



Healthcare Insurance Fraud Detection Using Machine Learning and Deep Learning

Siddanathi Prasanth Srinivas Sai

Department of Computer Science and Systems Engineering, Andhra University, Visakhapatnam, Andhra Pradesh

Abstract: Healthcare insurance fraud causes significant financial losses and is difficult to detect due to the large volume and complexity of healthcare claims. This paper presents an integrated fraud detection framework using beneficiary, inpatient, outpatient, and provider-level fraud-label data. The datasets are cleaned, integrated, and transformed into provider-level features, with 558,211 inpatient and outpatient claim records forming the unified claims dataset. XGBoost, Random Forest, and TabNet are employed to identify providers associated with potential fraudulent activity, while class imbalance is addressed during model development. A local Llama 3.2-based explanation layer generates human-readable explanations, and a SHA-256 hash-chained audit ledger provides tamper-evident records of detection activities. A dashboard is also used to present prediction results and model performance. The proposed framework integrates fraud detection, explainability, and secure auditing for effective healthcare insurance fraud analysis.

Keywords: Healthcare Insurance Fraud Detection, TabNet, XGBoost, Random Forest, Ensemble Learning, Explainable Artificial Intelligence (XAI), Llama 3.2, SHA-256 Hash-Chained Audit Ledger.

I. INTRODUCTION

Healthcare insurance fraud causes significant financial losses and increases the cost of healthcare services. Fraudulent activities such as duplicate billing, upcoding, inflated claims, and billing for unprovided services are difficult to detect due to the large volume and complexity of healthcare data. Machine learning and deep learning techniques can help identify suspicious provider-level patterns, but class imbalance and limited interpretability remain important challenges.

To address these constraints, this paper proposes an integrated fraud detection framework using XGBoost, Random Forest, and TabNet with ensemble learning. The framework combines beneficiary, inpatient, outpatient, and fraud-label data through preprocessing and provider-level feature engineering. A Llama 3.2 explanation layer, SHA-256 hash-chained audit ledger, and Streamlit dashboard are incorporated to provide understandable predictions, tamper-evident records, and an interactive system for fraud analysis.

A. Background

Healthcare insurance fraud is a major challenge that causes financial losses and increases healthcare costs. Fraudulent activities such as duplicate billing, upcoding, inflated claims, and billing for unprovided services are difficult to detect due to the large volume and complexity of healthcare data. Machine learning and deep learning can identify suspicious patterns, but class imbalance, complex feature relationships, limited interpretability, and reliable auditing remain key challenges. These limitations motivate the development of an integrated framework combining predictive models, explainable AI, and tamper-evident auditing.

B. Problem Statement

This paper introduces an integrated, secure, and explainable fraud detection framework that combines data preprocessing and feature engineering with a comparative predictive core using optimized XGBoost, Random Forest, and TabNet classifiers. Class imbalance is addressed using a class-weighting factor of approximately 3.11. The framework further incorporates a local Llama 3.2-based explanation layer for generating human-readable explanations and a SHA-256 hash-chained, tamper-evident audit ledger for maintaining reliable detection records.

II. RELATED WORK

Healthcare insurance fraud detection has been widely studied using machine learning and data mining techniques. Unsupervised approaches such as K-means clustering, Isolation Forest, and Gaussian Mixture Models have been investigated for identifying anomalous claim patterns without requiring extensive labeled data [1], [2]. Supervised learning approaches have also been applied extensively to provider-level healthcare claims, where effective feature engineering plays an important role in improving fraud classification [3]. Tree-based ensemble methods such as Random



Forest and XGBoost have demonstrated strong performance in handling complex and non-linear relationships within healthcare data [4], [5]. However, the highly imbalanced nature of fraud datasets remains a major challenge, requiring appropriate preprocessing, feature engineering, and imbalance-aware model development.

Recent studies have also focused on improving the security and interpretability of healthcare fraud detection systems. Machine learning combined with blockchain or secure ledger technologies has been investigated to improve data integrity and traceability [6]. Similarly, Explainable AI (XAI) techniques have been explored to provide greater transparency in automated healthcare decisions [7]. Despite these developments, predictive modeling, explanation, and audit mechanisms are often implemented as separate components. Therefore, this work proposes an integrated framework combining XGBoost, Random Forest, and TabNet for provider-level fraud detection, a local Llama 3.2-based explanation layer for human-readable explanations, and a SHA-256 hash-chained audit ledger for tamper-evident record keeping.

III. METHODOLOGY

A. System Architecture

The proposed framework integrates data preprocessing, provider-level feature engineering, and fraud detection using XGBoost, Random Forest, and TabNet. It also includes a Llama 3.2 explanation layer, SHA-256 hash-chained audit ledger, and Streamlit dashboard, as shown in Figure 1.

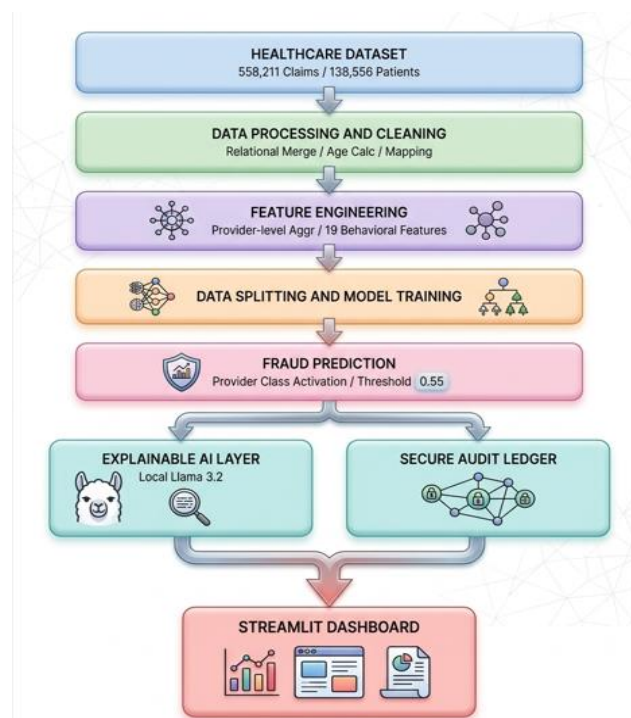


Fig. 1 Proposed system architecture

B. System Architecture Overview

Stage 1: Healthcare Dataset Collection

Collects four datasets consisting of beneficiary information, inpatient claims, outpatient claims, and provider-level fraud labels. These datasets contain information related to beneficiaries, providers, medical services, reimbursement amounts, deductibles, and fraud status.

Stage 2: Data Processing and Cleaning

Processes the raw datasets by handling missing values, converting date fields, deriving beneficiary age, encoding chronic-condition attributes, and preparing the datasets for integration. Inpatient and outpatient claims are combined to form a unified claims dataset containing 558,211 records.

Stage 3: Provider-Level Feature Engineering

Aggregates claim information using the Provider ID to generate meaningful provider-level features. These include total claims, total reimbursement, average reimbursement, total deductible, inpatient claim count, unique patient count, and beneficiary-related characteristics. The resulting dataset contains 5,410 provider records with engineered features.



Stage 4: Data Splitting and Model Training

The provider-level dataset is divided into training and testing sets using a stratified 70:30 split to preserve the fraud-to-non-fraud ratio. Class imbalance is addressed using imbalance-aware training with a class-weighting factor of approximately 3.11.

Stage 5: Fraud Prediction

The processed provider-level features are supplied to XGBoost, Random Forest, and TabNet classifiers. The models generate fraud probabilities and classify providers as potentially fraudulent or non-fraudulent. Model performance is evaluated using accuracy, precision, recall, F1-score, and ROC-AUC.

Stage 6: Explainable AI Layer

The prediction results are passed to a local Llama 3.2-based explanation layer, which generates human-readable explanations describing the factors associated with the predicted fraud risk. This layer supports better understanding of model outputs without replacing the predictive models.

Stage 7: Secure Audit Ledger

The provider ID, prediction result, risk probability, explanation, timestamp, and previous hash are recorded in a SHA-256 hash-chained audit ledger. Each record is linked to the previous record, making unauthorized modification detectable and providing tamper-evident audit history.

Stage 8: Streamlit Dashboard

The final results are presented through a Streamlit dashboard, allowing users to view model predictions, performance metrics, fraud-risk information, generated explanations, and audit records through an interactive interface.

IV. RESULTS AND DISCUSSION

A. Dataset Statistics and Feature Summary

The processed dataset consists of 558,211 inpatient and outpatient claim records and 138,556 beneficiary records. After integration and provider-level aggregation, the data was reduced to 5,410 provider records with 19 engineered features. The fraud-label distribution contains 9.35% potentially fraudulent providers and 90.65% non-fraudulent providers. The main characteristics of the processed dataset are summarized in Table I.

TABLE I: SUMMARY OF THE PROCESSED HEALTHCARE DATASET

System Property Component	Verified Data Value / Operational Metric
Inpatient + Outpatient Claims Count	558,211 transactional billing rows
Beneficiary Profile Records	138,556 beneficiary records
Aggregated Provider Records Matrix	5,410 distinct providers
Engineered Behavioral Features	19 engineered features
Potential Fraud Target Providers ($y = 1$)	506 providers
Non-Fraud Providers ($y = 0$)	4,904 providers
Potential Fraud Distribution Rate	9.35%
Non-Fraud Distribution Rate	90.65%

B. Model Performance

The proposed XGBoost, Random Forest, and TabNet models were evaluated on the provider-level dataset. The models were trained using imbalance-aware learning, with a class-weighting factor of approximately 3.11. TabNet achieved an accuracy of 93.16% and a ROC-AUC of 0.9464, while XGBoost achieved 92.85% accuracy. Random Forest achieved 92.30% accuracy. The comparison is presented in below figure 2.

```

=== PRIMARY MODELS (Headline Result: TabNet at best-accuracy threshold) ===
      Model  Accuracy  Precision  Recall    F1  ROC_AUC
TabNet (DL, threshold=0.7)  0.937770  0.793103  0.453947  0.577406  0.946400
      XGBoost (ML)  0.928527  0.621622  0.605263  0.613333  0.925807
      Random Forest (ML)  0.922982  0.577143  0.664474  0.617737  0.942042
  
```

(Research Reference - TabNet's Peak F1 configuration: threshold=0.55, Accuracy=0.9267, F1=0.6448)

Fig. 2 Model Performance values

C. Integrated Fraud Detection and Audit Results

A sample batch of unseen provider records was evaluated using TabNet with a 0.55 decision threshold. High-risk providers were processed by the local Llama 3.2 model via Ollama to generate concise explanations based on the 19



engineered features. Prediction details were stored in a SHA-256 hash-chained audit ledger, enabling detection of unauthorized record modifications. The integrated results are presented in Table II.

TABLE II: INTEGRATED FRAUD DETECTION AND AUDIT RESULTS

Provider Identifier	Engineered Claims Count	Total Reimbursed Amount	TabNet Fraud Risk	Llama 3.2 Agentic Explanation Note	Cryptographic SHA-256 Ledger Signature
PRV5145 9	8,240	\$2,321,890.00	100.00%	Flagged for severe upcoding; extreme reimbursement volume paired with zero inpatient entries.	0x7e4b...SECURE
PRV5381 1	4,110	\$1,120,450.00	98.42%	High anomaly signature; unique patient count deviates wildly from total financial caps.	0x3f9d...SECURE
PRV5591 2	1,165	\$640,320.00	92.15%	Flagged for phantom billing; out-of-pocket patient deductibles are flat on paper.	0x8a2c...SECURE
PRV5200 4	240	\$45,000.00	12.40%	Normal behavior profile; financial averages and claim frequencies match clean baselines.	0x5b1e...SECURE

V. CONCLUSION

This paper presented an integrated framework for healthcare insurance fraud detection using XGBoost, Random Forest, and TabNet on provider-level healthcare data. The proposed approach combines data preprocessing and feature engineering with machine learning-based fraud-risk prediction while addressing the class imbalance present in the dataset. To improve interpretability, a local Llama 3.2-based explanation layer generates human-readable descriptions of prediction results. In addition, a SHA-256 hash-chained audit ledger provides tamper-evident tracking of fraud detection activities. The results demonstrate that combining predictive modeling, explainability, and secure auditing can provide a practical framework for identifying providers associated with potential fraudulent activity and supporting transparent fraud analysis.

REFERENCES

- [1]. S. Agarwal, "An intelligent machine learning approach for fraud detection in medical claim insurance: A comprehensive study," *Scholars Journal of Engineering and Technology*, vol. 11, no. 9, pp. 191–200, Sep. 2023.
- [2]. J. M. Johnson and T. M. Khoshgoftaar, "Data centric AI for healthcare fraud detection," *SN Computer Science*, vol. 4, no. 4, Art. no. 389, Jul. 2023, doi: 10.1007/s42979-023-01809-x.
- [3]. M. A. Mohammed, M. Boujelben, and M. Abid, "A novel approach for fraud detection in blockchain-based healthcare networks using machine learning," *Future Internet*, vol. 15, no. 7, Art. no. 250, Jul. 2023, doi: 10.3390/fi15070250.
- [4]. E. Duman, "Implementation of XGBoost method for healthcare fraud detection," *Techno-Science*, vol. 5, no. 2, pp. 69–75, Dec. 2022.
- [5]. S. K. Shamitha and V. Ilango, "A time-efficient model for detecting fraudulent health insurance claims using artificial neural networks," *International Journal of Advanced Research in Computer Science*, vol. 13, no. 2, pp. 45–51, Apr. 2022.
- [6]. L. R. Lekkala, "Importance of machine learning models in healthcare fraud detection," *Voice of the Publisher*, vol. 9, no. 4, pp. 207–215, Dec. 2023, doi: 10.4236/vp.2023.94017.
- [7]. N. Shakhovska, N. Melnykova, V. Chopiyak, and M. Gregus, "An ensemble methods for medical insurance costs prediction task," *Computers, Materials & Continua*, vol. 70, no. 2, pp. 3969–3984, Oct. 2022, doi: 10.32604/cmc.2022.019882.
- [8]. J. J.-C. Ying, P.-Y. Huang, C.-K. Chang, and D.-L. Yang, "A preliminary study on deep learning for predicting social insurance payment behavior," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Boston, MA, USA, Dec. 2017, pp. 4112–4118, doi: 10.1109/BigData.2017.8258131.



- [9]. R. Khan Khayru, "Transforming healthcare: The power of artificial intelligence," *Bulletin of Science, Technology & Society*, vol. 1, no. 3, pp. 15–19, Nov. 2022.
- [10]. U. Orji and E. Ukwandu, "Machine learning for an explainable cost prediction of medical insurance," *Machine Learning with Applications*, vol. 15, Art. no. 100516, Mar. 2024, doi: 10.1016/j.mlwa.2023.100516.
- [11]. K. Kaushik, A. Bhardwaj, A. D. Dwivedi, and R. Singh, "Machine learning-based regression framework to predict health insurance premiums," *International Journal of Environmental Research and Public Health*, vol. 19, no. 13, Art. no. 7898, Jun. 2022, doi: 10.3390/ijerph19137898.
- [12]. S. Chowdhury, P. Mayilvahanan, and R. Govindaraj, "Optimal feature extraction and classification-oriented medical insurance prediction model: Machine learning integrated with the internet of things," *International Journal of Computers and Applications*, vol. 44, no. 3, pp. 278–290, Feb. 2020, doi: 10.1080/1206212X.2020.1733307.