RESEARCH ARTICLE OPEN ACCESS

# The Application of Federated Generative Adversarial Networks in Medical Insurance Anti-Fraud

WU Chenxi[1], FANG Qiquan[2], ZHANG Ying[3]

[1]*Science College, Zhejiang University of Science and Technology, Hangzhou, China*
[2]*Science College, Zhejiang University of Science and Technology, Hangzhou, China*
[3]*Science College, Zhejiang University of Science and Technology, Hangzhou, China*
[1]*charleswu23@qq.com*

Received: Revised: Accepted: Published:

***Abstract -*** *To address medical insurance fraud, this study developed a Federated Generative Adversarial Network (FGAN) model, integrating federated learning with GANs. Using the Shenzhen Cup inpatient reimbursement dataset, we performed standardized preprocessing and evaluated the model against logistic regression, random forest, a standalone GAN, and a baseline federated learning method. Results indicate that the FGAN model achieved balanced performance, with precision, recall, F1-score, and AUC reaching* 0.927, 0.892, 0.909 *and* 0.986 *respectively. Model performance also improved consistently with more training epochs. By enabling collaborative modeling across institutions without sharing raw data, FGAN mitigates sample limitations inherent in isolated federated learning setups. This work demonstrates the effectiveness and practical value of FGAN for detecting and preventing medical insurance fraud.*

***Keywords -*** *Federated Learning, Generative Adversarial Networks, Logistic Regression, Random Forest, Healthcare Insurance, Anti-Fraud.*

## 1. Introduction

As the basic institutional arrangement of our country's social security system, social medical insurance plays an irreplaceable role in protecting the health rights and interests of the whole people. However, with the continuous improvement of the coverage rate of the system, significant institutional defects have been exposed during its operation. According to statistics from the National Health Insurance Administration, from 2020 to 2024, the supervision of the national health insurance fund has investigated and dealt with more than 1.644 million violations of laws and regulations, causing more than 10 million economic losses per year. There are two main causes of this phenomenon: first, techno logistic iteration has spawned a dynamic fraud model based on deep learning, and the subject of fraud has significantly reduced the detection accuracy rate of traditional methods through forged medical documents, fictitious diagnosis and treatment paths, etc.; Second, there are significant data barriers to medical data governance, and medical data governance has significant data barriers. Due to compliance considerations, institutions are cautious about cross-agency data sharing, which makes it difficult for anti-fraud systems to build a complete process covering diagnosis and treatment.

Traditional anti-fraud methods rely on rule engines or supervised learning, but fraud is highly hidden and data silos are common, resulting in insufficient detection accuracy. This paper studies two main models, federated learning and generative adversarial network. Foreign scholars have made great progress in the research of the two. Abdul Salam[3]et al.'s research focuses on credit card transaction fraud detection. In view of the two core challenges of data privacy and category imbalance, a solution based on federated learning and hybrid resampling technology is proposed to ensure high accuracy while reducing computing costs., More suitable for deployment by financial institutions; Research by Wenyou Duet al. proposes a WOE coding method under the framework of vertical federated learning. By constructing an implicit transfer mechanism of cross-institutional tags, unstructured data tags are realized under the premise of protecting the privacy of participants. Numerical conversion of text features. The results show that the AUC value of this method is increased by 12.4% in insurance anti-fraud classification tasks compared

with traditional TF-IDF coding, which verifies the effectiveness of FL integration of multi-source heterogeneous data; Y. Tang and Z. Liu[5]et al. Conducted research on the detection of credit card transaction fraud, combining Transformer's time series modeling capabilities with the privacy protection mechanism of federated learning to break through the performance bottleneck of traditional models in data silos and category imbalance scenarios; Ugo Fiore[6]et al. used GAN to train a multilayer perceptron to output simulated minority samples, and merged these samples with the training data into an enhanced training set, thereby improving the effective of the classification. The failure of existing methods to break through mainly faces two major problems: one is the issue of privacy protection when operating in concert across institutions. Centralized methods have strong synergy but weak privacy. When using a single federated learning, the privacy protection ability is strong but the learning ability is weak; The second is that single federated learning faces the problem of scarce samples, and traditional methods rely on real samples, and the generation of a single GAN sample lacks cross-institutional adaptability. These contradictions have led to the inability of the existing technology to form a high-precision and compliance medical insurance anti-fraud system.

## 2. The Construction of federated generation adversarial network

### *2.1. Federated Learning*

*2.1.1. Definition of Federated Learning*

Federated learning (FL) is a kind of distributed machine learning technology. Its core feature is that the original data is always retained locally by each participant, and the training is only completed through the encrypted transmission and aggregation of the intermediate results of the model to achieve the core goal of data availability and invisibility. The model performance and data privacy protection requirements of machine learning. The simplified specific process is: first of all, multiple participants each hold local data, and the data is always stored locally and not transmitted outwards. Then the central server sends the initial model parameters to all participants. Then each participant trains the model with local data, and only uploads the trained model parameters to the central server. Finally, the central server aggregates and optimizes all the parameters, generates a global model, and then feeds it back to the participants. Repeat the cycle of "local training-parameter upload-global aggregation-model feedback" until the model achieves optimal results. The simplified algorithm flow chart is shown in Figure 1.

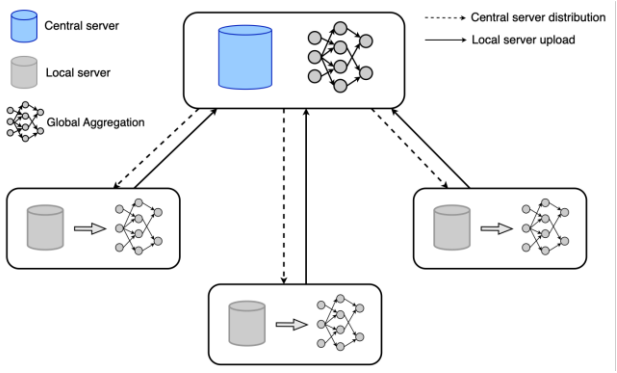


**Fig.1 Federated Learning Simplified Flow Chart**

*2.1.2. The algorithm process of federated learning*

Show the algorithm flow of federated learning as follows:

**Algorithm 1**: Federated learning

input: K client data sets $\{D_1, D_2, \cdots, D_k\}$, global model parameter $w_0$, learning rate $N$, training round $T$

Output: optimized global model parameter $w^*$

1: for $t = 1\ to\ T$

2: The server randomly selects a collection of $m$ clients $S_t$

3: for each client $k \in S_t$ to execute in parallel

4: Local update: $w_k^{T+1} \leftarrow w^T - \eta \nabla \mathrm{F_k}(w^t)$

5: end for

6: Server aggregation: $w^{t+1} \leftarrow \Sigma_k \left(\frac{|D_k|}{|D|}\right) \cdot w_k^{t+1}$

7: end for

### *2.2. Generative Adversarial Network*

*2.2.1. Definition of generator and discriminator*

The essence of the generative adversarial network (GAN) is "adversarial and collaborative training of two models", which contains two core components: Generator and Discriminator. The goal of the generator: To learn the distribution law of real data and generate "fake data" (such as simulated fraudulent samples, images, and text) that are as close to the real data as possible. The goal of the discriminator: Learn to distinguish whether the input data is "real data" or "fake data" (generated by the generator). The two are in a continuous game: the generator is constantly optimized to "fool" the discriminator, and the discriminator is constantly optimized to "see through" the false data until a balance is reached, that is, the discriminator cannot distinguish between the true and the false data. At this time, the generator can output high-quality simulation data, and the discriminator has accurate

Recognition ability. The simplified principle of generating an adversarial network is shown in Figure 2.

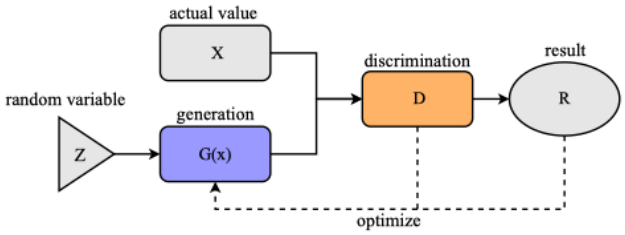


**Fig. 2 Generative Adversarial Network Simplified Diagram**

*2.2.2 Adversarial loss function*

The optimization goal of GAN is the minimum and maximum game problem:

$$minmaxV(D,G) = \sum_{x}^{p_{data}} [logD(x)] + E_{z\sim p_z}\left[\log\left(1 - D\left(G(z)\right)\right)\right] \quad (1)$$

Among them, the discriminant loss function:

$$L_D = -\sum_{x}^{p_{data}} [logD(x)] - \sum_{z}^{p_z}\left[\log\left(1 - D\left(G(z)\right)\right)\right] \quad (2)$$

Generator loss function:

$$L_G = -\sum_{z}^{p_z} [logD(\,G(z)\,)] \quad (3)$$

*2.2.3. Generative adversarial network training process*

Show the algorithm flow of the generative adversarial network as follows：

**Algorithm 2:** Generative adversarial network

input: real data distribution $p_{data}$, noise distribution $p_z$

output: trained generator G and discriminator D

1: initialize the parameters of G and D

2: for epoch = 1 to E

3: for k steps

4: from $p_z$ sampling noise $\{z_1, z_2, \ldots, z_m\}$

5: generate samples:$\tilde{x}_i = G(Z_i)$

6: $p_{data}$ sample real samples from p data$\{x_1, x_2, \ldots, x_m\}$

7: update the discriminator:$\theta g \leftarrow \theta d - \eta \nabla \mathrm{LG_G}$

8: end for

9: Sampling noise from $\{z_1, z_2, \cdots, z_m\}$

10: Update the generator: $\theta g \leftarrow \theta g - \eta \nabla \mathrm{L_G}$

11: end for

***2.3 Federated Generation Adversarial Network***

*2.3.1 Definition of FGAN*

The federated generated adversarial network (FGAN) is not simply superimposing two technologies, but enhancing each other in concert. Tomisin Awosika[1]et al. proposed the deep integration of FL with GAN and XAI technology, GAN based on the federated architecture can break through the unique data silos restrictions in the medical field, and allow the regulatory model to integrate the fraud characteristics of multiple institutions. At the same time, the embedding of the XAI module can also meet the traceability needs of ethical review and regulatory compliance decision-making in medical scenarios; on the other hand, GAN is used to generate virtual fraud samples locally in each participant, and synthesize them by generating synthetic samples. A small number of types of samples alleviate the problem of uneven categories in health insurance fraud detection, not only supplementing the scarce samples of a single organization, but also avoiding the privacy risks of cross-agency transmission of virtual samples. Finally, the dual goal of “guaranteed privacy protection and sufficient data for model training” is achieved. The simplified principle of federated generated adversarial network is shown in Figure 3.

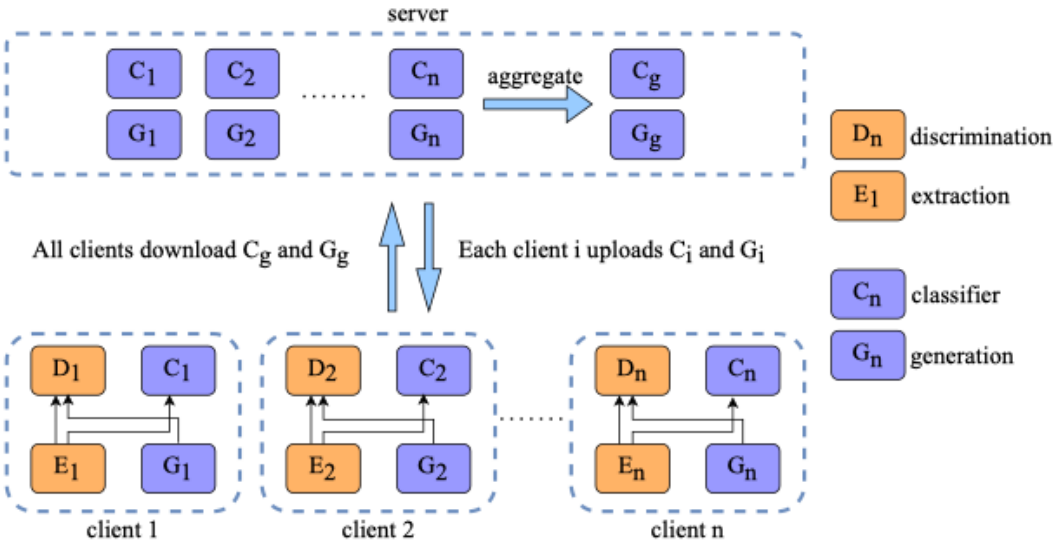


**Fig. 3 Federated Generative Adversarial Network Simplified Schematic Diagram**

*2.3.2. FGAN's Algorithm flow*

Show the algorithm flow of the federated generative adversarial network as follows:

**Algorithm 3:** Federated generative adversarial network

input: K client data $\{D_1, D, \ldots, D_k\}$, local training round L, global communication round T

output: global generator G and discriminator D

1: The server initializes the global model $G^{(0)}, D^{(0)}$

2: **for** t = 1 to T

3: The server selects the client set $S_t$

4:**for** each client $K \in S_t$ execution

5: Download the global model $G^{(t)}$, $D^{(t)}$

6: **for** l = 1 to L

7: locally train the GAN model

8: **end for**

9: upload the model to update $\nabla G_k$, $\Delta D_k$

10: **end for**

11: server aggregation: $G^{(t+1)} \leftarrow \mathrm{FedAvg}(\{G_k\})$, $D^{(t+1)} \leftarrow \mathrm{FedAvg}(\{D_k\})$

12: **end for**

## 3. Data preprocessing and feature index selection

### *3.1 Data Sources and Characteristics*

The data is derived from the 2025 Shenzhen Cup Mathematical Modeling Competition A submission data. The research in this paper is mainly aimed at the medical reimbursement of hospitalized patients. The required data mainly include age, gender, date of admission, date of discharge, type of disease, total amount of hospitalization and reimbursement. By selecting the data required for this study and processing the targets privately, the following results have been obtained. For the data, see Table 1:

**Table 1. Inpatient Status Data Table (Partial)**

| ID | Age | Gender | Date Admitted | Date Discharged | Diagnosis | Amount Billed | Fraud Type |
|---|---|---|---|---|---|---|---|
| 1 | 4 | Female | 2023/5/17 | 2023/5/20 | Cataract Surgery | 87417.22 | No Fraud |
| 2 | 87 | Male | 2024/5/14 | 2024/5/14 | Hypertension | 960571.45 | Phantom Billing |
| 3 | 30 | Male | 2023/2/12 | 2023/2/22 | Migraine | 151758.91 | No Fraud |
| 4 | 58 | Female | 2022/1/14 | 2022/1/16 | Gastroenteritis | 127758.53 | No Fraud |
| 5 | 20 | Male | 2023/11/21 | 2023/11/22 | Pregnancy | 48083.36 | No Fraud |
| 6 | 46 | Female | 2023/6/26 | 2023/6/26 | Routine Check-up | 48079.01 | No Fraud |
| 7 | 49 | Male | 2023/8/24 | 2023/9/3 | Diabetes | 30455.05 | No Fraud |
| 8 | 47 | Male | 2022/5/23 | 2022/5/23 | Pregnancy | 175911.71 | Wrong Diagnoses |
| 9 | 99 | Female | 2022/6/13 | 2022/6/16 | Pregnancy | 128200.7 | No Fraud |

For follow-up research, simplify the data, replace the check-in date and discharge date with the hospitalization time1, and replace the medical insurance reimbursement situation with $0-1$, as shown in Table 2:

**Table 2. Simplified Inpatient Status Data Table (Partial)**

| ID | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $Y$ |
|---|---|---|---|---|---|---|
| 1 | 4 | Female | 4 | Cataract Surgery | 87417.22 | 0 |
| 2 | 87 | Male | 1 | Hypertension | 960571.45 | 1 |
| 3 | 30 | Male | 11 | Migraine | 151758.91 | 0 |
| 4 | 58 | Female | 3 | Gastroenteritis | 127758.53 | 0 |
| 5 | 20 | Male | 2 | Pregnancy | 48083.36 | 0 |
| 6 | 46 | Female | 1 | Routine Check-up | 48079.01 | 0 |
| 7 | 49 | Male | 11 | Diabetes | 30455.05 | 0 |
| 8 | 47 | Male | 1 | Pregnancy | 175911.71 | 1 |
| 9 | 99 | Female | 4 | Pregnancy | 128200.7 | 0 |
| 10 | 59 | Female | 6 | Cesarean Section | 147453.06 | 0 |

The meanings of each variable are: $x_1$: the patient's age; $x_2$: the patient's gender; $x_3$: the patient's hospitalization time; $x_4$: the type of disease the patient suffers from; $x_5$: The total amount of the patient's hospitalization; $Y$: The patient's medical insurance reimbursement, the value is 0. It means that there is no fraud, and a value of 1 means that there is fraud.

### *3.2. Data Preprocessing*

*3.2.1. Descriptive statistics*

Use Python to perform descriptive statistics on numeric variables in the data set, the results are shown in Table 3:

**Table 3. Descriptive Statistics Table**

| | $x_1$ | $x_3$ | $x_5$ |
|---|---|---|---|
| Count | 500.00 | 500.00 | 500.00 |
| Mean | 48.88 | 136275.44 | 5.93 |
| Std | 28.71 | 124295.67 | 3.20 |
| Min | 3.00 | 20993.98 | 1.00 |
| 25% | 22.75 | 57607.98 | 3.75 |
| 50% | 49.50 | 118046.57 | 5.50 |
| 75% | 71.25 | 169339.48 | 8.25 |
| Max | 100.00 | 960571.45 | 11.00 |

*3.2.2. Handling of outlier values*

Outlier values are detected based on the $3\sigma$ principle and boxplot analysis. After identification, the mean value of the data is used instead of a field of data, and the data set is checked again after processing to ensure that outlier values are properly processed.

*3.2.3. Data standardization*

According to the descriptive statistical table in Table 3 above, it can be seen that due to the large order-of-magnitude difference between different indicators, the distribution of the data is biased. In order to avoid its impact on subsequent research, Z-score standardization is used for numerical characteristics. The standardized formula is as follows:

$$X_i^* = \frac{X_i - E(X_i)}{\sqrt{D(X)}}, i = 1, 2, \cdots, p \quad (4)$$

Where $E(X_i)$ is the mean of the data$and$ $D(X_i)$ is the variance of the sample.

## 4. Example analysis

### *4.1. Experimental Setup*

In order to verify the predictive performance of the model, the user hospitalization data is used as the starting point for the experiment. After the data is preprocessed, the user data is imported into three models, namely, logistic regression (LR), federated generative adversarial network (FGAN), federated learning (FL), and random forest (RF), and the user data is imported into three models, such as logistic regression (LR), federated generative adversarial network (FGAN), federated learning , and random forest (RF).Predictive analysis of user data for fraud cases. Finally, the fraud cases predicted by each model are compared with the real fraud cases, and the accuracy rate of the model prediction is calculated, so that the advantages and disadvantages of each model can be tested by comparison. The simplified experimental flow chart is shown in Figure 4:

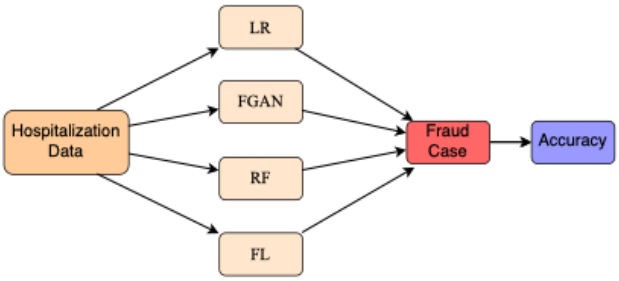


**Fig. 4 Experiment Flow Chart**

### *4.2. Model Comparison*

Four models were selected for experiments and the prediction results were compared. In addition to the main algorithms studied in this article: federated learning and single GAN, logistic regression and random forest algorithms were also selected. The following will briefly introduce logistic regression and random forest algorithms.

*4.2.1. Logistic regression*

The logistic regression algorithm is a classic two-way classification model in the field of machine learning. Although it is called "regression", its essence is a probabilistic classification algorithm. Logistic regression is based on linear regression and realizes classification tasks by introducing the nonlinear mapping function Sigmoid. It has the characteristics of simple model, strong interpretability, and high training efficiency. Its core idea is: to construct a linear prediction model based on the eigenvectors, and then Sigmoid map the linear output to the $[0,1]$ interval through the Sigmoid function. The output value represents the probability that the sample belongs to the positive category, and finally the classification decision is made according to the probability threshold. The mathematical principle of logistic regression is mainly divided into the following 4 parts:

1. Introduce the linear prediction part, the mathematical expression is as follows:

$$z_i = \boldsymbol{w}^T \boldsymbol{x}_i + b \tag{5}$$

2. Introduce the Sigmoid function, the mathematical formula is as follows:

$$\sigma(z) = \frac{1}{1+e^{-z}} \tag{6}$$

3. With the mapping of the Sigmoid function, the probability that the sample belongs to the positive class ($y_i = 1$):

$$\hat{y_i} = P(y_i = 1|\boldsymbol{x_i}; \boldsymbol{w}, b) = \sigma(z_i) = \frac{1}{1+e^{-(w^T x_i + b)}} \tag{7}$$

4. Probability that the sample belongs to the negative category ($y_i = 0$):

$$P(y_i = 0|\boldsymbol{x_i}; \boldsymbol{w}, b) = 1 - \hat{y_i} = \frac{e^{-(w^T x_i + b)}}{1+e^{-(w^T x_i + b)}} \tag{8}$$

*4.2.2. Random forest*

The random forest algorithm is a classic algorithm in the field of integrated learning. It is based on a decision tree as the basic learner. It is an integrated model composed of multiple independently trained decision trees. Through the dual randomness of random sample sampling and random feature selection, it effectively reduces the risk of overfitting a single decision tree and significantly improves the model. Generalization ability and prediction stability. The core idea of random forest is: put back sampling of the original data set to generate multiple different sub-data sets; when each decision tree is trained, some features are randomly selected to build the optimal split node; the classification task integrates the prediction results of all decision trees by the majority voting method, and the regression task integrates by the mean method. result. The mathematical principle of random forest is mainly divided into the following 2 steps:

1. Classification task: The random forest consists of M decision trees, each tree $h_m(x)$ makes a prediction of the input x, and the final prediction is the result of the majority vote:

$$\hat{y_i} = mode\{h_1(x), h_2(x), \ldots, h_M(x)\} \tag{9}$$

2. Regression task: Average the predicted values of all trees, which is the final prediction:

$$\hat{y} = \frac{1}{M}\sum_{m=1}^{M} h_m(x) \tag{10}$$

### *4.3. Forecast Results*

*4.3.1. Accuracy*

Accuracy refers to the proportion of correct predictions made by a model out of all samples, providing a direct measure of a model's predictive performance. The calculation formula is as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{11}$$

*4.3.2. F1 score*

In this paper, the prediction performance of each model is evaluated by calculating the F1 score. The F1 score is an important indicator to evaluate the performance of the classification model, especially when the data category is unbalanced. It can reflect the performance of the model more comprehensively than the accuracy rate. It combines the accuracy rate and recall rate of the model, and is the reconciled average of the two. Precision refers to the accuracy of the

model's prediction results. It indicates how many of all samples predicted by the model to be positive are really positive. The calculation formula is as follows:

$$Precision = \frac{TP}{TP + FP} \quad (12)$$

The Recall rate refers to the coverage rate of the real situation. It indicates how many of all true positive samples have been successfully found by the model. The calculation formula is as follows:

$$Recall = \frac{TP}{TP + FN} \quad (13)$$

The formula for calculating the F1 score is as follows:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (14)$$

Where TP is a true positive, which represents the number of samples that the model correctly predicts to be positive; TN is a true negative, which represents the number of samples that the model correctly predicts to be negative; FP is a false positive, which represents the number of samples that the model incorrectly predicts to be positive, which can be understood as the model Misjudgment, a negative sample is judged to be a positive class; FN is a false negative, indicating that the model incorrectly predicts the number of samples of the negative class, that is, the model missed the sample that should have been a positive class but was not found. Because the tuning average is characterized by being very sensitive to extremes, the F1 score will only be high if both the accuracy rate and the recall rate are high. This makes the F1 score a very balanced indicator, suitable for use in scenarios where both accuracy and recall rates need to be taken into account.

*4.3.3. AUC*

AUC refers to the area below the ROC curve, and the ROC curve is a curve drawn by the true positivity rate (recall rate) and false positivity rate under different thresholds. AUC can measure the overall ability of a model to distinguish between positive and negative categories, especially suitable for multi-model comparison and thresholds. Uncertain scene. The formula for calculating the false positive rate is as follows:

$$FPR = \frac{FP}{FP + TN} \quad (15)$$

Finally, the comparison of the prediction performance of each model on the test set is calculated as shown in Table 4 below:

**Table 4. Model Prediction Performance Comparison Table**

| Model | Accuracy | Precision | Recall | F1 | AUC |
|---|---|---|---|---|---|
| LR | 0.9400 | 0.8540 | 0.9194 | 0.8855 | 0.9885 |
| RF | 0.9553 | 0.9601 | 0.8587 | 0.9066 | 0.9876 |
| GAN | 0.9643 | 0.9656 | 0.8904 | 0.9265 | 0.9923 |
| FL | 0.9417 | 0.8601 | 0.9181 | 0.8882 | 0.9885 |
| FGAN | 0.9550 | 0.9272 | 0.8917 | 0.9091 | 0.9863 |

The table above visually shows the values of each algorithm on the two key indicators of F1 score and AUC. From the tabular data, it can be seen that in terms of accuracy, a single GAN algorithm is the highest and the logistic regression algorithm is the lowest; in terms of recall, the logistic regression algorithm is the highest and the random forest algorithm is the lowest; in terms of F1 score, a single GAN, the algorithm is the highest and the logistic regression algorithm is the lowest; in AUC, the single GAN algorithm is the highest and the FGAN algorithm is the lowest. In summary, the overall performance of a single GAN is in the leading position, reflecting its comprehensive advantages in classification tasks.

*4.4.4. Analysis of results*

In order to more intuitively show the performance differences of each model, a radar map of the algorithm is drawn, as shown in Figure 5:

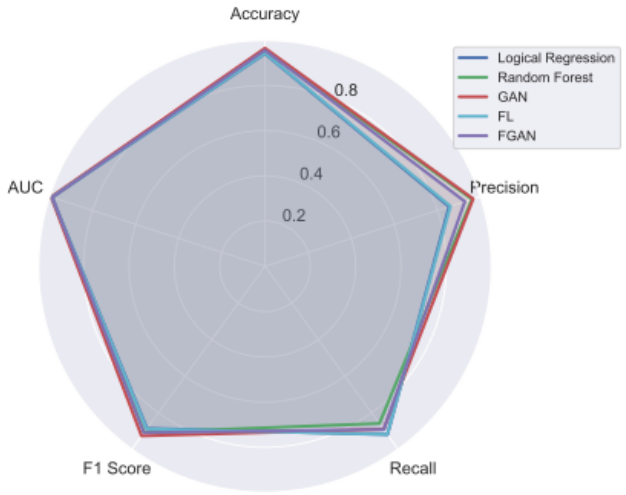


**Fig. 5 Algorithm Radar Chart**

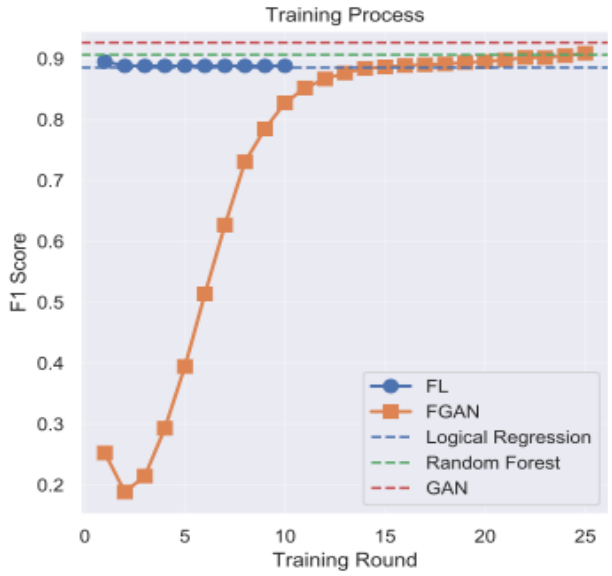


**Fig. 6 F1 Score Variation Line Chart**

The radar map of the algorithm analyzes the dimensions (accuracy rate, accuracy rate, recall rate, F1, AUC), and compares the equilibrium of the algorithm. The coverage area and shape of the radar map can reflect the comprehensiveness of the algorithm. From the figure, it can be seen that a single GAN may have the most comprehensive performance in multiple dimensions. In addition, as the training rounds are increased, the prediction performance of each algorithm will also change. From this, the F1 score is drawn. See Figure 6. The F1 score line chart can show the convergence speed and stability of each algorithm's F1 score during the training process. By analyzing the line chart, it can be concluded that under low training rounds of FGAN, the F1 score is always difficult to reach a stable state, and it is significantly lower than 0.8 . As the training rounds exceed 10 times, the F1 score of FGAN gradually tends to stabilize, showing a trend of surpassing other algorithms. FGAN achieved a significant increase in F1 scores in a very small number of training rounds, which reflects the high learning efficiency of the FGAN algorithm. For computing resource or time-sensitive application scenarios, this fast convergence feature is very valuable, which means that core performance can be obtained with less training cost.

## 5. Conclusion

By combining the distributed sample generation technology with cross-agency collaboration, the Federated Generation Adversarial Network (FGAN) has successfully resolved the three problems of data isolation, sample acquisition and privacy in the health insurance anti-fraud. The anti-fraud system based on the federated generated network has greatly enhanced the fraud detection rate, achieving an accuracy rate of up to 95.5%, and has successfully identified a number of cross-agency collaborative fraud protection cases. With the actual use of the FGAN algorithm, its prediction performance will improve, and FGAN can achieve the expected prediction accuracy with very little training, greatly reducing the actual cost of use. Putting FGAN into actual use is expected to save millions of yuan in health insurance funds, which verifies the feasibility of its technology and its huge social value.

## Funding Statement

This research was funded by the National-level College Students' Innovation and Entrepreneurship Training Program. The project number is 0101108516.

## References:

[1] T. Awosika, R. M. Shukla and B. Pranggono, "Transparency and Privacy: The Role of Explainable AI and Federated Learning in Financial Fraud Detection," in *IEEE Access*, vol. 12, pp. 64551-64560, 2024. https://ieeexplore.ieee.org/abstract/document/10509682

[2] Brisimi T S, Chen R, Mela T, et al. Federated learning of predictive models from federated electronic health records[J]. International journal of medical informatics, 2018, 112: 59-67. https://doi.org/10.1016/j.ijmedinf.2018.01.007

[3] Abdul Salam, M., Fouad, K.M., Elbably, D.L.et al. Federated learning model for credit card fraud detection with data balancing techniques.

Neural Comput & Applic 36, 6231–6256 (2024). https://link.springer.com/article/10.1007/s00521-023-09410-2

[4] W. Du, H. Wang, J. Shen, G. Meng, H. Li and W. Zhou, "Insurance Anti-fraud based on FL-WOE Encoding for Vertical Federated Learning," 2024 IEEE International Conference on Big Data (BigData), Washington, DC, USA, 2024, pp. 7717-7724. https://link.springer.com/article/10.1007/s11433-019-1528-y

[5] Y. Tang and Z. Liu, "A Credit Card Fraud Detection Algorithm Based on SDT and Federated Learning," in IEEE Access, vol. 12, pp. 182547-182560, 2024. https://ieeexplore.ieee.org/document/10742353

[6] Ugo Fiore, Alfredo De Santis, Francesca Perla, Paolo Zanetti, Francesco Palmieri，Using generative adversarial networks for improving classification effectiveness in credit card fraud detection, Information Sciences, Volume 479,2019,Pages 448-455,ISSN 0020-0255. https://api.semanticscholar.org/CorpusID:59528925

[7] XIAO Yi , XIAO Jin , LIU John , WANG Shouyang. A MULTISCALE MODELING APPROACH INCORPORATING ARIMA AND ANNS FOR FINANCIAL MARKET VOLATILITY FORECASTING. *Journal of Systems Science and Complexity*, 2014, 27(1): 225-236. https://doi.org/10.1007/s11424-014-3305-4

[8] Li T, Sahu A K, Talwalkar A, et al. Federated learning: Challenges, methods, and future directions[J]. IEEE Signal Processing Magazine, 2020, 37(3): 50-60. https://ieeexplore.ieee.org/document/9084352

[9] Sheller M J, Edwards B, Reina G A, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data[J]. Scientific reports, 2020, 10(1): 1-12. https://doi.org/10.1038/s41598-020-69250-1

[10] Chawla N V, Bowyer K W, Hall L O, et al. SMOTE: synthetic minority over-sampling technique[J]. Journal of artificial intelligence research, 2002, 16: 321-357. https://api.semanticscholar.org/CorpusID:1554582

[11] 支宇晗,胡强,谢肖飞,等. 联邦学习中基于数据选择的模型修复方法研究[J/OL].中国科学:信息科学, 1-26[2026-05-26]. https://link.cnki.net/urlid/11.5846.TP.20260522.1104.004

[12] 杜倩, 刘鸿宇, 胡琦. 社会医疗保险基金欺诈行为的扎根理论研究——基于 58 个医保欺诈刑事案件分析[J]. 法制与经济, 2020(8): 70-71+75.

[13] 菅银龙,陈学斌,景忠瑞,等.基于条件生成对抗网络的联邦学习中数据增强方案[J/OL].计算机应用,1-16[2025-05-16].

[14] 陈志强.基于 GAN 模型的电网用电数据安全生成研究[D].华北电力大学(北京),2023.

[15] 余锋,林庆新,林晖,等.基于生成对抗网络的隐私增强联邦学习方案[J].网络与信息安全学报,2023,9(03):113-122.

[16] 瞿祥谋.基于生成对抗网络的联邦学习非独立同分布数据问题研究[D].重庆大学,2022.

[17] 葛广为. 联邦学习在医疗数据隐私保护中的应用研究[J].信息记录材料,2026,27 (10):181-183.DOI:10.16009/j.issn.1009-5624.2026.10.058.

[18] 蒋大锐,吕峻闽,徐胜超. 基于生成式对抗网络的网络安全态势感知[J/OL].计算机测量与控制, 1-10[2026-05-26]. https://link.cnki.net/urlid/11.4762.tp.20260122.1740.002

[19] 陈石轩,王博琛,赵志林. 基于 GAN 的恶意 PE 文件对抗样本生成模型研究[J].电脑与信息技术,2025,33 (6):69-72+127.DOI:10.19414/j.cnki.1005-1228.2025.06.004.

[20] 考兴楷,苗莉,郑远硕. 基于联邦学习的安全边缘缓存优化机制[J/OL].计算机系统应用, 1-16[2026-05-26]. https://doi.org/10.15888/j.cnki.csa.010197