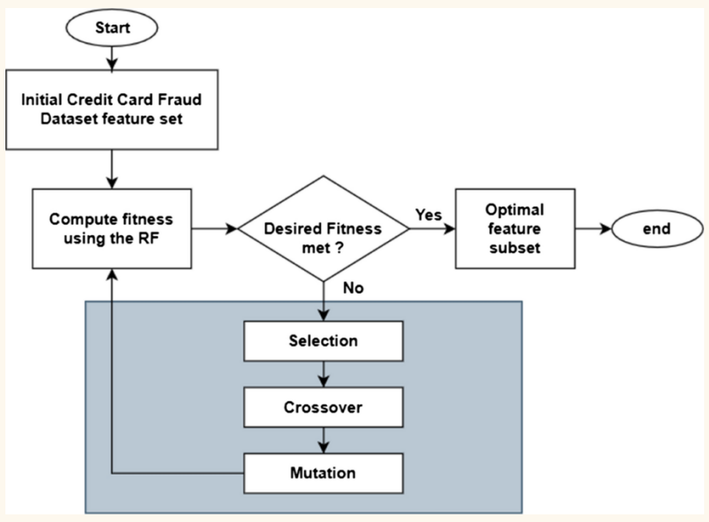
CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING TECHNIQUES.

1 Introduction

Credit card fraud poses a massive financial and cyber security risk, causing banks, merchants, and consumers billions of dollars in losses each year. As online transactions rise, scammers create new sophisticated techniques, rendering conventional rule-based detection useless. Such detection tends to fall short in catching evolving fraud schemes by failing to learn and improve with new fraud patterns, and hence, experiences high false positives (marking legitimate transactions as suspicious) or false negatives (failing to catch fraud).

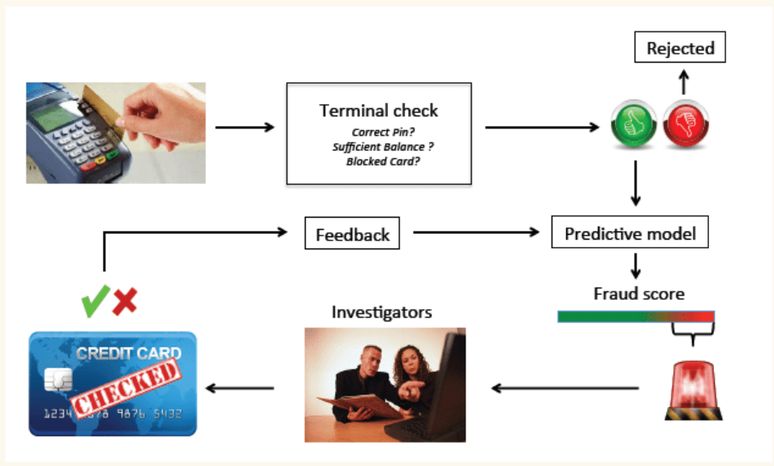


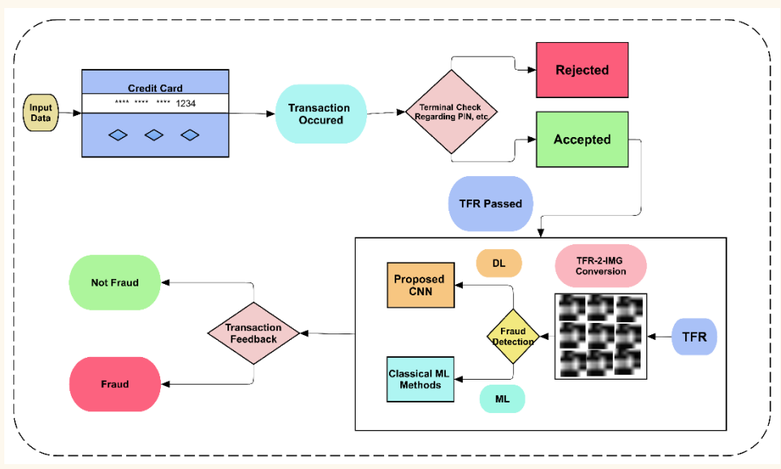
4 Methodology

Supervised Learning Models fraud patterns.
Logistic Regression (LR): Simple baseline model, interpretable but limited performance.
Random Forest (RF): Handles imbalanced data well, uses multiple decision trees.
XG Boost: Reduces overfitting and improves fraud detection accuracy.
Artificial Neural Networks (ANN): Deep learning model that captures complex revised Learning Models
• Isolation Forest: Detects anomalies without labeled data.
• Local Outlier Factor (LOF): Identifies fraud based on transaction density.

2

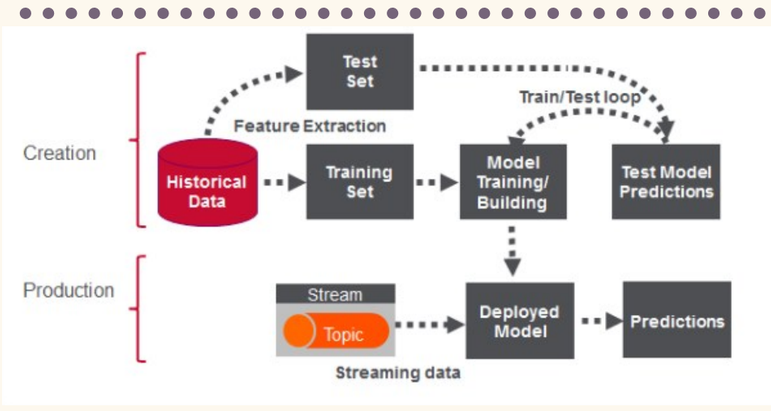
Study	Algorithm(s)	Data Characteristics	Key Techniques	Performance	Main Contribution
Alarfaj et al. (2022)	CNN, LSTM, RF, SVM	284,807 transactions (0.17% fraud)	SMOTE, Feature normalization	F1: 92.7%, AUC: 0.989	Deep learning models outperformed traditional ML methods
Ileberi et al. (2021)	LR, DT, RF with SMOTE & AdaBoost	Severe class imbalance dataset	SMOTE + AdaBoost	AUC: 0.987, Acc: 99.9%	Boosting + oversampling improved detection by 9.6%
Madhurya et al. (2022)	Isolation Forest, LOF, SVM	Anonymous features (V1-V28)	Anomaly detection, PCA	AUPRC: 0.83, Prec: 0.85	Isolation Forest excelled at anomaly detection with minimal FPs
Bin Sulaiman et al. (2022)	Ensemble methods (RF, XGBoost, Stacking)	European dataset, 0.173% fraud	Hybrid undersampling	AUC: 0.991, Prec: 0.93	Ensemble methods reduced FP rates by 23%
Khalid et al. (2024)	Voting ensemble (LR, RF, XGBoost)	30 features dataset	ADASYN, RFECV	AUC: 0.993, F1: 0.93	Voting ensemble reduced FPs by 27%





3 Literature Review

• Deep learning models (CNN, LSTM) perform well but require high computational power.
• Ensemble learning (XG Boost, Stacking, Voting Classifier) improves accuracy and reduces false positives.
• Anomaly detection methods (Isolation Forest, LOF) identify new fraud trends without labeled data.
• Resampling techniques (SMOTE, ADASYN) balance fraud datasets, improving model learning.



5 Early Findings

The combination of ensemble learning models produces encouraging outcomes that help decrease incorrect positive alerts.
The way features are selected has a direct influence on the performance achieved by the model.
Time-based features make it possible to achieve better fraud detection abilities.
When supervised approaches join forces with unsupervised solutions the detection of fresh fraud patterns becomes more attainable.