



**Submitted in fulfillment of the requirements for the degree
Master of Science in Artificial Intelligence**

A Comparative Study of Machine Learning and Adaptive LSTM Models for Credit Card Fraud Detection

Prepared by

Name: Ahmad Mostafa Abutalib

ID No.: 243002097

Supervisor Name: Dr. Amjad Qtaish



Agenda Overview

Introduction & Research Motivation

Problem Definition & Objectives

Literature Review

Dataset & Preprocessing

Models & Proposed Approach

Experimental Setup

Results & Discussion

Conclusion & Recommendations



Introduction & Research Motivation

Introduction & Research Motivation



Introduction

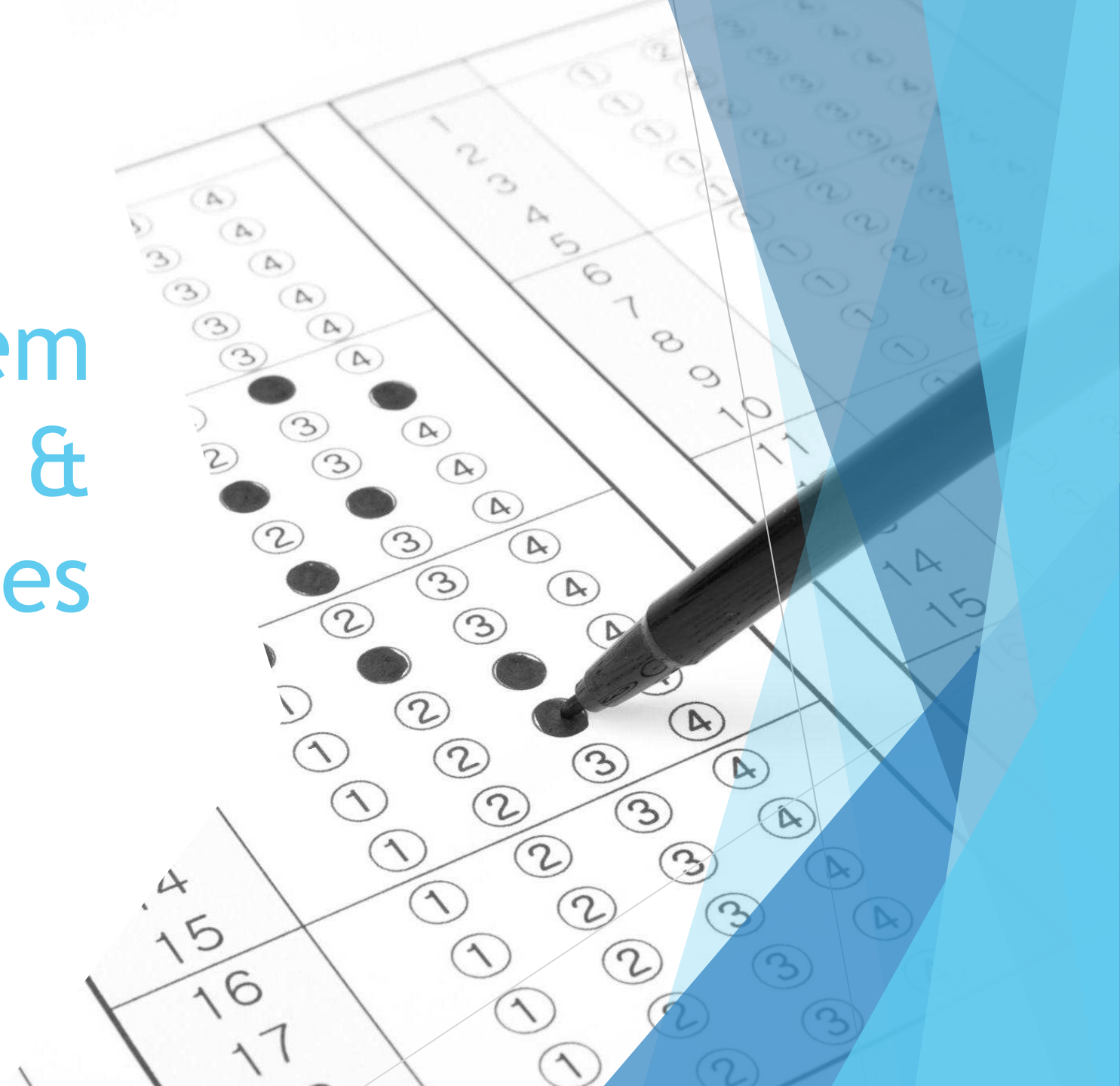
Credit-card fraud continues to rise globally, creating major financial losses and increasing pressure on banks and payment systems to detect fraudulent transactions in real time. Traditional rule-based systems struggle because fraud patterns change quickly and fraudulent transactions represent a very small minority, making detection increasingly difficult.



Research Motivation

This study investigates whether modern Machine Learning and Deep Learning models—especially sequence-based architectures—can improve fraud detection accuracy under realistic imbalance. In particular, the research explores how **adaptive learning-rate scheduling** can enhance LSTM convergence and significantly boost minority-class recall, enabling earlier and more reliable detection of rare fraudulent activities.

Problem Statement & Objectives



- ▶ Credit-card fraud is increasing in volume and complexity. Traditional rule-based systems fail to detect evolving fraud patterns, especially under highly imbalanced datasets where fraudulent cases represent a very small minority. This creates a challenge in maintaining high recall without sacrificing model stability or increasing false alarms.

Problem Statement

Objectives

Main Objective

- Design and evaluate a credit-card fraud detection framework using Machine Learning and Deep Learning models, with a specific focus on improving minority-class recall through adaptive LSTM learning-rate scheduling.

Specific Objectives

- **Evaluate ML and DL performance** (Logistic Regression, XGBoost, RNN, LSTM, Adaptive LSTM) on balanced (50%) and imbalanced (3%) datasets.
- **Analyze the impact of data imbalance** on model accuracy, precision, recall, F1-score, and AUC to understand real-world behavior.
- **Assess the contribution of Adaptive LSTM** and how learning-rate scheduling improves stability and detection of rare fraud cases.
- **Identify the most effective model** for deployment in practical fraud-monitoring environments based on predictive performance and robustness.

Literature Review



Related Works Overview

Recent studies (2020-2024) applied ML and DL models such as CNN, CNN-SVM, BiLSTM-GRU, and LSTM to detect credit-card fraud using Kaggle datasets. While these models achieved reasonable accuracy (83%-96%), their performance varied depending on the model architecture and preprocessing techniques, especially under imbalanced datasets.



Imbalanced datasets were not evaluated systematically → Most studies trained on a single distribution and did not compare balanced vs imbalanced settings.



LSTM models lacked adaptive training methods → No study tested learning-rate scheduling such as ReduceLROnPlateau to improve convergence and minority-class recall.



Limited comparison across ML and sequential DL models under identical conditions → Few papers provided a controlled benchmark between classical ML (LR, XGBoost) and DL models (RNN, LSTM).



Insufficient analysis of recall for rare fraud cases → Most studies prioritized accuracy over minority-class detection.

Gap of Previous Studies

Samples of Related Works

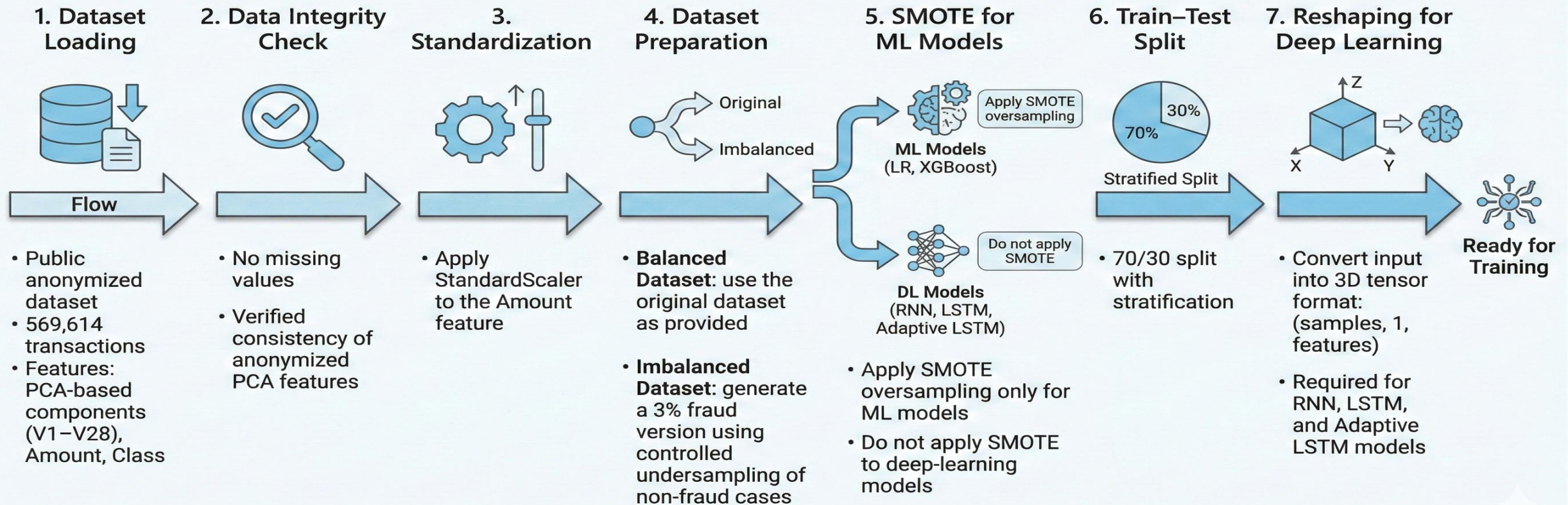
Study Title & Source	Year	Method	Dataset	Result	Limitation
Credit Card Fraud Detection Using CNN without Pooling	2023	CNN	Kaggle 2013	83.72%	Ineffective with PCA-based features; weak recall for fraud cases.
Hybrid CNN-SVM for Fraud Detection	2023	CNN-SVM	Kaggle 2013	91.05%	Limited handling of data imbalance; lacks sequence modeling.
BiLSTM-GRU-MaxPooling	2020	BiLSTM-GRU	Kaggle 2013	91.37%	Sensitive to imbalance; no adaptive learning-rate strategies.
LSTM Approach for Fraud Detection	2024	LSTM	Kaggle 2013	96.3%	High accuracy but lower recall under imbalanced data; fixed learning rate reduces stability.

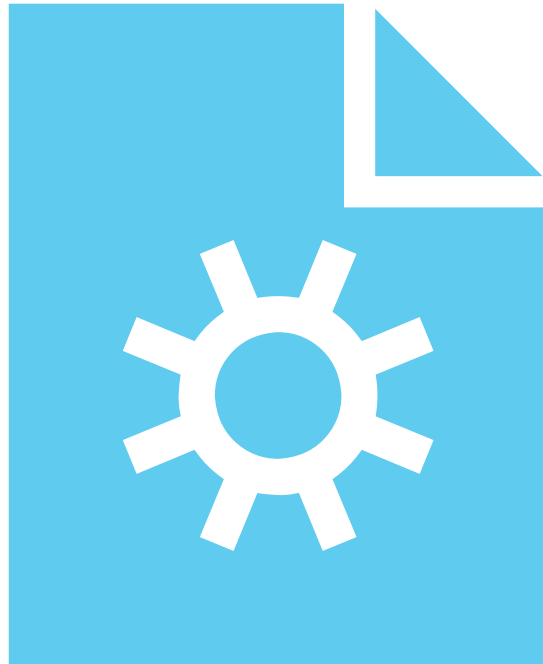
Dataset & Preprocessing



Preprocessing Steps

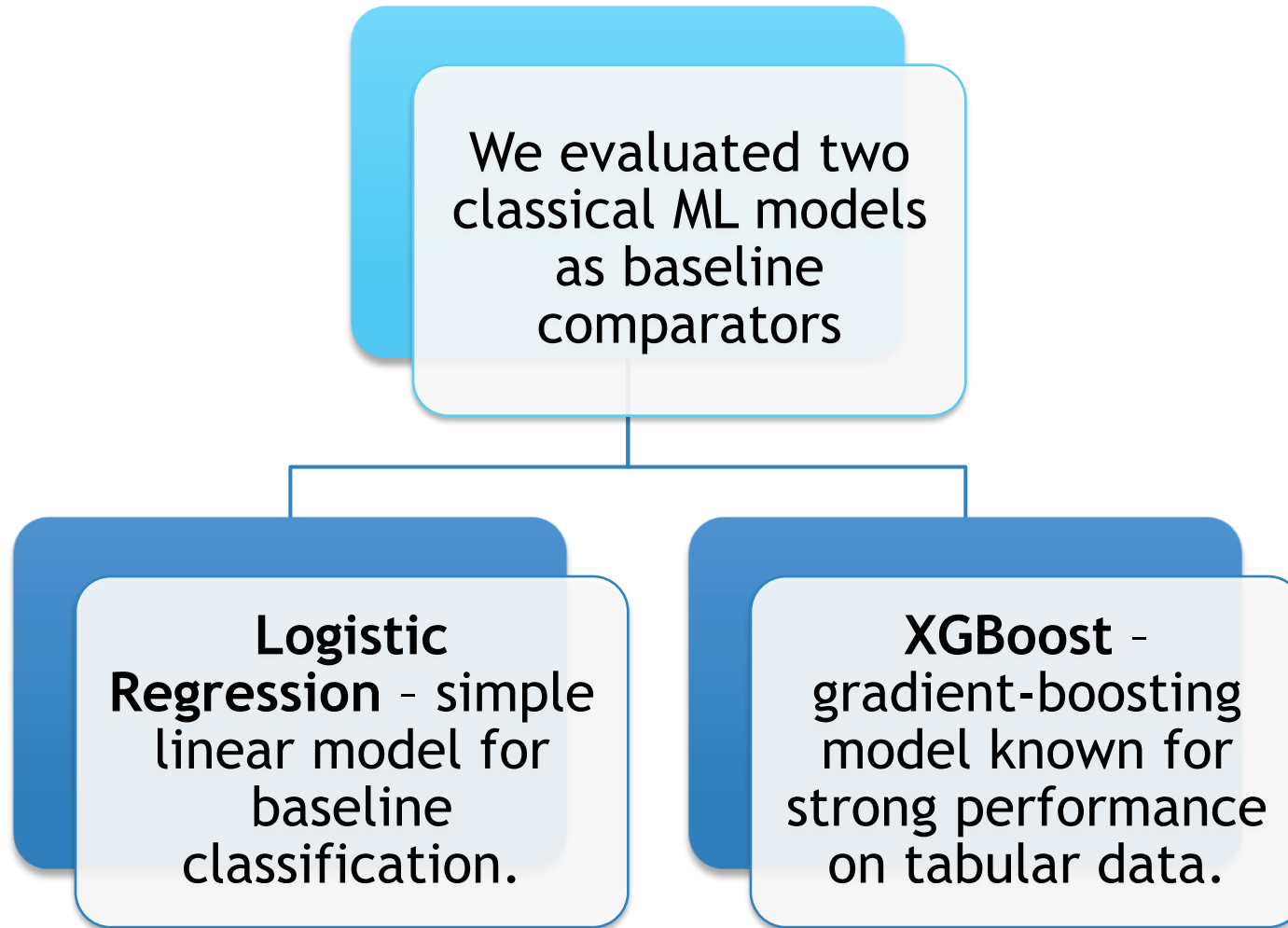
Preprocessing Workflow for Credit-Card Fraud Detection Dataset



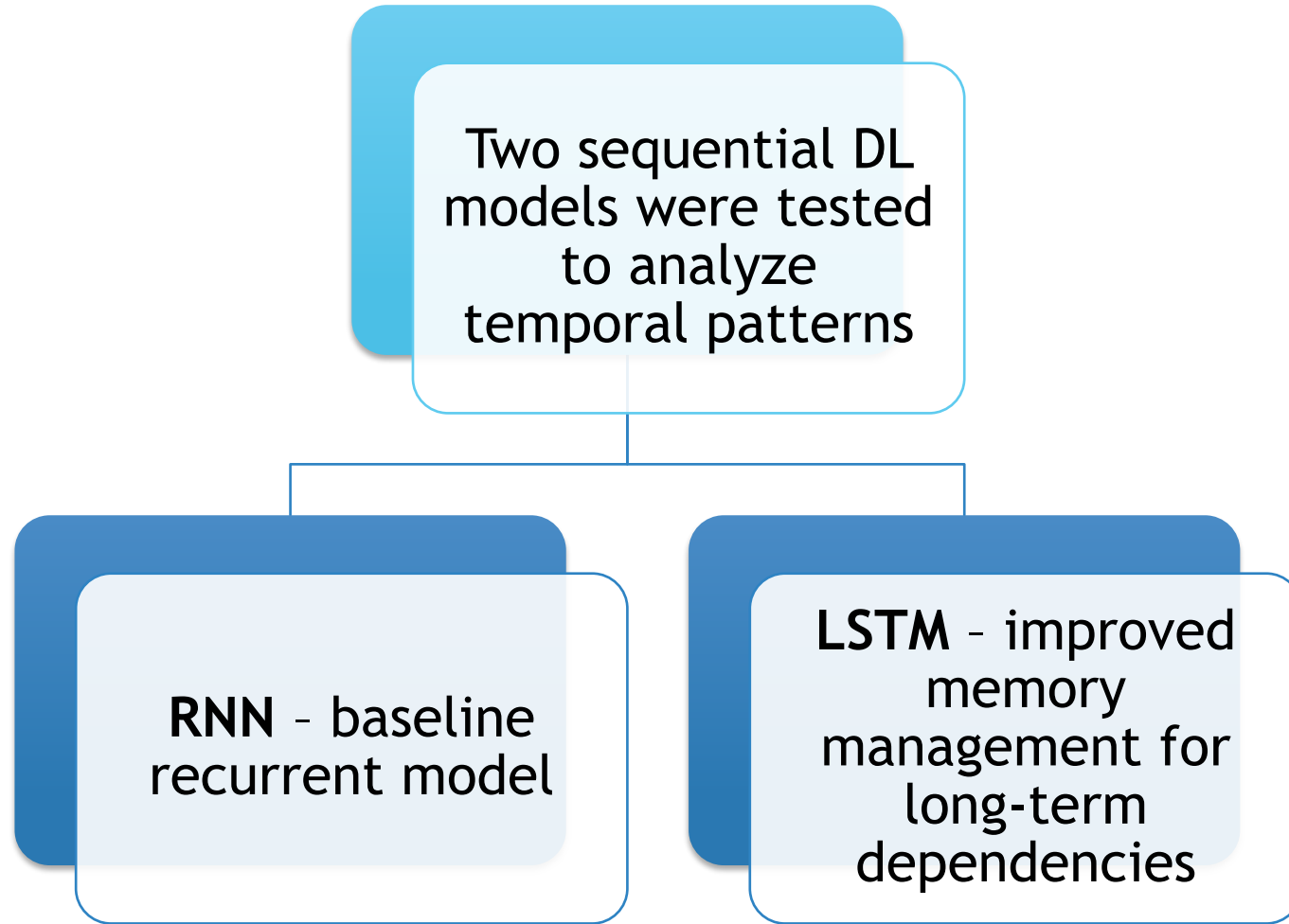


Models & Proposed Approach

Machine Learning Models



Deep Learning Models



Deep Learning Models - Adaptive LSTM

- ▶ **Adaptive LSTM** - LSTM enhanced with **ReduceLROnPlateau** to improve stability and minority-class recall.



Why Adaptive LSTM?

Adaptive LSTM

1

Fraud data is **sparse and unstable**.

2

Standard LSTM uses a **fixed learning rate**, causing unstable training when fraud is rare.

3

Adaptive scheduling **reduces learning rate automatically** when improvement stalls → better convergence and higher recall.

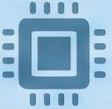
Machine Learning Model Performance on a Balanced Credit Card Fraud Dataset

 Model	 Precision (Fraud)	 Recall (Fraud)	 F1-Score (Fraud)	 Accuracy	 Best Aspect
Logistic Regression	0.98	0.95	0.96	0.9645	Strong balanced performance 
XGBoost	1.00	1.00	1.00	0.9997	Perfect performance – strongest ML model 

Deep Learning Model Performance on a Balanced Credit Card Fraud Dataset

 Model	 Precision (Fraud)	 Recall (Fraud)	 F1-Score (Fraud)	 Accuracy	 Best Aspect
RNN	1.00	1.00	1.00	1.00	Performs perfectly on balanced data 
LSTM	1.00	1.00	1.00	1.00	Strong temporal modeling; excellent on balanced data 

ML Model Performance on Imbalanced Dataset (3% Fraud) after Controlled Undersampling

 Model	 Precision (Fraud)	 Recall (Fraud)	 F1-Score (Fraud)	 Accuracy	 Best Aspect
Logistic Regression	0.98	0.86	0.92	0.9954	Stable baseline 
XGBoost	1.00	0.98	0.99	0.9993	Best rare-fraud  detection among ML models

Deep Learning Model Performance on an Imbalanced Dataset (3% Fraud) after Controlled Undersampling

 Model	 Precision (Fraud)	 Recall (Fraud)	 F1-Score (Fraud)	 Accuracy	 Best Aspect
RNN	0.00	0.00	0.00	0.97	Fails completely on rare fraud cases 
LSTM	0.99	0.45	0.62	0.98	Unstable with imbalanced data 
Adaptive LSTM	0.98	0.85	0.91	1.00	Massive recall improvement (+40%) due to adaptive learning rate  

ML and DL Model Performance: Balanced vs. Imbalanced Datasets

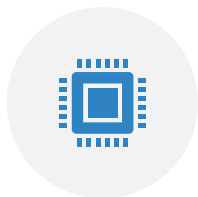
Model 	Balanced F1 	Imbalanced F1 	Balanced Recall 	Imbalanced Recall 	Best Aspect 
Logistic Regression	0.96 	0.92 	0.95 	0.86 	Strong, stable baseline 
XGBoost	1.00 	0.99 	1.00 	0.98 	Best overall ML performance 
RNN	1.00 	0.00 	1.00 	0.00 	Works only on balanced data  
LSTM	1.00 	0.62 	1.00 	0.45 	Sensitive to imbalance 
Adaptive LSTM	1.00 	0.91 	1.00 	0.85 	Most robust DL model under imbalance 

Deep Learning Model Performance: Credit Card Fraud Detection



Comparative Analysis of RNN, LSTM, and Adaptive LSTM

 Model	 Precision (Fraud)	 Recall (Fraud)	 F1-Score (Fraud)	 Accuracy	 Best Aspect
 RNN	0.00	0.00	0.00	0.97	 Fails under imbalance
 LSTM	0.99	0.45	0.62	0.98	 Unstable with imbalanced data
 Adaptive LSTM	0.98	0.85	0.91	1.00	 Best DL model – large recall improvement (+40%) due to adaptive LR scheduling  +40%



The project developed a robust fraud-detection framework using both Machine Learning and Deep Learning models on balanced and highly imbalanced credit-card datasets.



ML models—especially **XGBoost**—demonstrated exceptional performance, achieving perfect accuracy and recall on the balanced dataset and maintaining strong detection capability even after applying controlled undersampling.



Classical DL models (RNN, LSTM) performed extremely well on balanced data but **struggled significantly under imbalance**, highlighting their sensitivity to rare-class distributions.



The proposed **Adaptive LSTM** introduced learning-rate scheduling that substantially improved recall (+40%) compared to standard LSTM, making it the most reliable deep model for rare-fraud detection.



The framework showed that combining resampling strategies with carefully selected ML/DL architectures can dramatically improve fraud detection while reducing false negatives.

Conclusion



Thank you
for your
attention