Proceedings of the 6th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV-2025)
DVD Part Number: CFP25ONG-DVD; ISBN: 979-8-3315-1174-6

Credit Card Security with Generative Adversarial Networks: A Personalized Fraud Detection Framework

Nanthini R
Research scholar, Department of computer science
Vels institute of Science, Technology and Advanced Studies
Chennai, Tamil Nadu, India.
nanthini.11d@gmail.com

K. Kumutha
Assistant Professor, PG Department of Computer Applications,
Vels institute of Science, Technology and Advanced Studies
Chennai, Tamil Nadu, India.
kkumutha.scs@vistas.ac.in

***Abstract*— Banks and online retailers are being forced to use computerized fraud detection systems that mine vast transaction histories due to the recent spike in credit card-based online payment card scams. Machine learning (ML), which uses supervised binary classification algorithms that have been appropriately trained on pre-screened sample datasets in order to discriminate between fraudulent and legitimate events, appears to be one of the most effective means of identifying suspicious transactions. Generative Adversarial Networks (GANs) for content tailored generation in detecting credit card fraud (CCF) is discussed in this study. It compares the detection of CCF with and without employing GAN on four ML classifiers: Logistic Regression (LR), Random Forest (RF), XGBoost (XGB), and Naïve Bayes (NB). The results reveal that GAN with XGBoost performs better with accuracy at 98.6 %. On a general basis, GAN-created data improves the ability of ML models to identify fraud more accurately, minimizing cases of fraud not caught and improving system dependability.**

Keywords : ***Credit Card Security, Generative adversarial Networks, Personalized Content Generation, Machine Learning***

I. INTRODUCTION

The pervasiveness of digital transactions in today's world is a reflection of our strong reliance on the Internet for tasks like social media, e-learning, mobile banking, recreational activities, and e-commerce. Fraud detection is increasing its importance in today's world. The lack of data indicating fraudulent activities makes it difficult for traditional fraud detection systems to handle unbalanced data. Generative artificial intelligence (AI) can identify tiny irregularities in financial records by employing generative adversarial networks (GANs), but standard methods frequently miss them. Its ability to produce synthetic data protects the confidentiality of the actual data while also allowing for accurate evaluation of detection systems (Kulkarni et al., 2025).

Most CCFs take place online, where criminals obtain payment information through fraud, phishing, or data breaches. The main problem with the several strategies proposed to combat CCF in online transactions is the significant class imbalance in the data, which makes it more difficult to develop efficient detection algorithms. Most of the existing algorithms intended to solve class imbalance overestimate the minority class distribution. This tends to generate noisy, unrepresentative, and highly overlapping features, resulting in overfitting and inexact learning results(Ghaleb et al. (2023)

Imbalanced data is a pervasive problem in most industries, particularly in fraud detection and medical diagnosis. These data sets have an imbalanced class distribution with one class overwhelmingly more frequent than the other. This kind of imbalance will result in machine learning models developing biases toward the majority class and losing important patterns within the minority class. In an effort to curb this, various industries have opted to use deep neural networks to generate data as well as enhance model performance.Studies in these fields, however, show that balanced data outperforms imbalanced data in deep neural networks. Few studies have been conducted on the effectiveness of deep generative methods, such as GANs, in augmenting high-dimensional data, and the current methods have limitations. (Strelcenia et al., 2024). The main objectives of this study is to

- Assessing the efficiency of GANs in detecting CCF and examine how fraud detection performance is enhanced by GAN-generated synthetic data.
- Evaluate the performance of LR, RF, XGB and NB with and without GAN-generated data to determine how GANs affect various ML classifiers.

II. LITERATURE STUDY

The identification of CCF has been an issue for most researchers for a long time. Supervised learning techniques using ML and DL are used to identify CCF. A case study of CCF identification utilizing GANs to create synthetic samples and considering client distributions is offered by Langevin et al. in 2022. Analysis of two different cooperating party situations provides four possible client distributions by credit quality. Companies that have customers with better credit ratings are likely to gain more from GAN augmentation, according to this study. Relative gains of synthetic data sharing also seem to benefit banks lower on the credit spectrum disproportionately even without feature set heterogeneity. Scholars hypothesize that this is so because individuals with various patterns of consumer expenditure lack equivalent variations in interest rates. To improve the accuracy of the CCF rate, Kaur et al. (2021) develop an ensemble model based on GANs and RFs. Large-scale testing reveals that the algorithm outperforms existing methods.

Shafik et al. (2024) discuss how generative AI may transform e-commerce fraud detection and prevention. It highlights how the threat of fraudulent activity is heightened by the growing rate of data activity and online transactions across different sectors such as finance, e-commerce, and healthcare. Traditional rule-based systems lag behind in keeping pace with evolving fraud techniques, though GAI, leveraging tools like GANs and variational autoencoders, may deliver lifelike yet artificial data to detect advanced fraud schemes.In 2024, Zhao et al. introduce the use of self-attention GANs (SAGANs) to the detection of CCF. The ability of SAGANs to recognize significant patterns and features in huge transaction datasets through the application of self-attention processes results in a better understanding

and more precise identification of credit card fraud. The model may generate data that mimics actual fraudulent conduct by using GANs, which enhances and improves fraud detection algorithms.

Aftabi et al. (2023) propose a novel approach based on ensemble models and GAN that can both address the problem of a lack of non-fraudulent samples and handle the high complexity of feature space. Three distinct feature categories suggested in this study are then extracted from the annual financial statements of ten Iranian banks to produce a new dataset. A GAN and RF are used by Singh et al. (2022) to increase the precision of CCF detection rates. Several tests have shown that the method offered is better than the ones that are currently in use.

High-dimensional data has been effectively enhanced through the use of GANs and other deep generating techniques. The classifiers based on RF, Nearest Neighbour, LR, MLP, and Adaboost were trained using the novel K-CGAN technique. The classifiers outperformed earlier oversampling methods in terms of F1 score performance metrics. Experiments revealed that classifiers trained on the augmented set outperformed those trained on the original data, resulting in an effective fraud detection technique (Sterlencia et al., 2022).

III. Methodology

GAN-based personalized fraud detection architecture can be implemented through a series of steps, including data collection, preprocessing, training the GAN model, identifying anomalies, and personalized fraud risk scores. The figure 1 shows the workflow of the CCF detection process.

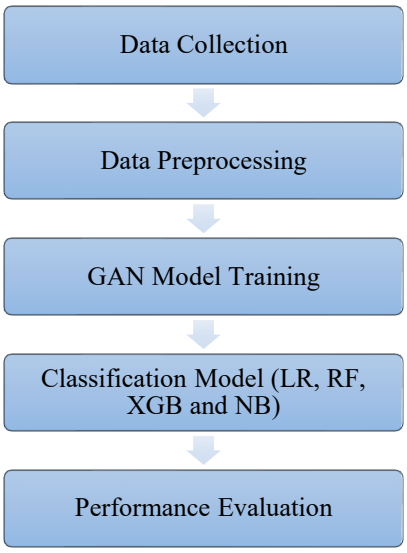

Figure 1. CCF Prediction Model with GAN

A. Data Collection

The core component of detecting CCF is transactional data. To properly detect suspicious trends, a customized fraud detection system based on GANs needs a wide range of transaction factors. Table 1 provides the attributes related to basic transactional, behavioral and external data sources.

Table 1: Credit Card Attributes

Core Transactional Attributes		
Attributes	Description	Relevancy in Fraud Detection
Transaction ID	Each transaction has a unique identifier	Maintain data integrity
Time	Transaction data and time	Time based anomalies can be detected
Amount	Transaction amour	Alert if it is unusually high or low
Merchant ID	Unique identifier for each merchant	Identifies the fraud merchants
Location	Physical or digital location	Identifies any unusual location
Device information	Mobile , Desktop etc.,	Flags unusual device
Payment Method	Swipe, Online or digital wallet like Gpay, Phonepe, Paytm etc.,)	Identifies high risk payment method
Transaction status	Approved, declined, pending,	Fraudulent transactions results in decline or chargeback etc.
Behavioral & Historical Data		
Spending trends	Frequency of transactions and volume	Sudden high in spending indicates fraud
Usual Merchants	Frequently used merchants	Unknown merchants may be detected
International transactions	Frequency of international transactions	International purchases requires verification
Time based behavior	Transaction times	Unusual hours transactions could be suspicious
Travel history of users	Past purchase locations	Detect location based fraud
Device usage history	Common devices used by the card holder	Alerts if transactions from unrecognized devices
External and derived Data Sources		
IP Address & VPN Usage	Internet address used for online transactions	VPNS and proxies may indicate fraud usage
Social Media and Open Source data	Behavioural insights form public social media	Helps verify location and activity consistency
Black list Databases	Lists of known fraudulent merchants	Blocks high-risk transactions before they occur
Machine Learning Risk Scores	AI-generated fraud probability scores	Helps prioritize high-risk transactions

B. Data Pre-processing

The key to an effective GAN-based tailored CCF detection system is preprocessing. The process of normalization involves scaling numerical transaction parameters, such as the transaction value and the time between transactions, to a uniform range. It keeps models from becoming overfit to expensive transactions, which

might not always be a sign of fraud. For standardization, this study used Z-Score normalization, which is carried out as follows, with the mean set to 0 and the variance set to 1.

$$X_{standardized} = \frac{X-\mu}{\sigma} \qquad (1)$$

The practice of developing novel, noteworthy attributes from raw transaction data to improve the accuracy of fraud detection systems is known as feature engineering.

Credit card transactions involve the transmission of sensitive personal information. Sensitive data can be tokenized in order to replace it with different tokens. It is feasible to both oversample and undersample the data in order to address the imbalance. Fraud transactions that already exist are replicated by oversampling. and fewer non-fraudulent transactions result from undersampling. Preprocessing methods such as normalization, feature engineering, anonymization, and handling skewed data are crucial for improving the accuracy and reliability of GAN-driven fraud detection models.

C. GAN Model Training

The primary goal is to further enhance the model by training a GAN to produce realistic fraudulent transactions. There are two main components to the GAN architecture:

- Generator learns to create fraudulent transactions that closely resemble real fraud patterns.
- Classifying transactions as authentic or fraudulent improves the discriminator's capacity to detect fraud.

The training method uses the preprocessed transactional data as input. The Discriminator learns to recognize the real fraud cases that the Generator tries to create.

GAN is capable of producing extremely realistic fraudulent transactions. Finally, to improve fraud detection models, the dataset has been enriched to include the simulated fraudulent transactions. The whole GAN system is shown in figure 2

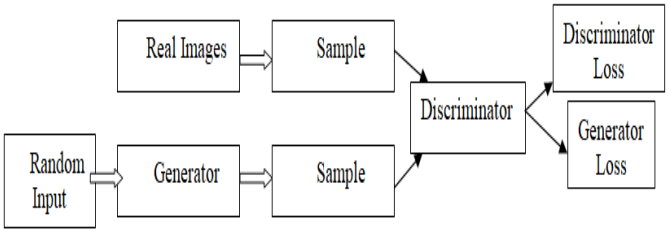

Figure 2 GAN Architecture

Both discriminator and generator are neural networks. The output of the generator goes directly into the input of the discriminator. The weights of the generator are updated by backpropagating a signal from the discriminator's classification. The discriminator is trained on two sources of data:

- The actual data samples, such as real credit card information. Such events are helpful training samples for the discriminator.
- Some instances of fake data that the generator produced serve as negative training examples for the discriminator.

The overall two "Sample" boxes in Figure 2 illustrate these two sources of data that are fed into the discriminator. The generator consists of no training data when training the discriminator. Its weights are kept static by generating samples for the discriminator to learn from. The discriminator has two loss functions connected by it. As discriminator training takes place, the discriminator does not consider the generator loss and makes use of only the discriminator loss. The discriminator classifies both valid and fake data produced by the generator.

The discriminator is punished by the discriminator loss when a false instance is misclassified as real or an actual instance as false. The discriminator updates its weights by backpropagating the discriminator loss through the discriminator network (Figueira et al., 2022).

D. Classification Model

This study classifies transactions as either real or fraudulent using supervised machine learning methods such as logistic regression, Random Forests, XGBoost, and Naïve Bayes.

Logistic Regression(LR): Logistic Regression(LR): LR method is used to forecast a binary outcome variable. Explanatory variables do not have to be related or normally distributed to use this method. LR methods adopt a qualitative outcome variable. Explanatory variables can be measured in numbers or categories. Widespread referred to as a logit model, this form of statistical model is extensively used in predictive analytics and classification.This logistic function, often known as the natural logarithm of odds, is shown by the formulas below.

$$Logit(Pi) = \frac{1}{1+\exp(-pi)} \qquad (2)$$

$$Ln\left(\frac{pi}{1-pi}\right) + \beta_0 + \beta_1 * X_1 + \ldots\ldots B_k + K_k \qquad (3)$$

he dependent variable ranges from 0 to 1 since the outcome is a probability. The odds, or success probability divided by failure probability, are logit transformed in LR. Dependent variable is Logit(Pi) and independent variable is X in Equations (1) and (2).

The beta parameter or coefficient of this model is typically estimated by maximum likelihood estimation (MLE). This procedure runs over a set of beta values usually to make the process efficient for the optimum fit of log odds. These iterates develop a log likelihood function, and LR tries to maximize it in order to have the best optimum parameter estimate. Therefore, by calculating the conditional probability for each observation, prediction probability can be achieved. One is forecasted by a value of probability greater than 0 under binary classification, whereas zero is forecasted by a probability less than.5 (Alarfaj et al., 2022)

Random Forests(RF): RFs are an effective, simple ML technique that tends to give good outcomes even without tuning the hyperparameters. Its simplicity and versatility mean it can be used in both classification and regression tasks. It is a supervises learning algorithm. A collection of decision trees, typically trained with the use of bagging, forms the "forest" it generates. The core concept behind the bagging technique is that the averaging of a large number of learning models increases the result.

As a training dataset with characteristics and names is fed into a decision tree, it will produce a set of instructions that will be used to produce predictions. An RF classifier is trained with n_estimators=100 in order to identify fraud tendencies in historical data. Following training, the model determines whether fresh transactions are fraudulent, and its effectiveness is evaluated using the F1-score, accuracy, precision, and recall. It is a dependable fraud detection system since high recall guarantees fewer false negatives, or missed fraud situations. RF reduces bias and volatility and generalizes better than a single decision tree. Some criteria

that the model could select up are "If transaction amount > $5000 and a new device are used, flag as high fraud risk." Overall, by better managing intricate patterns and unbalanced data, our ensemble approach improves fraud detection(Saeed et al., 2024).

XGBoost (XGB): One of the widely used ML algorithms for detecting credit card fraud, XGBoost (Extreme Gradient Boosting) is powerful and strong as it can handle unbalanced data, tune performance, and avoid overfitting. This decision tree-based ensemble ML algorithm uses a gradient boosting framework. Therefore, artificial neural networks tend to perform better than all other frameworks or algorithms if unstructured data (text, etc.) is used for prediction tasks. XGBoost is very effective in identifying subtle fraud patterns since it builds decision trees sequentially, fixing the errors of the last one. The approach uses measures such as accuracy, recall, F1-score, and AUC-ROC to make high fraud detection with low false negatives and false positives. The model makes an estimation of whether transactions are fraudulent or not. XGBoost is more scalable, efficient, and effective in finding complex fraud patterns than traditional ML models(Ango et al., 2024).

Naïve Bayes(NB): Naive Bayes(NB) classifiers generate fast and precise predictions by applying Bayes' Theorem. Their ease of use and effectiveness in text-based categorization tasks have made them popular. NB classifiers are ideal for real-time applications since they are supervised ML techniques. For data scientists and analysts, they are important. They are widely utilized because of their capacity to manage high-dimensional data and function effectively with tiny datasets. NB classifier uses previous information to determine the likelihood of an event. It assumes feature independence and uses conditional probability to predict class labels. This streamlines computations but may ignore difficulties in the real world.In NB, the fundamental components of the Bayes theorem are:

- Prior probability: A class's initial likelihood before any supporting data is taken into account
- Probability under condition: The likelihood of seeing particular characteristics in a class
- Probability in the posterior: The revised class probability after taking the evidence into account

These probabilities are used to make predictions by NB classifiers. Each class's posterior probability is determined, and the highest one is selected as a prediction(Khan et al.,2024).

IV. Results and Discussions

. The experiments were carried out in python with a dataset available on Kaggle for CCF detection. The GAN model is evaluated using four ML classifiers like LR, RF, XGB and NB.

A. Dataset Description

The "Credit Card Transactions Fraud Detection Dataset" is a synthetic dataset of credit card transactions from January 1, 2019, to December 31, 2020, that comprises both legitimate and fraudulent transactions. One thousand clients' credit cards that are making purchases from eight hundred retailers are covered.

B. Performance Metrics

Accuracy: A popular performance indicator called accuracy measures how frequently a model classifies cases properly.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Precision: determines the proportion of transactions that are mistakenly identified as fraudulent. Fewer false positives legitimate transactions that are mistakenly flagged—are the result of high precision. It is crucial to minimize consumer frustration by not obstructing legitimate transactions.

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

Recall: calculates the number of real frauds that were accurately detected. Fewer false negatives (missed fraud instances) are associated with high recall. It is crucial for identifying as many fraudulent transactions

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

F1-Score : balances recall and precision, which makes it helpful for datasets that are unbalanced. A good trade-off between preventing false alarms and identifying fraud is indicated by a high F1-score.

$$F1 - Score = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (7)$$

Table 1(a) and (b) examines the efficiency of four ML classifiers for detecting CCF with and without personalized content generation using GAN: LR, RF, XGB and NB in terms of accuracy, precision, recall, and F1-score

Table 1(a) Performance Analysis: GAN with and Without ML Models

Methods	Accuracy (%)		Precision(%)	
	With GAN	Without GAN	With GAN	Without GAN
Logistic Regression(LR)	96.3	93.1	85.6	81.2
Random Forests(RF)	97.7	95.4	91.2	87.5
XGBoost(XGB)	98.6	95.8	94.7	90.1
Naïve Bayes(NB)	92.1	88.0	81.3	75.4

Table 1(b) Performance Analysis: GAN with and Without ML Models

Methods	Recall (%)		F1-Score (%)	
	With GAN	Without GAN	With GAN	Without GAN
Logistic Regression(LR)	76.5	70.9	82.4	74.7
Random Forests(RF)	83.9	76.3	87.8	81.4
XGBoost(XGB)	86.4	78.4	90.3	84.3
Naïve Bayes(NB)	75.2	66.5	77.2	70.5

From table 1(a) and (b), it is found that the GAN-based data augmentation enhances performance in all metrics (Accuracy, Precision, Recall, F1-Score) for all the ML models used in this study. XGBoost performs more efficiently than any other model, with 98.6% accuracy, 94.7% precision, and 90.3% F1-score. The greatest improvement is in recall (86.4% compared to 78.4% without

GAN), suggesting that GAN is a useful tool for identifying fraud trends.

With GAN, RFs' precision reaches 91.2% and their accuracy increases from 95.4% to 97.7% . It has a significant recall of 83.9% vs. 76.3% indicates that RF gains from the synthetic data. The accuracy of the LR technique rises from 93.1% to 96.3%. From 74.7% to 82.4%, the F1-score increases, suggesting a more equitable trade-off between recall and precision. Despite having the worst performance, Naïve Bayes (NB) still gains from GAN . It is the least accurate model (92.1%) and the least precise (81.3%). But recall increases from 66.5% to 75.2%, indicating that GAN aids NB in identifying more fraud incidents.

Figure 3 displays the performance of the classifiers for LR, RF, XGB, and NB for CCF detection with personalized content generation using GAN and without GAN. The four evaluation metrics that are shown are accuracy, precision, recall, and F1-score.

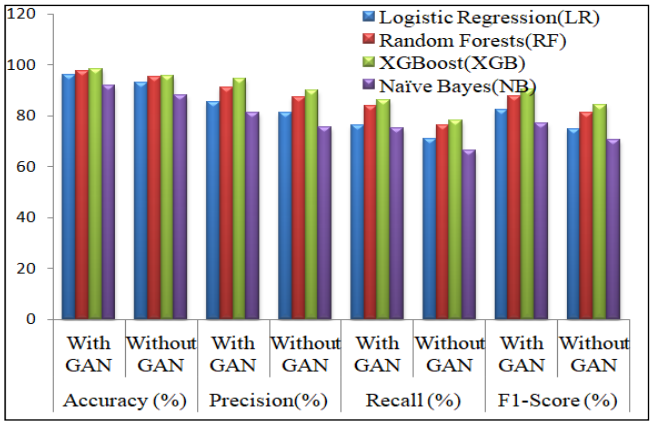

Figure 3: Performance Analysis: GAN with and Without ML Models

Figure 3 shows that the "With GAN" is consistently greater than the "Without GAN" for all models and metrics. This demonstrates how GAN-generated synthetic data improves fraud detection. With and without GAN, XGB has the best F1-score, Accuracy, Precision, and Recall. Recall and F1-score show the greatest gain with GAN, suggesting improved fraud detection capabilities. XGB is the most effective model, followed by RF NB and LR. GAN-generated data significantly improves recall, especially for XGB and RF. Because it is more costly to ignore illegal transactions than to flag a few false positives, this is crucial for fraud detection. The NB model is the least effective since its assumptions do not take into consideration the complex relationships involved in fraud detection.

V. Conclusion

The study is conducted to enhance the CCF detection using GAN model. ML classifiers are used to evaluate the model efficiency. The findings conclusively show that synthetic data produced by GANs improves fraud detection performance across all assessed models and indicators proving that utilizing synthetic data can enhance model learning. It shows that XGBoost with GAN outperforms the LR, RF and NB models with an accuracy of 98.6 in detecting the CCF. GAN-generated data, in general, improves ML models' capacity to identify fraud more precisely, lowering the number of fraudulent cases that are overlooked and boosting system dependability. In the future, more sophisticated GAN architectures like conditional or adaptive GANs will be developed to produce fraud patterns that are even more realistic. Further, investigating how various data augmentation methods interact with GAN-generated data may offer possibilities for improving model performance.

References

[1] F. K. Alarfaj, I. Malik, H. U. Khan, N. Almusallam, M. Ramzan and M. Ahmed, "Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms," in IEEE Access, vol. 10, pp. 39700-39715, 2022, doi: 10.1109/ACCESS.2022.3166891

[2] Langevin, A., Cody, T., Adams, S., & Beling, P. (2022). Generative adversarial networks for data augmentation and transfer in credit card fraud detection. Journal of the Operational Research Society, 73(1), 153-180.

[3] Kaur, S., Singh, K. D., Singh, P., & Kaur, R. (2021). Ensemble model to predict credit card fraud detection using random forest and generative adversarial networks. In Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2020, Volume 2 (pp. 87-97). Springer Singapore.

[4] Kulkarni, P., Pathak, P., Pillai, S., & Tigga, V. (2025). Role of Generative AI for Fraud Detection and Prevention. Generative Artificial Intelligence in Finance: Large Language Models, Interfaces, and Industry Use Cases to Transform Accounting and Finance Processes, 175-198.

[5] Shafik, W. (2024). The Role of Generative Artificial Intelligence in E-Commerce Fraud Detection and Prevention. In Strategies for E-Commerce Data Security: Cloud, Blockchain, AI, and Machine Learning (pp. 430-469). IGI Global.

[6] Strelcenia, E. (2024). A New Generative Adversarial Network for Improving Classification Performance for Imbalanced Data (Doctoral dissertation, Bournemouth University).

[7] Khalid, A. R., Owoh, N., Uthmani, O., Ashawa, M., Osamor, J., & Adejoh, J. (2024). Enhancing credit card fraud detection: an ensemble machine learning approach. Big Data and Cognitive Computing, 8(1), 6.

[8] F. A. Ghaleb, F. Saeed, M. Al-Sarem, S. N. Qasem and T. Al-Hadhrami, "Ensemble Synthesized Minority Oversampling-Based Generative Adversarial Networks and Random Forest Algorithm for Credit Card Fraud Detection," in *IEEE Access*, vol. 11, pp. 89694-89710, 2023, doi: 10.1109/ACCESS.2023.3306621.

[9] Zhao, C., Sun, X., Wu, M., & Kang, L. (2024). Advancing financial fraud detection: Self-attention generative adversarial networks for precise and effective identification. Finance Research Letters, 60, 104843.

[10] Aftabi, S. Z., Ahmadi, A., & Farzi, S. (2023). Fraud detection in financial statements using data mining and GAN models. Expert Systems with Applications, 227, 120144.

[11] Singh, K. D., Singh, P., & Kang, S. S. (2022, October). Ensembled-based credit card fraud detection in online transactions. In AIP Conference Proceedings (Vol. 2555, No. 1). AIP Publishing.

[12] E. Strelcenia and S. Prakoonwit, "Generating Synthetic Data for Credit Card Fraud Detection Using GANs," *2022 International Conference on Computers and Artificial Intelligence Technologies (CAIT)*, Quzhou, China, 2022, pp. 42-47, doi: 10.1109/CAIT56099.2022.10072179.

[13] Figueira, A., & Vaz, B. (2022). Survey on synthetic data generation, evaluation methods and GANs. Mathematics, 10(15), 2733.

[14] Saeed, V. A., & Abdulazeez, A. M. (2024). Credit Card Fraud Detection using KNN, Random Forest and Logistic Regression Algorithms: A Comparative Analysis. The Indonesian Journal of Computer Science, 13(1).

[15] Ango, R., Masih, R. K., Reddy, C. K. K., Shuaib, M., Singh, M., & Alam, S. (2024, December). Fraud Detection in Banking using the Kaggle Credit Card Dataset and XGBoost Model. In 2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS) (pp. 968-973). IEEE.

[16] Khan, S., & Palaniswamy, B. (2024, June). Enhancing Credit Card Fraud Detection: A Comparative Analysis of Artificial Neural Networks, k-NN, and Naive Bayes Classifiers. In 2024 IEEE Students Conference on Engineering and Systems (SCES) (pp. 1-6). IEEE.