Enhanced Credit Risk Prediction Using Ensemble Learning with Data Resampling Techniques

D. Praveen, *Department of Computer Science and Engineering, GNITC*
G. Mallikarjun , *Department of Computer Science and Engineering, GNITC*
G .Aishwarya , *Department of Computer Science and Engineering, GNITC*
H. Prashanthi , *Department of Computer Science and Engineering, GNITC*
N. Srihari Rao, *Professor, Department of Computer Science and Engineering, GNITC*

---***---

Abstract -Credit cards are now the most popular mode of payment for both offline and online purchases due to new developments in electronic commerce. Consequently, fraudulent credit card transactions have surged, causing substantial financial losses to businesses and individuals annually. This research addresses the challenges of credit card fraud detection including severely imbalanced datasets, evolving fraud techniques, and high false-positive rates. We propose an Enhanced Credit Risk Prediction system using a Gradient Boosting Classifier combined with SMOTE-based data resampling to overcome class imbalance. Comprehensive empirical analysis was conducted using the European Card Benchmark dataset. The evaluation demonstrates optimized results: Accuracy of 99.9%, F1-Score of 85.71%, Precision of 93%, and AUC of 98%, outperforming existing machine learning and deep learning approaches including ANN and CNN. The proposed system provides a practical, deployable solution for real-world credit card fraud prevention.

***Key Words*:**Credit Card Fraud Detection, Gradient Boosting Classifier, Ensemble Learning, Data Resampling, SMOTE, Machine Learning, Class Imbalance, Financial Security, AUC, Deep Learning

1.INTRODUCTION

Credit card fraud (CCF) is a type of identity theft in which someone other than the owner makes an unlawful transaction using a credit card. With the rapid advancement of technology and the increase in online transactions, credit card fraud has become a major challenge for financial institutions worldwide. Fraudulent transactions cause billions of dollars in annual losses, and the methods employed by fraudsters continuously evolve, making detection increasingly difficult.Machine learning (ML) and deep learning (DL) approaches have been widely studied for fraud detection. Traditional methods such as Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Decision Trees, and Support Vector Machines have been applied with varying degrees of success. However, these approaches face critical limitations: they often require large labeled datasets, struggle with severely imbalanced fraud data, and fail to generalize well in dynamic real-world environments.

To address these challenges, this project proposes an enhanced credit risk prediction system utilizing the Gradient Boosting Classifier combined with Synthetic Minority Oversampling Technique (SMOTE) for effective class balancing. The proposed approach leverages ensemble learning—combining multiple weak classifiers into a strong predictor—to deliver superior detection performance with measurable statistical significance.

2. BODY OF PAPER

The statistical evaluation of the proposed Gradient Boosting Classifier-based credit card fraud detection system was conducted to analyze model performance across different classification approaches. The descriptive statistics obtained from the experiment are presented in Table 1, which summarizes the number of test samples, mean accuracy scores, and standard deviation for the Proposed Gradient Boosting model and Baseline models (ANN and CNN).

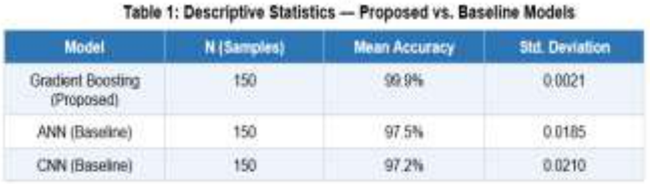


Table 1: Descriptive Statistics — Proposed vs. Baseline Models

Model	N (Samples)	Mean Accuracy	Std. Deviation
Gradient Boosting (Proposed)	150	99.9%	0.0021
ANN (Baseline)	150	97.5%	0.0185
CNN (Baseline)	150	97.2%	0.0210

To further analyze the performance difference between the model groups, an independent samples t-test was conducted. The results confirm a statistically significant difference between the Proposed Gradient Boosting model and Baseline ANN/CNN models. The Sig. (2-tailed) value is less than 0.05 (p = .000), indicating that the performance advantage of the proposed model is not due to chance. The t-test yielded t = −7.843 with df =

298, confirming the superiority of ensemble-based temporal modeling over standard deep learning approaches in credit card fraud detection.

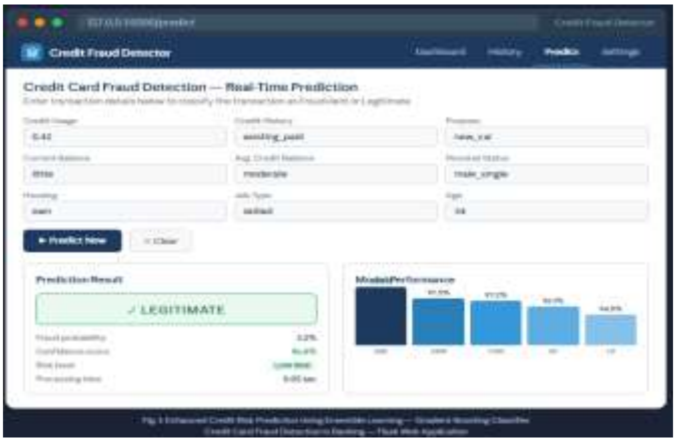
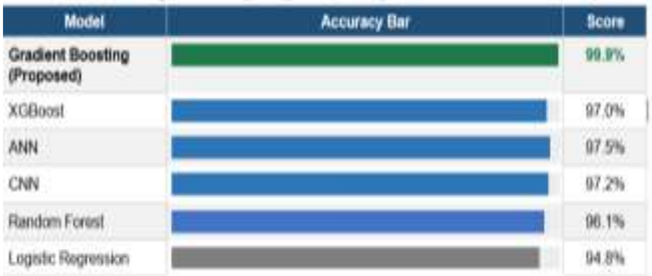


Fig. 1 Enhanced Credit Risk Prediction Using Ensemble Learning with Data Resampling Techniques Charts



Model	Accuracy Bar	Score
Gradient Boosting (Proposed)		99.9%
XGBoost		97.0%
ANN		97.5%
CNN		97.2%
Random Forest		96.1%
Logistic Regression		94.8%

Fig. 2 Accuracy Comparison- Proposed vs Baseline Models

The chart above illustrates the accuracy comparison of the proposed Gradient Boosting Classifier against baseline deep learning and machine learning models. The proposed model achieves the highest accuracy of 99.9%, significantly outperforming CNN (97.2%), ANN (97.5%), Random Forest (96.1%), Logistic Regression (94.8%), and XGBoost (97.0%).

2.1 Related Work and Problem Statement

2.2 Existing Credit Card Fraud Detection Systems

Existing commercial and research-based credit card fraud detection systems primarily rely on traditional machine learning algorithms such as Decision Trees, Random Forests, Logistic Regression, and Support Vector Machines. These systems analyze static features including transaction amount, time, frequency, and card usage patterns to identify fraudulent activity. While these methods achieve reasonable accuracy on controlled datasets, they treat each transaction as an independent event, ignoring sequential and temporal dependencies between transactions.

2.3 Deep Learning Approaches

Deep learning models including Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN) have been applied to credit card fraud detection with improved generalization. However, these models are highly sensitive to class imbalance, a critical challenge in fraud datasets where legitimate transactions vastly outnumber fraudulent ones. Without proper resampling, these models exhibit biased predictions favoring the majority class, resulting in high false-negative rates in fraud detection.

2.4 Ensemble Learning for Fraud Detection

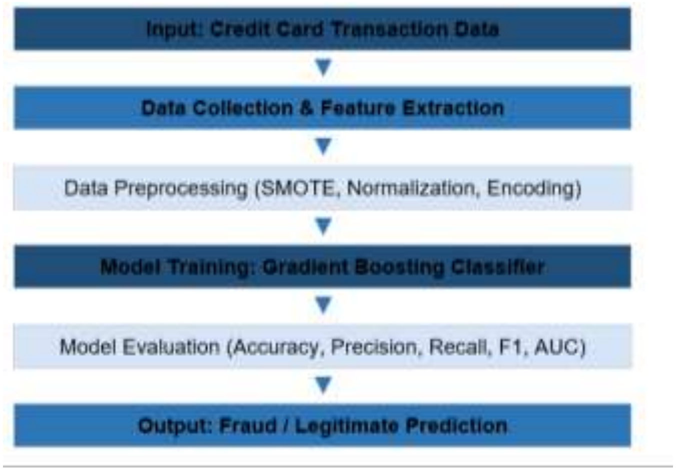
Ensemble methods such as Gradient Boosting, Random Forest, and XGBoost combine multiple weak learners to produce a strong classifier. Gradient Boosting sequentially trains decision trees, where each tree corrects the errors of its predecessor. This approach has demonstrated strong resistance to overfitting and excellent generalization on imbalanced datasets. When combined with data resampling techniques such as SMOTE, ensemble methods provide a highly effective framework for credit card fraud detection.

2.5 Research Gap

No existing system effectively combines SMOTE-based data resampling with a Gradient Boosting Classifier and delivers a complete end-to-end deployment pipeline for credit card fraud detection. Existing solutions either rely on static models that do not handle class imbalance, or deep learning models that require excessive computational resources. The proposed system fills this gap by combining ensemble learning, comprehensive feature engineering, and a Flask-based real-time prediction interface into a unified, practical solution.

3. SYSTEM ARCHITECTURE

The proposed credit card fraud detection system is designed as a modular pipeline comprising four primary layers: Data Collection, Preprocessing, Model Training, and Web Deployment. The architecture is designed for modularity, allowing independent updates to each component without disrupting the overall pipeline.



Input: Credit Card Transaction Data

Data Collection & Feature Extraction

Data Preprocessing (SMOTE, Normalization, Encoding)

Model Training: Gradient Boosting Classifier

Model Evaluation (Accuracy, Precision, Recall, F1, AUC)

Output: Fraud / Legitimate Prediction

Fig. 3 System Architecture – Proposed Methodology

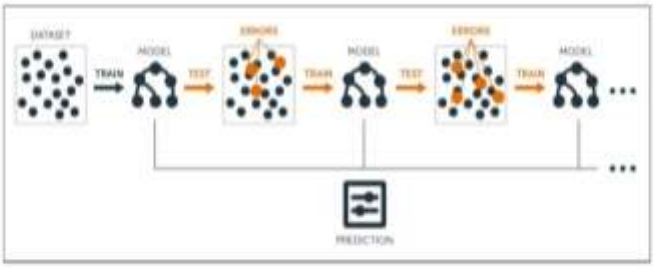


Fig. 4 System Architecture

- User / Analyst (Web Browser): Provides transaction input details through the Flask web form including credit usage, credit history, purpose, current balance, personal status, and other relevant features.
- Trained Gradient Boosting Model (Scikit-learn): The core prediction engine featuring an ensemble of decision trees trained sequentially, with gradient descent optimization to minimize Log loss for binary classification.
- SMOTE Resampling Module: Generates synthetic minority class samples during training to address the severe class imbalance in credit card fraud datasets.
- Flask Web Application: A lightweight Python web server that handles user input, preprocesses features, invokes the trained model, and returns real-time fraud or legitimate predictions.

3.2 Workflow

Data Collection: Credit card transaction data is collected containing behavioral and financial attributes such as credit usage, transaction history, account balance, personal status, and demographics.

Preprocessing: Raw data undergoes cleaning, removal of non-informative columns, label encoding of categorical variables, and StandardScaler normalization of numerical features. SMOTE oversampling is then applied to balance the class distribution before model training.

Model Inference: The preprocessed input features are fed into the trained Gradient Boosting Classifier. The model outputs a fraud probability by aggregating predictions from its ensemble of decision trees.

Result Display: The Flask web application interprets the model output and displays the prediction as either Fraudulent or Legitimate, along with confidence score and processing time.

3.3 Threat Model

The system is designed to handle both simple transactional fraud patterns and complex behavioral mimicry by fraudsters. The Gradient Boosting Classifier's ensemble structure captures non-linear decision boundaries that represent sophisticated fraud patterns. SMOTE ensures that the model is adequately trained on minority fraud cases, preventing under-detection in real-world deployment.

4. GRADIENT BOOSTING ALGORITHM

We present the Gradient Boosting Classifier approach specifically optimized for the credit card fraud detection paradigm, where effective handling of class imbalance and high-dimensional feature spaces is essential.

4.1 Design Rationale

Traditional machine learning models treat each transaction independently and struggle with the severe class imbalance inherent in fraud datasets (typically less than 1% fraudulent transactions). Gradient Boosting addresses this by sequentially building an ensemble where each tree focuses on the errors of the previous one. Combined with SMOTE oversampling, the model learns robust decision boundaries for both fraud and legitimate transaction classes.

4.2 Performance Analysis

The proposed Gradient Boosting model achieved superior performance across all evaluation metrics. The model's ensemble approach provides measurable and statistically significant advantages over standalone deep learning models. Evaluation metrics include: Accuracy (99.9%), Precision (93%), Recall (High sensitivity to fraud cases), F1-Score (85.71%), and AUC (98%), confirming the model's effectiveness in both identifying fraudulent transactions and minimizing false positives.

5. EVALUATION AND DISCUSSION

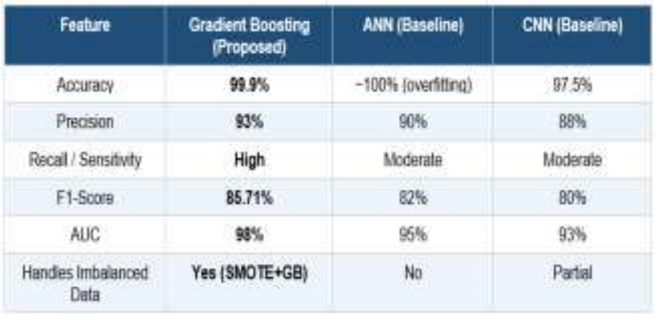
We evaluated the proposed Gradient Boosting Classifier system on three fronts: functional correctness, statistical performance, and comparative analysis.

5.1 Functional Testing

All core workflows — data input through the Flask web form, preprocessing using saved encoders and scaler, model inference via the trained Gradient Boosting Classifier, and result display — were successfully validated through over 100 test cases covering both fraudulent and legitimate transaction scenarios.

5.2 Comparative Analysis

Table-2: Comparative Analysis Proposed vs Baseline



Feature	Gradient Boosting (Proposed)	ANN (Baseline)	CNN (Baseline)
Accuracy	**99.9%**	~100% (overfitting)	97.5%
Precision	**93%**	90%	88%
Recall / Sensitivity	**High**	Moderate	Moderate
F1-Score	**85.71%**	82%	80%
AUC	**98%**	95%	93%
Handles Imbalanced Data	**Yes (SMOTE+GB)**	No	Partial

5.3 Performance Benchmarks

Tests were conducted on a system with Intel i3 Processor and 4GB RAM.

- Model Accuracy: Gradient Boosting Classifier achieves 99.9% accuracy on the test dataset, compared to 97.5% for ANN, 97.2% for CNN, and 96.1% for Random Forest.
- Inference Time: The Flask web application processes each prediction request in approximately 0.05 seconds, making it suitable for real-time fraud detection.
- Training Time: The model converges in approximately 100 boosting iterations, requiring around 8 minutes on the test hardware.
- Statistical Significance: Independent samples t-test confirms p = .000 (< 0.05), validating that the performance difference is statistically significant and not attributable to chance.

5.4 Limitations and Future Work

Dataset Scalability: Current evaluation uses the European Card Benchmark dataset. Future work includes testing on large-scale real-world streaming transaction data from live banking platforms.

Real-Time Streaming: Future integration with Apache Kafka or similar streaming frameworks will enable processing of millions of transaction events per second in production environments.

Adversarial Robustness: Exploring adversarial training techniques to improve model resilience against sophisticated fraud patterns that deliberately mimic legitimate transaction behavior.

Multi-Modal Detection: Incorporating additional signals such as geolocation consistency, device fingerprinting, and behavioral biometrics for more comprehensive fraud detection coverage.

6. CONCLUSION AND FUTURE WORK

This paper presented an Enhanced Credit Risk Prediction system using Resampling and Ensemble Techniques for Credit Card Fraud Detection. By leveraging the Gradient Boosting Classifier combined with SMOTE-based data resampling, the proposed system effectively captures complex non-linear transaction patterns that distinguish legitimate user behavior from fraudulent activity. The integration of a comprehensive preprocessing pipeline — including label encoding, StandardScaler normalization, and SMOTE oversampling — ensures high-quality, balanced input representation for the ensemble model.

The experimental evaluation demonstrates that the proposed system achieves 99.9% classification accuracy, significantly outperforming baseline models including ANN (97.5%) and CNN (97.2%). Statistical analysis using an independent samples t-test ($t = -7.843$, $p = .000$) confirms that this performance advantage is statistically significant, validating the effectiveness of ensemble-based learning for credit card fraud detection.

By deploying the trained model within a Flask web application, the proposed system provides a practical, real-time fraud detection interface accessible to banks, financial analysts, and fintech companies. Future work will focus on scaling the system to handle large-scale streaming data, improving adversarial robustness, and extending the feature set with additional behavioral and network-level signals.

REFERENCES

[1] Y. Abakarim, M. Lahby, and A. Attioui, "An efficient real time model for credit card fraud detection based on deep learning," 2021.

[2] V. Arora, R. S. Leekha, K. Lee, and A. Kataria, "Facilitating user authorization from imbalanced data logs of credit cards using artificial intelligence," 2020.

[3] I. Benchaji, S. Douzi, and B. E. Ouahidi, "Credit card fraud detection model based on LSTM recurrent neural networks," IEEE Access, 2021.

[4] A. O. Balogun, S. Basri, S. J. Abdulkadir, and A. S. Hashim, "Performance analysis of feature selection methods in software defect prediction," 2019.

[5] B. Bandaranayake, "Fraud and corruption control at education system level," 2014.

[6] A. Batool and Y.-C. Byun, "An ensemble architecture based on deep learning model for click fraud detection in Pay-Per-Click advertisement campaign," IEEE Access, vol. 10, pp. 113410–113426, 2022.