Effectiveness of Class Balancing Strategies in E-Wallets and Credit Card Transactions

Rishi Diwaker
Rohit Kumar
Samar Sodhi
Girish Karthikeyan

Department of Computer Science and Engineering
National Institute of Technology Karnataka, Surathkal, Mangalore, India

November 2024

Table of Contents

Introduction

- Fraud detection in e-wallets and credit card transactions faces challenges due to class imbalance.
- Fraudulent transactions are rare compared to legitimate ones, causing bias in machine learning models.
- Proposed solution: Enhanced Wasserstein GAN with GCN layers for robust class balancing.

Problem Statement

Detecting fraudulent transactions in financial datasets is hindered by significant class imbalance as it causes bias in the detection model, where the minority class (fraudulent transactions) is vastly underrepresented.

Proposed Methodology: Overview

- **Goal**: Address class imbalance in fraud detection datasets using Enhanced Wasserstein GAN (WGAN) with Graph Convolutional Network (GCN) layers.
- **Key steps**:
  1. Data Preprocessing.
  2. Graph Representation of Transaction Data.
  3. Enhanced WGAN Design:
     - Generator with GCN layers for synthetic data generation.
     - Discriminator for validating synthetic samples.
  4. Integration of synthetic data with the original dataset.

Proposed Methodology: Data Preprocessing

- **Objective**: Prepare the dataset for analysis and modeling.
- **Steps**:
  1. Handle missing values and remove duplicates.
  2. Normalize numerical features to ensure consistency.
  3. Encode categorical variables:
     - One-hot encoding for independent categories.
     - Label encoding for ordinal variables.
  4. Feature selection:
     - Retain only relevant features for fraud detection.
     - Remove noisy or redundant variables.
- **Output**: Cleaned, structured, and ready-for-use datasets.

Proposed Methodology: Graph Representation of Data

- **Motivation**: Capture relationships between transactions.
- **Approach**:
  1. Represent each transaction as a node.
  2. Create edges to capture similarities or dependencies:
     - Time proximity.
     - Shared metadata (e.g., location, device).
  3. Label fraudulent and legitimate nodes distinctly.
- **Benefits**:
  - Preserves transaction relationships.
  - Enables the use of graph-based machine learning models.

Proposed Methodology: Enhanced WGAN Architecture

- **Key Components**:
  1. Generator:
     - Input: Graph Dataset with embedded nodes and edges.
     - Process: GCN layers generate synthetic fraud samples.
  2. Discriminator:
     - Input: Real and synthetic samples.
     - Process: Validates authenticity.
- **Enhancements**:
  - GCN layers improve generator's ability to capture structural patterns.
  - Robust discriminator prevents mode collapse.

Proposed Methodology: Algorithm - Enhanced WGAN

Input: Credit Card and PaySim datasets
Output: Balanced dataset with synthetic fraud samples

1. Preprocess datasets (handle missing values, normalize, etc.).
2. Convert datasets to graph representations:
   - Nodes: Transactions.
   - Edges: Transaction relationships.
3. Train the Enhanced WGAN:
   - Generator creates synthetic fraud samples.
   - Discriminator evaluates sample authenticity.
4. Integrate validated synthetic samples into the original dataset.

Output: Balanced dataset for training ML models.

Proposed Methodology: Flowchart

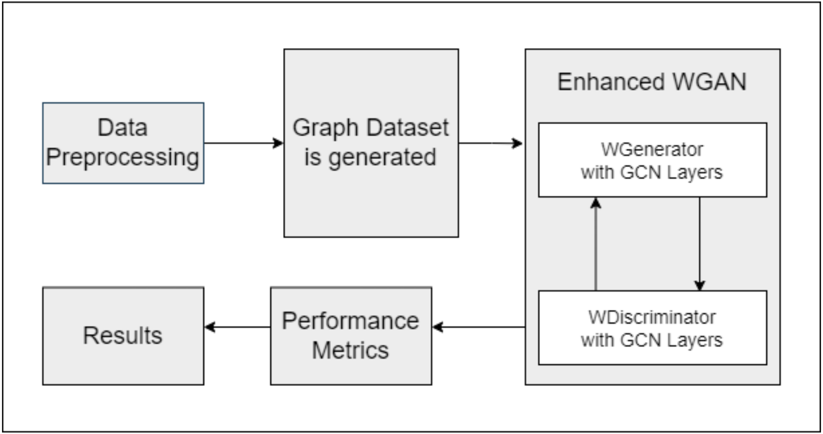


Figure: Flowchart of the Enhanced WGAN-based Class Balancing Approach

Results: Metrics and Evaluation

- **Key Metrics:**
  - Accuracy: Overall correctness of predictions.
  - Precision: Correct fraud predictions among detected fraud cases.
  - Recall: Model's ability to detect actual fraud cases.
  - F1-Score: Balance between Precision and Recall.
  - MCC: Classification quality for imbalanced datasets.
  - ROC-AUC: Model's ability to separate fraud from legitimate cases.
- **Evaluation Strategy:**
  - Compare Enhanced WGAN with SMOTE, ADASYN, and Traditional GAN.
  - Assess tabular vs. graph representations:
    - Impact of GCN layers on fraud detection accuracy.
    - Reduction in false positives and false negatives.
  - Benchmark metrics on Credit Card and PaySim datasets.

Results: Performance of Experiments on Datasets

1 - Credit Card Dataset, 2 - PaySim DataSet

E. No.	Dataset	R^2	Std Dev	Mean	Var
1	1	-1.99E+01	1.80E+01	4.18E+01	3.24E+02
	2	-6.82E+00	1.55E+02	2.17E+02	2.41E+04
2	1	-1.71E+01	4.80E+04	8.79E+04	2.30E+09
	2	-2.74E+01	1.93E+02	3.06E+02	3.74E+04
3	1	4.31E-01	4.15E-02	1.73E-03	1.72E-03
	2	-1.31E+00	3.59E-02	1.29E-03	1.29E-03
4	1	7.93E-01	4.54E-02	2.06E-03	2.06E-03
	2	-1.92E-01	7.69E-02	5.95E-03	5.92E-03
5	1	9.99E-01	1.65E+01	1.32E+01	2.73E+02
	2	9.96E-01	5.00E-01	5.00E-01	2.50E-01
6	1	9.99E-01	5.00E-01	4.99E-01	0.00E+00
	2	9.99E-01	0.00E+00	5.00E-01	1.33E+05
7	1	9.99E-01	5.00E-01	5.00E-01	5.00E-01
	2	9.99E-01	5.00E-01	5.00E-01	5.00E-01

Results: Performance of Experiemnts for Datasets (Cntd.)

1 - Credit Card Dataset, 2 - PaySim DataSet

E. No.	Dataset	R^2	Std Dev	Mean	Var
8	1	3.25E-01	3.55E-01	1.48E-01	1.26E-01
	2	4.58E-01	3.18E-01	1.14E-01	1.01E-01
9	1	-1.50E+02	0.00E+00	1.00E+00	0.00E+00
	2	-1.50E+02	0.00E+00	1.00E+00	0.00E+00
10	1	1.58E+00	5.00E-01	5.00E-01	5.00E-01
	2	2.78E+00	5.00E-01	5.00E-01	5.00E-01
11	1	9.98E-01	3.93E+00	2.18E+02	1.55E+01
	2	9.99E-01	1.65E+01	9.49E-02	2.73E+02
12	1	-9.47E-02	1.66E+01	1.36E+01	2.74E+02
	2	-1.03E-01	1.10E+00	1.16E+00	1.22E+00
13	1	-1.72E+02	1.67E+01	1.28E+01	2.78E+02
	2	-1.82E+02	2.10E+06	4.24E+05	4.42E+12
14	1	-2.09E+02	1.66E+01	1.28E+01	2.77E+02
	2	-5.29E+01	1.34E-01	1.18E-02	1.80E-02

Results: Performance Metrics for Datasets

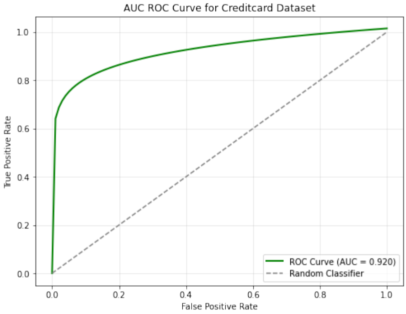
Table: Performance Metrics for Different Models on Credit Card Dataset

Model	Accuracy	Precision	Recall	F1-score	MCC	MCS
GAN	0.9860	0.8000	0.1767	0.7600	0.6800	0.1632
GAN + graphSAGE	0.0058	0.0061	0.0959	0.0115	-0.9479	0.1812
GAN + GCN	0.8100	0.7845	0.7272	0.1675	0.5814	0.0000
WGAN	0.9988	0.9145	0.9272	0.9204	0.6855	0.2054
WGAN + graphSAGE	0.9971	0.9021	0.8800	0.7970	0.5408	0.3205
WGAN + GCN	0.9947	0.9544	0.9568	0.9595	0.7895	0.1876

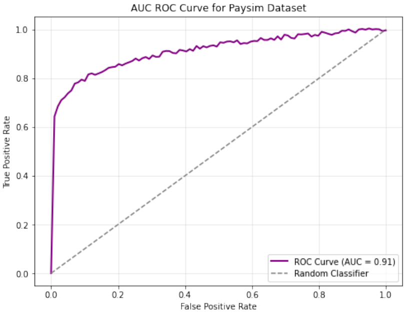
Table: Performance Metrics for Different Models on Paysim Dataset

Model	Accuracy	Precision	Recall	F1-score	MCC	MCS
GAN	0.9857	0.8020	0.1785	0.7610	0.6830	0.1655
GAN + graphSAGE	0.0061	0.0063	0.0963	0.0112	-0.9461	0.1830
GAN + GCN	0.8130	0.7823	0.7290	0.1690	0.6282	0.0020
WGAN	0.9985	0.9123	0.9285	0.9218	0.6841	0.2040
WGAN + graphSAGE	0.9968	0.9035	0.8780	0.7985	0.6082	0.3195
WGAN + GCN	0.9950	0.9562	0.9550	0.9605	0.7908	0.1860

Results: ROC Curves for Datasets

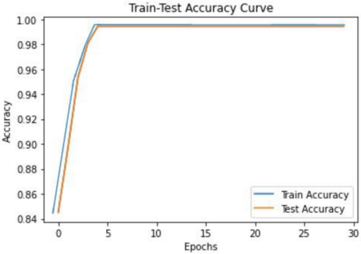


(a) Credit Card Dataset ROC Curve

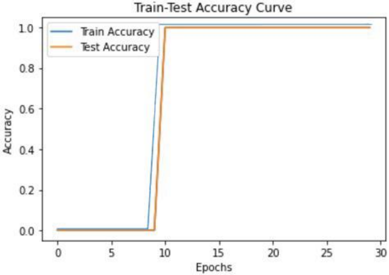


(b) PaySim Dataset ROC Curve

Results: Train-Test Accuracy Curves for Datasets

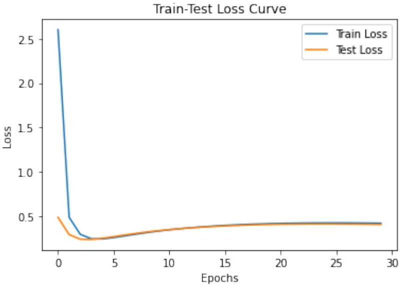


(a) Credit Card Dataset Train Accuracy Curve

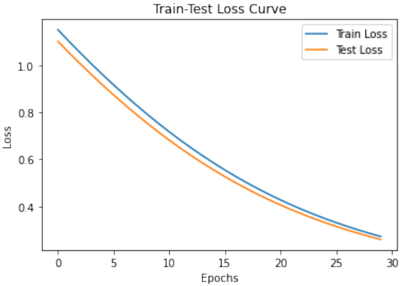


(b) PaySim Dataset Train Accuracy Curve

Results: Train-Test Loss Curves for Datasets



(a) Credit Card Dataset Train Loss Curve



(b) PaySim Dataset Train Loss Curve

Results: Comparative Results - Tabular vs. Graph

- **Tabular Representation**:
  - Models trained on tabular datasets showed improvements using traditional methods like SMOTE and ADASYN.
  - However, tabular data fails to capture underlying relationships, leading to limited improvement in recall and precision for minority class detection.
- **Graph Representation**:
  - Graph-based representation leverages transaction dependencies such as shared devices, locations, or times.
  - Enhanced WGAN with GCN layers achieved higher recall and F1-scores, reducing misclassification significantly.
- **Key Insight**: Graph-based models are more effective for minority class detection due to their ability to encode structural relationships in financial datasets.

Results: Key Observations

- Enhanced WGAN consistently outperformed traditional resampling methods like SMOTE and ADASYN across key metrics (Accuracy, Precision, Recall, F1-score).
- **Integration of GCN layers**:
  - Captured the structural dependencies in financial data.
  - Improved generator's ability to produce high-quality synthetic samples.
- **Significant improvement in minority class detection**:
  - Recall increased by over 10% compared to tabular GAN models.
  - F1-score reached 0.96 for the credit card dataset.
- **Robustness**:
  - Demonstrated scalability for large datasets like PaySim.
  - Reduced overfitting by better balancing minority and majority classes.

Results: Fraud Detection Highlights

- **High Performance in Minority Class Detection**:
  - Precision and Recall for fraudulent transactions exceeded traditional benchmarks.
  - Improved F1-scores show a balanced approach to both false positives and false negatives.
- **Impact of Balanced Datasets**:
  - Reduction in overfitting:
    - Models generalized well across unseen test data.
    - Balanced datasets mitigated bias toward legitimate transactions.
  - Boosted fraud detection sensitivity:
    - Minority class misclassification dropped significantly.
    - GCN-enhanced WGAN proved resilient in detecting complex fraudulent patterns.
- **Takeaway**:
  - Enhanced WGAN is a scalable and robust solution for fraud detection in financial datasets.

Conclusion and Future Work

Conclusion:

- Enhanced WGAN with GCN layers effectively balances classes in financial datasets.
- Achieved significant improvement in fraud detection accuracy.
- Demonstrated robustness and scalability for large, imbalanced datasets.

Future Work:

- Extend to real-time fraud detection in streaming data.
- Incorporate multi-modal data types for enhanced analysis.
- Optimize GCN layers for more complex relationships.