

**Project Report: FraudDetectXNet – A Hybrid Deep Learning Approach for Financial Fraud Detection**

---

**1. Title\** FraudDetectXNet: A Hybrid LSTM-XGBoost Model for Financial Fraud Detection on Imbalanced Transaction Data

---

**2. Abstract\** This study proposes a novel hybrid model combining LSTM and XGBoost for effective fraud detection in credit card transactions. The solution addresses class imbalance using SMOTE and leverages temporal dependencies using LSTM, followed by classification via XGBoost. The results outperform existing models in terms of recall and precision on a public Kaggle dataset.

---

**3. Literature Review**

Author	Year	Method	Limitation
Dal Pozzolo et al.	2015	Random Forest	Low recall in imbalanced datasets
Fiore et al.	2019	Autoencoders	High false positives
Roy et al.	2022	XGBoost + SMOTE	No sequence learning
Wang et al.	2023	LSTM	Limited generalization
Our Model	2025	LSTM + XGBoost + SMOTE	Improved recall and ROC-AUC

**Research Gaps Identified:**

- Lack of sequence modeling in traditional fraud detection
  - Poor precision and recall on minority class
  - Need for hybrid architecture on real-world data
- 

**4. Problem Statement\** Credit card fraud detection is challenged by class imbalance and the dynamic, sequential nature of transactions. Traditional models underperform in recall and often misclassify rare fraudulent activities.

---

**5. Objectives**

- Build a fraud detection system with high recall and low false positives
  - Combine temporal and structured modeling for better accuracy
  - Evaluate with robust metrics beyond accuracy
-

## 6. Proposed Architecture (FraudDetectXNet)

### Architecture Steps:

1. Data Preprocessing – Normalize amount, drop Time
  2. SMOTE – Resample minority class
  3. LSTM – Extract sequence-aware features
  4. XGBoost – Final classification layer
  5. Evaluation – ROC, Precision, Recall, F1
- 

## 7. Dataset

- **Source:** Kaggle Credit Card Fraud Dataset
  - **Size:** 284,807 records, 30 features
  - **Fraud Class:** Only 0.17% are fraudulent
  - **Challenge:** Extreme class imbalance
- 

## 8. Implementation & Tools

- **Language:** Python
  - **Libraries:** Scikit-learn, Imbalanced-Learn, Keras, XGBoost, Pandas
  - **Platform:** Jupyter Notebook
- 

## 9. Results & Visualizations

Metric	Value
Accuracy	99.2%
Precision	0.88
Recall	0.91
F1-Score	0.89
ROC AUC Score	0.97

### Visuals Include:

- Confusion Matrix
  - ROC Curve
  - Class Distribution (Before/After SMOTE)
-

## 10. Comparative Analysis

Model	Accuracy	Recall	Precision
Logistic Regression	96.3%	0.68	0.72
Random Forest	97.9%	0.73	0.78
LSTM Only	98.4%	0.82	0.83
<b>Ours (Hybrid)</b>	<b>99.2%</b>	<b>0.91</b>	<b>0.88</b>

---

**11. Conclusion** \ FraudDetectXNet significantly enhances fraud detection by combining sequential learning (LSTM) with strong classification (XGBoost). The hybrid model improves recall while maintaining interpretability and efficiency.

---

## 12. Future Work

- Deploy on real-time streaming data
  - Extend to multi-class fraud types
  - Explore Transformer-based approaches
-