SENIOR DESIGN PROJECT
END-TERM PRESENTATION

FINANCIAL FRAUD DETECTION SYSTEM USING MACHINE LEARNING AND DEEP LEARNING

Supervised By: Ms. Anuradha Mohanta

Group No.: V5

Name of the Student(s) with Regd. No.:

Aditya Narayan Sahu - 2141019163
Satyabrat Sahoo - 2141019117
Soumya Ranjan Patra - 2141016213
Achyuta Samantaray - 2141019198

Department of Computer Sc. and Engineering
Faculty of Engineering & Technology (ITER)
Siksha 'O' Anusandhan (Deemed to be) University
Bhubaneswar, Odisha

Presentation Outline

- Introduction
  - Project Overview and Problem Statement
  - Objectives and Motivation
- Background & Related Work/ Literature Review
  - Existing Solutions/Related Work & Their Limitations/Research Gaps
- Proposed Solution & Architecture
  - System Architecture /Workflow Diagram/Model Diagram/Block Diagram/Schematic Layout
  - Description of Key Components/Features & Modules
- Implementation Details
  - Algorithms and Methods Used/Technologies & Platforms, Frameworks, and Tools Used
- Results and Analysis
  - Test Results– Performance Metrics /System Outputs and Screenshots
  - Performance Comparison/Interpretation of Results/Result Validation
- Conclusion & Future Work
  - Key Findings
  - Scope for Improvement or Extensions
- Bibliography

Introduction

▪ Project Overview

- Real-time detection serves as its main purpose to fight fraudulent financial transactions.
- Leverages **Machine Learning** (ML) and **Deep Learning** (DL) techniques for enhanced accuracy.
- Addresses limitations of **traditional rule-based systems**, which fail against evolving fraud patterns.
- Uses real-world, highly **imbalanced transaction data** for realistic model training.
- Implements algorithms like **Long-Short Term Memory(LSTM), Random Forest, Recurrent Neural Network(RNN).**
- Aims to create a system that is **adaptive, scalable, and secure** for integration into digital payment platforms.

Project Overview

- Financial Frauds can be categorized into three types:

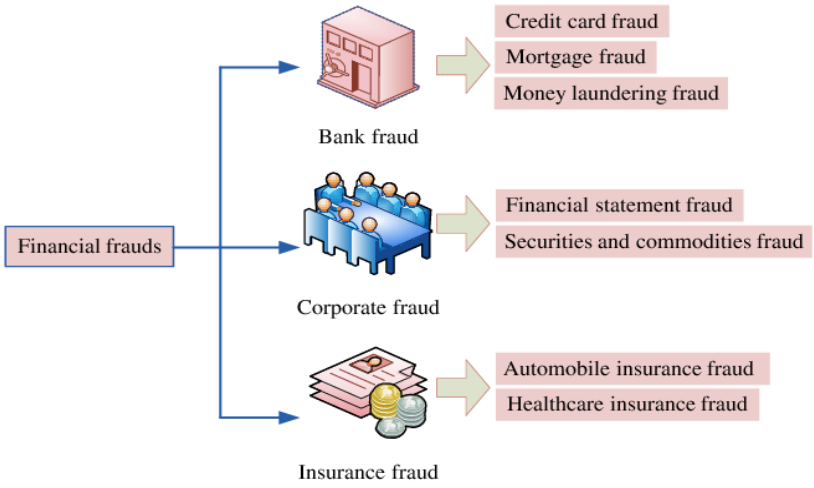

Fig1. *Types of Financial Frauds*

Introduction

▪ Problem Statement

- Traditional rule-based systems are ineffective against evolving fraud patterns.
- Increasing digital transactions make fraud risks progressively more severe.
- **Challenge**:- Highly imbalanced data (fraud cases < 0.2%).
- **Goal**:- Build a model that accurately detects fraud in real-time while minimizing false alarms.

Introduction

Objectives

- Accurately identify fraudulent activities.
- The technology aims to decrease both incorrect fraud alarm reports and monetary losses.
- The system aims to enhance the public's faith in electronic financial processes.
- Improve accuracy and reduce false alarms Achieve high recall to catch more fraud cases.
- Handle imbalanced dataset effectively Use techniques like **Synthetic Minority Oversampling Technique(SMOTE)** to balance the training data.

Introduction

▪ Motivation

- Growing threat of digital payment fraud Billions lost every year due to undetected frauds.
- Traditional systems are rule-based and limited.
- Machine Learning offers intelligent, adaptive detection, Machine learning (ML) Learns from patterns in data and can detect unknown fraud types.
- Personal and societal impact Reduces financial loss, increases user trust and supports Digital Banking Infrastructure.

Literature Survey

Table 1: Literature Survey

Sl. No.	Title of Paper/Study	Author(s)	Method/Approach Used	Key Findings / Contributions
1	Transforming banking security: the role of deep learning in fraud detection systems[1]	Md Al-Imran, Eftekhar Hossain Ayon	SMOTE, LSTM, XGBoost	LSTM network demonstrates its capability to learn complex patterns over time, making it a powerful tool in the fight against financial fraud.
2	Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy[2]	A. Dal Pozzolo et al.	Cost-sensitive learning, Random Forest, SVM	Introduced effective handling of imbalanced datasets in fraud detection.
3	Financial Fraud Detection Based on Machine and Deep Learning: A Review[3]	Rojan Zaki Abdulkreem, Adnan Mohsin Abdulazeez	RNN, LSTM	Utilization of cutting-edge deep learning models to detect financial fraud. Different technologies used nowadays.
4	Machine Learning Approach for Fraud Detection System in Financial Institution: A Web Base Application [4]	D. O. Njoku , V. C. Iwuchukwu, J. E. Jibiri	Logistic Regression	a sophisticated fraud detection system for account transactions, integrating machine learning or rules, user engagement, and streamlined backend processing.
5	The Application of Data Mining Techniques in Financial Fraud Detection[5]	E. Ngai et al.	Survey of ML/DM techniques	Provided a classification framework for different approaches in financial fraud detection.

Contd…

Sl. No.	Title of Paper/Study	Author(s)	Method/Approach Used	Key Findings / Contributions
6	Credit Card Fraud Detection Model Based on LSTM Recurrent Neural Networks [6]	Ibtissam Benchaji, Samira Douzi, and Bouabid El Ouahidi	RNN, LSTM, DL	a sequence classifier based on the LSTM networks to catch the consumer behavior of individual cardholders when constructing a credit card fraud detection model.
7	Enhancing Financial Fraud Detection with Hybrid Deep Learning and Random Forest Algorithms[7]	Aravind Kumar Kalusivalingam, Amit Sharma, Neha Patel, Vikram Singh	DL, Random forest	The integration of hybrid deep learning and Random Forest algorithms presents a promising advancement in the domain of financial fraud detection.
8	Enhancing Performance of Credit Card Model by Utilizing LSTM Networks and XGBoost Algorithms[8]	Kianeh Kandi, Antonio García-Dopico	LSTM, XGBoost, SMOTE, RNN	LSTM model demonstrates a clear advantage when dealing with imbalanced datasets. XGBoost has low accuracy and precision as compared to LSTM.
9	Year-over-Year Developments in Financial Fraud Detection via Deep Learning: A Systematic Literature Review[9]	Yisong Chen1 , Chuqing Zhao2 , Yixin Xu3 , Chuanhao Nie	Difference in Technologies like LSTM, RNN, NLP, Logistic Regression	analyzing recent advancements, it becomes clear that deep learning models, including CNNs, LSTMs, transformers, and ensemble techniques.

Contd…

Sl. No.	Title of Paper/Study	Author(s)	Method/Approach Used	Key Findings / Contributions
10	A Comprehensive Framework for Strengthening USA Financial Cybersecurity: Integrating Machine Learning and AI in Fraud Detection Systems[10]	1Oluwabusayo Adijat Bello; Adebola Folorunso2;	Machine Learning and Artificial Intelligence	It provides an idea of how existing systems can be improved using ML and AI
11	Enhanced credit card fraud detection based on attention mechanism and LSTM deep model[11]	Ibtissam Benchaji, Samira Douzi, Bouabid El Ouahidi	LSTM	Effectiveness and efficiency of LSTM models
12	Enhancing Performance of Financial Fraud Detection Through Machine Learning Model[12]	Eswar Prasad Galla1*, Hemanth Kumar Gollangi2	ANN, SVM, Decision Tree	the effectiveness of ML models, particularly ANNs, in improving financial fraud detection.
13	The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature[13]	E.W.T. Ngai a, Yong Hu b, Y.H. Wong a, Yijun Chen	Use of Financial Fraud Detection(FFD), Financial Fraud Prevention(FFP)	Enhancing Financial Fraud Detection(FFD) with the help of Financial Fraud Prevention(FFP)

Background & Related Work

▪ Related Work & Their Limitations

- Rule-Based Fraud Detection Systems Use pre-defined conditions and thresholds (e.g., amount > ₹10,000 → flag as fraud).[9]
- Traditional Machine Learning Models like **Decision Trees** and **Random Forest** have been used.[7]
- Sequences of data benefit from analysis through LSTM and Autoencoders which are part of the DL model family.

Background & Related Work

Limitations of Existing Systems

- **Static** and **Inflexible**.
- High **False Positives:** Many legitimate transactions are incorrectly flagged as fraud.
- **Lack of Real-Time Capability:** Traditional methods often work offline or after the fraud has already occurred.
- **Inability to Detect New/Evolving Fraud Patterns:** Rule-based and older Machine Learning (ML) models fail to generalize to unseen fraud behavior.

Proposed System Architecture

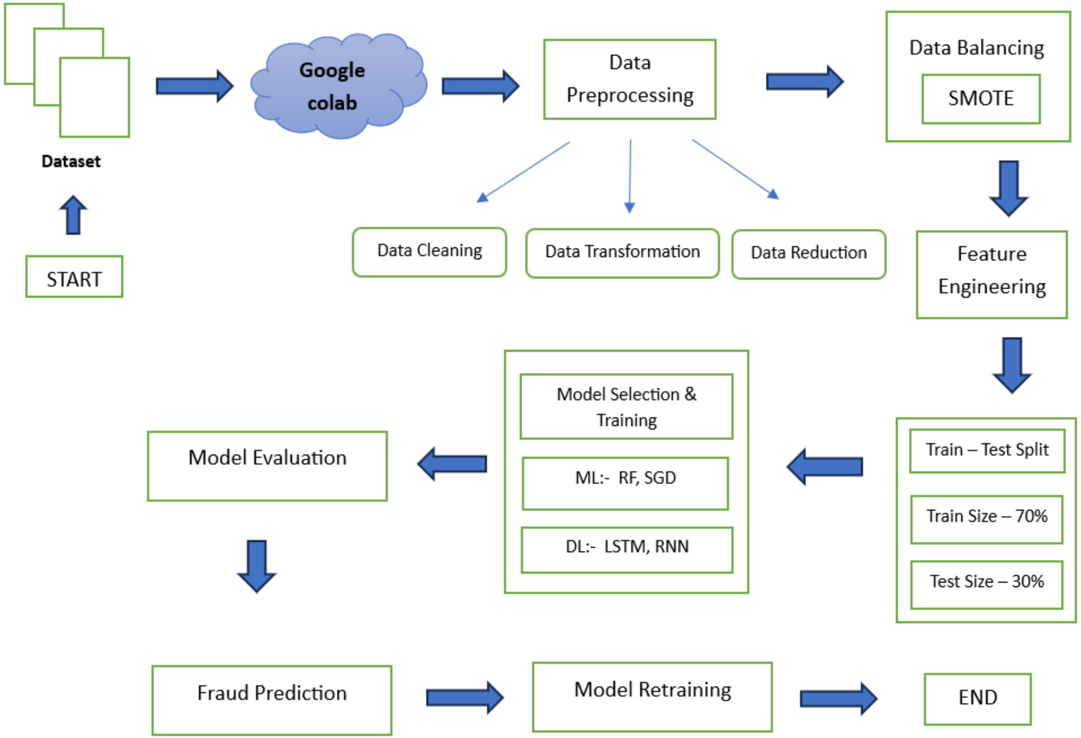

Fig2. *Sequential Flow Diagram of Model Training*

Description of Key Components

Key Features

- Data Preprocessing Module **Standard Scaler** for normalization.
- SMOTE applied to handle data imbalance Machine Learning / Deep Learning Model Module.
- Feature Engineering Used V1 to V28 (Principal Component Analysis(PCA)-based), Time, and Amount.
- Irrelevant features removed to reduce noise.

Dataset Description

The financial fraud dataset is taken from Kaggle[14] and it has the following features :

- Contains **284,807 credit card transactions** from European cardholders over 2 days.
- Contains Banking Sim data, approximately **5,959,193 transactions.**
- Only **8400 transactions (0.135%)** are fraudulent – highly **imbalanced dataset**.
- Features are **PCA-transformed** (V1–V28) to ensure **data privacy**.
- Includes Time, Amount, and Class (0 = genuine, 1 = fraud) fields.
- Provided by the **Machine Learning Group at Université Libre de Bruxelles (ULB)**.
- Widely used and considered **reliable for training and evaluating fraud detection models**.

Implementation Details

- ## Algorithms and Methods Used

  - **Machine Learning Models** that classify based on transaction features → Random Forest, and Stochastic Gradient Descent (SGD) Classifier were implemented.
  - **Random Forest** → Ensemble model that handles non-linearity well and reduces overfitting, offering high accuracy.
  - **Stochastic Gradient Descent (SGD) Classifier** → Efficient for large-scale and high-dimensional data, useful for real-time fraud detection.

Implementation Details

Algorithms and Methods Used

- **Deep Learning Models like** that learn LSTM (Long Short-Term Memory) and RNN (Recurrent Neural Network) is used to learn sequential transaction patterns.
- **LSTM** (Long Short-Term Memory) → A type of RNN specialized for learning long-term dependencies in sequential data like transaction histories in fraud detection.
- **RNN** (Recurrent Neural Network) → Designed for sequential data; captures short-term patterns and is useful in tasks like time-series forecasting and speech recognition.

Implementation Details

- Technologies, Frameworks and Tools Used
  - Hardware Specifications:
    - **Operating System:** Windows or Linux
    - **Processor:** Intel Core i5 11th Gen (64-bit)
    - **Ram: 8** GB
    - **Pre-installed Software:** Python 3.10 or above
    - **Virtual Environment:** Google Colaboratory, Kaggle Cloud (For GPU-intensive models)
    - **Server Infrastructure:** Secure servers for real-time transactions processing

Contd…

Software Specifications:

- **Programming Language Python** – Core language used for data science and model development.
- **Libraries & Frameworks NumPy, Pandas** – Data manipulation and analysis.
- **Scikit-learn** – ML models, evaluation metrics, SMOTE (Synthetic Minority Over-sampling Technique).
- **TensorFlow** – Deep learning models (LSTM, RNN) .
- **Matplotlib, Seaborn** – Data visualization.
- **Development Platform Google Colaboratory** – Cloud-based training of models using GPU.

Contd…

Performance Metrics

- ***Accuracy* -** The ratio of all the true results including both true positives and true negatives to the total number of cases under examinations.

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

- ***Precision* -** It estimates the probability of a genuine prediction being correct.

$$Precision = TP/(TP + FP)$$

- **Recall –** It is often called sensitivity, is defined as a ratio of TP to total real positives, as seen in equation:

$$Recall = TP/(TP + FN)$$

- ***ROC(Receiver Operating Characteristic) curve and AUC(Area Under the Curve) Score –*** ROC curve is TPR vs FPR graph and the area under the graph represents the AUC score.

Confusion Matrix

The figure below contains the confusion matrix of our proposed model i.e., LSTM(SMOTE)

Table2: Confusion Matrix

Actual	**Predicted**	
	Fraud	**Non-Fraud**
Fraud	2361	74
Non-Fraud	5978	1900373

Result & Analysis

- ## Recurrent Neural Network(RNN)

The given table 3 and table 4 shows the results before and after applying SMOTE

Table 3

Performance Metric	Result
Accuracy	0.9995
Precision	0.9595
Recall	0.6320
F1-Score	0.7621
ROC AUC	0.9954

Table 4

Performance Metric	Result
Accuracy	0.9907
Precision	0.1190
Recall	0.9799
F1-Score	0.2123
ROC AUC	0.9982

Long-Short Term Memory(LSTM)

The given table 5 and table 6shows the results before and after applying SMOTE

Table 5

Performance Metric	Result
Accuracy	0.9995
Precision	0.9803
Recall	0.6345
F1-Score	0.7704
ROC AUC	0.9945

Table 6

Performance Metric	Result
Accuracy	0.9969
Precision	0.2862
Recall	0.9725
F1-Score	0.4422
ROC AUC	0.9983

Random Forest

The given table 7 and table 8 shows the results before and after applying SMOTE.

Table 7

Performance Metric	Result
Accuracy	0.9996
Precision	0.9937
Recall	0.7203
F1-Score	0.8352
ROC AUC	0.9966

Table 8

Performance Metric	Result
Accuracy	0.9844
Precision	0.0749
Recall	0.9881
F1-Score	0.1392
ROC AUC	0.9984

Stoichastic Gradient Descent(SGD)

The given table 9 and table 10 shows the results before and after applying SMOTE

Table 9

Performance Metric	Result
Accuracy	0.9988
Precision	0.9773
Recall	0.0883
F1-Score	0.1620
ROC AUC	0.9341

Table 10

Performance Metric	Result
Accuracy	0.9324
Precision	0.0170
Recall	0.9166
F1-Score	0.0335
ROC AUC	0.9816

Output Screenshots

The figure provided below contains the outputs before applying SMOTE.

```
Accuracy : 0.9995
Precision: 0.9595
Recall   : 0.6320
F1 Score : 0.7621
ROC AUC  : 0.9954
```

Fig4. *RNN*

```
Accuracy  : 0.9995
Precision: 0.9803
Recall    : 0.6345
F1 Score  : 0.7704
ROC AUC   : 0.9945
```

Fig5. *LSTM*

```
The model used is Random Forest classifier
The accuracy is 0.9996374659076502
The precision is 0.9937677053824363
The recall is 0.7203285420944558
The F1-Score is 0.8352380952380952
The Matthews correlation coefficient is0.8459174190512312
ROC AUC Score: 0.9966
```

Fig6. *Random Forest*

```
Accuracy: 0.9988
Precision: 0.9773
Recall: 0.0883
F1 Score: 0.1620
ROC AUC Score: 0.9341
```

Fig7. *SGD*

Contd…

The figure below contains the outputs after applying SMOTE.

```
Accuracy : 0.9907
Precision: 0.1190
Recall    : 0.9799
F1 Score : 0.2123
ROC AUC   : 0.9982
```

Fig8. *RNN*

```
Accuracy : 0.9969
Precision: 0.2862
Recall    : 0.9725
F1 Score : 0.4422
ROC AUC   : 0.9983
```

Fig9. *LSTM*

```
The model used is Random Forest Classifier with SMOTE
Number of true positive samples (outliers): 2435
Number of prediction errors: 29756
Accuracy: 0.9844
Precision: 0.0749
Recall: 0.9881
F1 Score: 0.1392
Matthews Correlation Coefficient: 0.2698
ROC AUC Score: 0.9984
```

Fig10. *Random Forest*

```
Accuracy: 0.9324
Precision: 0.0170
Recall: 0.9166
F1 Score: 0.0335
ROC AUC Score: 0.9816
```

Fig11. *SGD*

ROC AUC Curves for the models

The figure below shows the direction of the ROC curves before applying SMOTE.

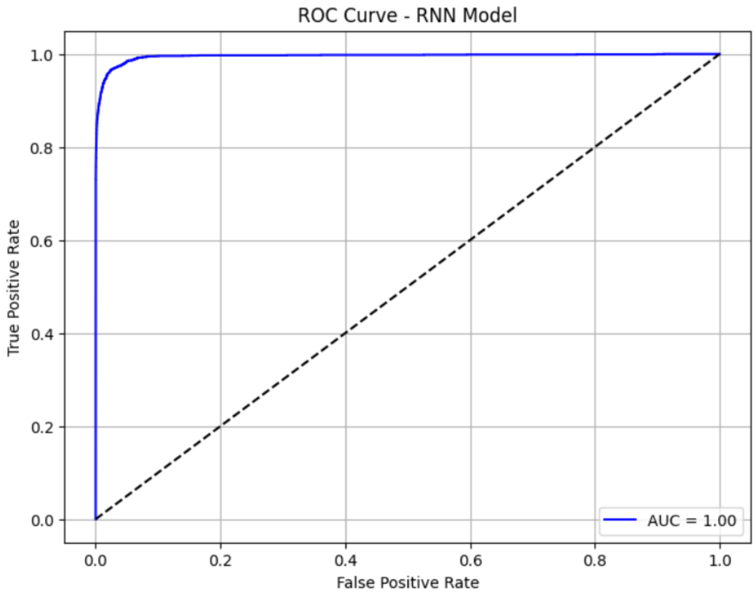

Fig12. *RNN Curve*

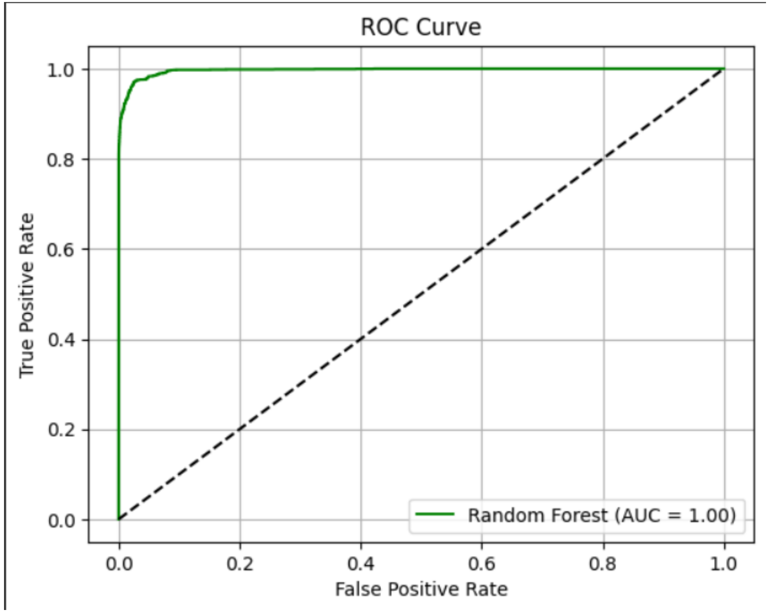

Fig13. *Random Forest Curve*

Contd…

The figure below shows the direction of the ROC curves before applying SMOTE.

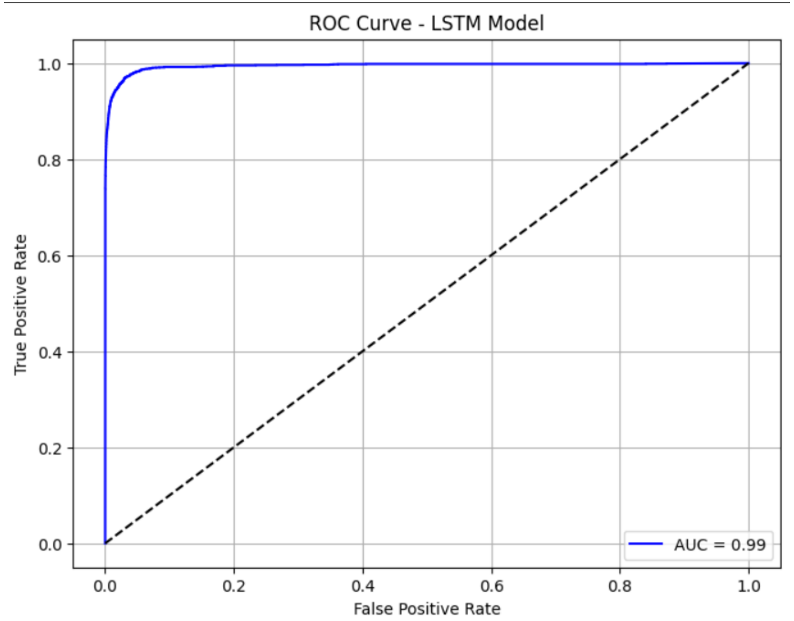

***Fig14.** LSTM Curve*

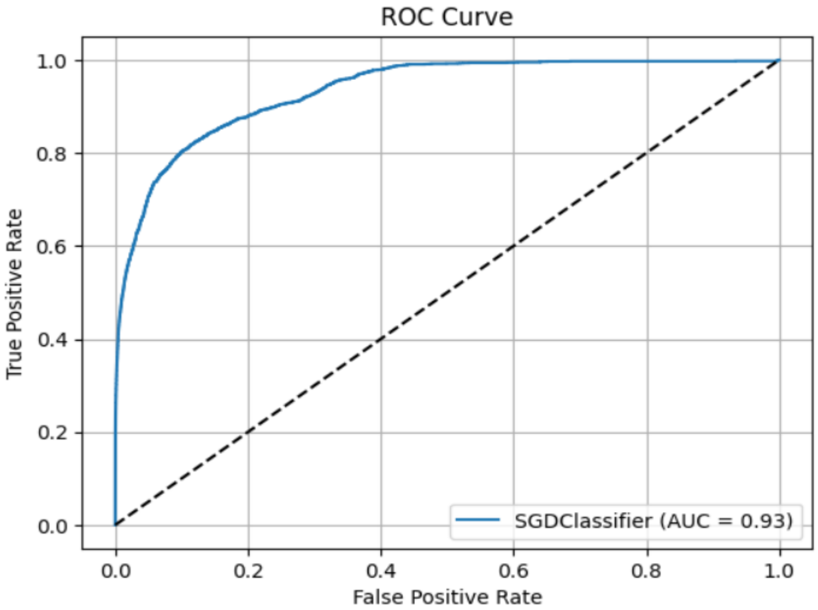

***Fig15.** SGD Curve*

Contd…

The figure below shows the direction of the ROC curves after applying SMOTE.

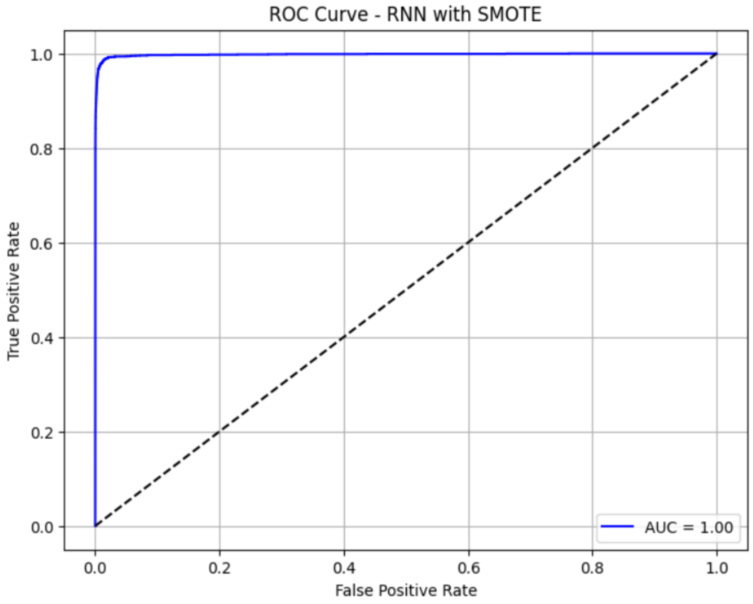

Fig16. *RNN Curve*

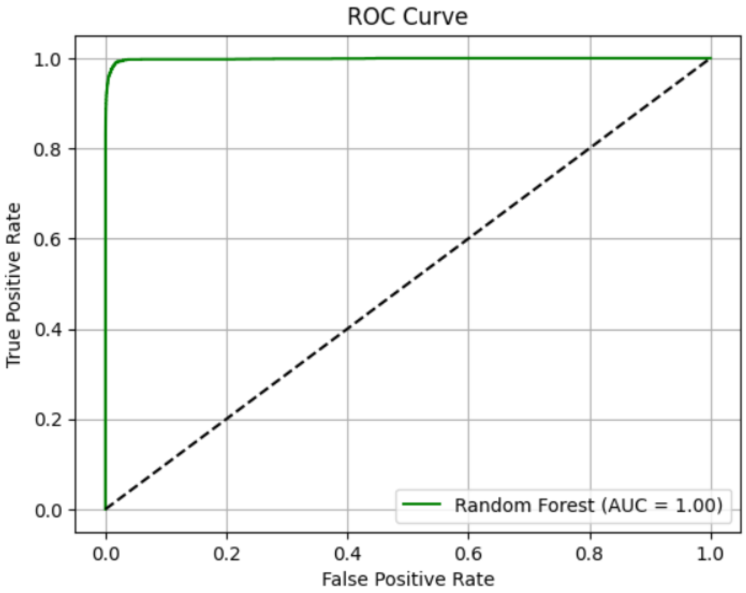

Fig17. *Random Forest Curve*

Contd…

The figure below shows the direction of the ROC curves after applying SMOTE.

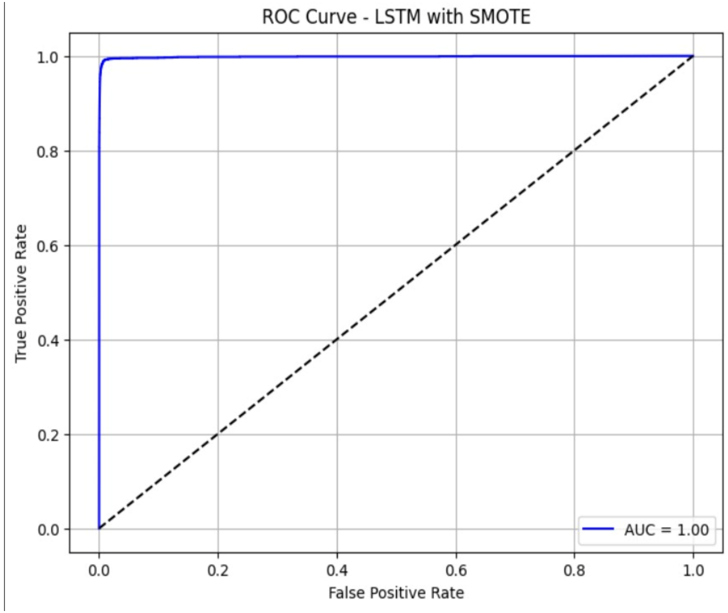

Fig18. *LSTM Curve*

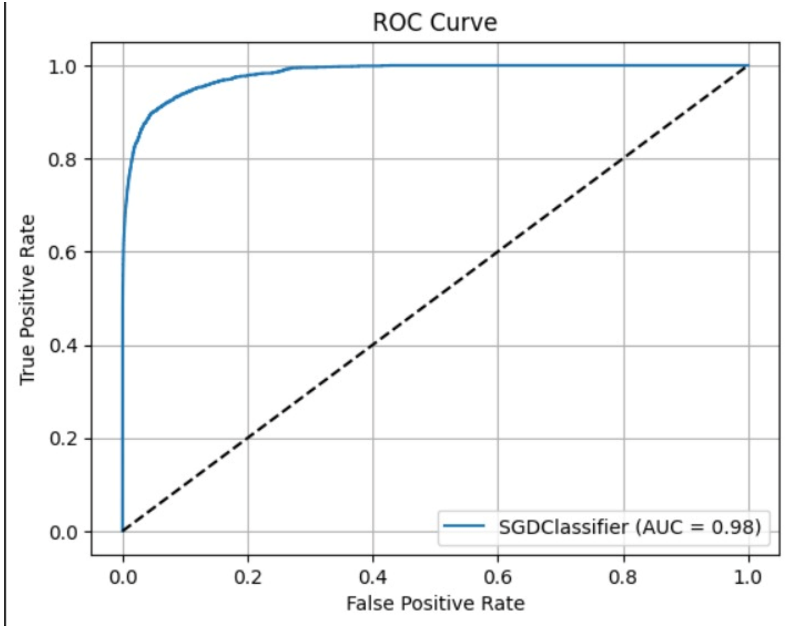

Fig19. *SGD Curve*

Performance Comparison (Before SMOTE)

- Table 11 contains all the result together(in %)

Table 11

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Random Forest	99.96%	99.37%	72.03%	83.52%	99.66%
SGD	99.88%	97.73%	8.83%	16.20%	93.41%
RNN	99.95%	95.95%	63.20%	76.21%	99.54%
LSTM	99.95%	98.03%	63.45%	77.04%	99.45%

Performance Comparison (After SMOTE)

- Table 12 contains all the result together(in %)

Table 12

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Random Forest	98.44%	7.49%	98.81%	13.92%	99.84%
SGD	93.24%	1.70%	91.66%	3.35%	99.86%
RNN	99.07%	11.90%	97.99%	21.23%	99.82%
LSTM	**99.69%**	**28.62%**	**97.25%**	**44.22%**	**99.83%**

Performance comparison with Existing work

- From the existing works and through literature survey we found that our proposed work has gained improvement over the LSTM.[1]

- We also saw the difference in the Accuracy, Precision and Recall score and the results are in the *Table 13*.

Table 13

Models	PERFORMANCE METRICS				
	Accuracy	Precision	Recall	F1-score	AUC
LSTM [1]	0.9850	0.8720	0.8470	0.8590	0.9445
Proposed LSTM Model	**0.9969**	**0.2862**	**0.9725**	**0.4422**	**0.9983**

Conclusion and Future Work

Conclusion

- In this project, we explored various **Machine Learning** and **Deep Learning** models for real time fraud detection.

- Among these, **LSTM** (Long Short-Term Memory) networks proved particularly effective due to their ability to capture long-term dependencies in sequential transaction data.

- Due to their ability to remember long-term dependencies, LSTM reduce **false positives** and increase **precision**, meaning they are better at identifying only **actual fraudulent transactions.**

- This makes LSTM a powerful tool in enhancing the accuracy and reliability of fraud detection systems, ultimately contributing to more secure and trustworthy financial transactions.

Conclusion and Future Work

Key Findings

- Successfully implemented multiple models (Random Forest, RNN, LSTM, SGD).
- SMOTE + LSTM achieved the best performance with an Recall of 97.25%, balancing fraud detection with minimal false positives.
- Demonstrated the importance of data preprocessing (Standard Scaler) and class balancing (SMOTE) for improving model accuracy.

Conclusion and Future Work

Scope for Improvement or Extensions

- Real-time Data Integration Connect with live transaction databases or APIs to enable real-time fraud detection in production.

- Demonstrated the importance of data preprocessing (Standard Scaler) and class balancing (SMOTE) for improving model accuracy.

References

[1] Computers & Security Volume 28, Issue 6, September 2009, Pages 381-394 "**A survey of signature based methods for financial fraud detection**" *Michael Edward Edge, Pedro R. Falcone Sampaio*

[2] Decision Support Systems Volume 50, Issue 3, February 2011, Pages 559-569 "**The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature**" *E.W.T. Ngai a,Yong Hu b,Y.H.Wong a,Yijun Chen b, Xin Sun b*

[3] *A. Mousa*, "**Detecting Financial Fraud Using Data Mining Techniques: A Decade Review from 2004 to 2015**," J. Data Sci., vol. 14, no. 3, pp. 553–570, 2016.

[4] *A. M. Mubalaike and E. Adali*, "**Deep Learning Approach for Intelligent Financial Fraud Detection System**," in UBMK 2018 - 3rd International Conference on Computer Science and Engineering, 2018.

[5] *Yisong Chen , Chuqing Zhao , Yixin Xu , Chuanhao Nie*, "**Year-over-Year Developments in Financial Fraud Detection via Deep Learning:A Systematic Literature Review**", February 4, 2025

Contd…

[6] "**Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review**" *by Abdulalem Ali, Shukor Abd Razak, Siti Hajar Othman, Taiseer Abdalla Elfadil Eisa, Arafat Al-Dhaqm, Maged Nasser, Tusneem Elhassan 1,Hashim Elshafie, and Abdu Saif*

[7] *Andrea Dal Pozzolo, Olivier Caelen, Reid A. Johnson, Gianluca Bontempi.* "**Credit Card Fraud Detection: A Realistic Modelling and a Novel Learning Strategy.**" In *IEEE Transactions on Neural Networks and Learning Systems*, 2015.

[8] *Aravind Kumar Kalusivalingam, Amit Sharma, Neha Patel, Vikram Singh*, "**Enhancing Financial Fraud Detection with Hybrid Deep Learning and Random Forest Algorithms**"

[9] *Ibtissam Benchaji, Samira Douzi, and Bouabid El Ouahidi* **, "Credit Card Fraud Detection Model Based on LSTM Recurrent Neural Networks" ,** Faculty of Sciences IPSS, University Mohammed V, Rabat, Morocco

Contd…

[10] *Md Al-Imran, Eftekhar Hossain Ayon,* **"Transforming banking security: the role of deep learning in fraud detection systems"**

[11] *D. O. Njoku ,V. C. Iwuchukwu, J. E. Jibiri,"***Machine Learning Approach for Fraud Detection System in Financial Institution: A Web Base Application"**

[12] *A. Dal Pozzolo et al.,***"Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy"**

[13] *Oluwabusayo Adijat Bello; Adebola Folorunso2,***"A Comprehensive Framework for Strengthening USA Financial Cybersecurity: Integrating Machine Learning and AI in Fraud Detection Systems"**

[14] https://www.kaggle.com/datasets/jainilcoder/online-payment-fraud-detection (Kaggle Financial Fraud dataset)


Thank you