belaton, W. P. Wong, and X. Chew, "Credit card fraud detection using a new hybrid machine learning architecture," Mathematics, vol. 10, no. 9, p. 1480, 2022.

[12] M. S. Derweesh, S. A. H. Alazawi, and A. H. Al-Saleh, "Multi level deep learning model for network anomaly detection," Journal of Al-Qadisiyah for Computer Science and Mathematics, vol. 15, no. 4, pp. 8–19, 2023.

[13] W. S. Mahdi and A. T. Maolood, "Banking intrusion detection systems based on customers behavior using machine learning algorithms: Comprehensive study," Journal of Al-Qadisiyah for computer science and mathematics, vol. 12, no. 4, pp. Page–1, 2020.

[14] A. G. de Sa, A. C. Pereira, and G. L. Pappa, "A customized classification algorithm for credit card fraud detection," Engineering Applications of Artificial Intelligence, vol. 72, pp. 21–29, 2018.

[15] Z. Li, M. Huang, G. Liu, and C. Jiang, "A hybrid method with dynamic weighted entropy for handling the problem of class imbalance with overlap in credit card fraud detection," Expert Systems with Applications, vol. 175, p. 114750, 2021.

[16] J. F. Roseline, G. Naidu, V. S. Pandi, S. A. alias Rajasree, and N. Mageswari,"Autonomous credit card fraud detection using machine learning approach," Computers and Electrical Engineering, vol. 102, p. 108132, 2022.

[17] M. A. Islam, A. A. Imran, M. H. Rahman, M. A. H. Pabel, B. K. Mishra, and K. Basu, "Analysis and performance evaluation of credit card fraud by multi-model ml," in 2024 3rd International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE). IEEE, 2024, pp. 1–7.

[18] Y. Tang and Y. Liang, "Credit card fraud detection based on federated graph learning," Expert Systems with Applications, vol. 256, p. 124979, 2024.

[19] V. V. K. Reddy, R. V. K. Reddy, M. S. K. Munaga, B. Karnam, S. K. Maddila, and C. S. Kolli, "Deep learning-based credit card fraud detection in federated learning," Expert Systems with Applications, p. 124493, 2024.

[20] G. Charizanos, H. Demirhan, and D. ̇Icen, "An online fuzzy fraud detection framework for credit card transactions," Expert Systems with Applications, vol. 252, p. 124127, 2024.

[21] H. Huang, B. Liu, X. Xue, J. Cao, and X. Chen, "Imbalanced credit card fraud detection data: A solution based on hybrid neural network and clustering-based undersampling technique," Applied Soft Computing, vol. 154, p. 111368, 2024.

[22] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in 2008 eighth ieee international conference on data mining. IEEE, 2008, pp. 413–422.

[23] "The european cardholders credit card dataset," accessed: 2024-09-03. [Online]. Available: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

[24] A. Dal Pozzolo, O. Caelen, R. A. Johnson, and G. Bontempi, "Calibrating probability with undersampling for unbalanced classification," in 2015 IEEE symposium series on computational intelligence. IEEE, 2015, pp. 159–166.

[25] S. Jiang, R. Dong, J. Wang, and M. Xia, "Credit card fraud detection based on unsupervised attentional anomaly detection network," Systems, vol. 11, no. 6, p. 305, 2023.

[26] A. Alharbi, M. Alshammari, O. D. Okon, A. Alabrah, H. T. Rauf, H. Alyami, and T. Meraj, "A novel text2img mechanism of credit card fraud detection: A deep learning approach," Electronics, vol. 11, no. 5, p. 756, 2022.

[27] A. M. Babu and A. Pratap, “Credit card fraud detection using deep learning,”in 2020 IEEE Recent Advances in Intelligent Computational Systems (RAICS). IEEE, 2020, pp. 32–36.

[28] I. Sadgali, N. Sael, and F. Benabbou, “Bidirectional gated recurrent unit for improving classification in credit card fraud detection,” Indonesian Journal of Electrical Engineering and Computer Science, vol. 21, no. 3, pp. 1704–1712,

2021.

[29] M. L. Ali, A. R. Ampojwala, D. R. Sudugu, J. Marousis, T. Guerrier, and M. R. Narasareddygari, “Analysis of various machine learning models in detecting credit card fraud activities,” in Fifth International Conference on Computer Vision and Computational Intelligence (CVCI 2024), vol. 13169. SPIE, 2024, pp. 114–122.

[30] M. M. Mijwil and I. E. Salem, “Credit card fraud detection in payment using machine learning classifiers,” Asian Journal of Computer and Information Systems (ISSN: 2321–5658), vol. 8, no. 4, 2020.

[31] K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi, “Credit card fraud detection using adaboost and majority voting,” IEEE access, vol. 6, pp. 14 277–14 284, 2018.

[32] 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, 2023.

[33] L. Zheng, G. Liu, C. Yan, and C. Jiang, “Transaction fraud detection based on total order relation and behavior diversity,” IEEE Transactions on Computational Social Systems, vol. 5, no. 3, pp. 796–806, 2018.

[34] A. Abd El Naby, E. E.-D. Hemdan, and A. El-Sayed, “Deep learning approach for credit card fraud detection,” in 2021 International Conference on Electronic Engineering (ICEEM). IEEE, 2021, pp. 1–5.