Monica G, 2026, 14:6
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Fraudulent Health Insurance Claims Detection Using Machine Learning

Monica G[1] Prasanna S[2]
Department Of Applied Computing & Emerging Technologies
Vels Institute Of Science Technology And Advanced Studies,
Monicalali21@Gmail.Com
Department Of Computer Applications – Pg
Vels Institute Of Science Technology And Advanced Studies,
Prasanna.Scs@Vistas.Ac.In

Abstract Fraudulent health insurance claims increase healthcare costs, reduce the quality of service, and undermine public trust in insurance systems. Machine learning provides a data-driven way to identify suspicious patterns in large volumes of claim records more effectively than manual review and static rule-based approaches. This paper presents a comprehensive framework for detecting fraudulent health insurance claims using supervised machine learning on structured claims data. The proposed system includes data ingestion, preprocessing, feature engineering, handling of class imbalance, model training, evaluation, and deployment considerations. Multiple algorithms, including logistic regression, decision trees, random forests, gradient boosting, and support vector machines, are compared using fraud-oriented metrics such as precision, recall, F1-score, and AUC-ROC. The framework is designed to be modular and scalable, and it explicitly addresses interpretability, integration with human investigators, and model governance. The work illustrates how insurers can use machine learning to prioritize high-risk claims, reduce investigation costs, and improve fraud management in healthcare.

Keyword— Health insurance, fraud detection, machine learning, imbalanced data, healthcare analytics.

I. INTRODUCTION

Health insurance enables risk pooling of medical expenses and provides financial protection to individuals and families against unexpected healthcare costs. However, as claim volumes and healthcare expenditures rise, fraudulent and abusive claims have become a major challenge for insurers, governments, and regulators. Fraudulent behaviour wastes resources that should be allocated to genuine patients, and it threatens the sustainability of healthcare systems.

Fraud in health insurance can occur at various stages of the claim lifecycle, including patient enrollment, pre-authorisation, billing and post-payment review. Typical schemes involve billing for services not rendered, exaggerating the severity of diagnoses, upcoding procedures to more expensive categories and collusion between providers and policyholders. Because genuine claims heavily outnumber fraudulent ones, exhaustive manual review is infeasible, and simple rule-based systems often fail to keep pace with evolving fraud patterns.

Machine learning (ML) offers an alternative that can learn complex relationships from historical claim data, adapt to new patterns, and support automated claim screening. By integrating ML into claim processing workflows, insurers can focus investigation resources on the most suspicious cases while allowing low-risk claims to pass through with minimal delay.

II. OBJECTIVES

- Design a modular pipeline that covers data ingestion, preprocessing, feature engineering and model training.
- Address class imbalance and evaluate multiple supervised learning algorithms using fraud-relevant performance metrics.

© 2026 Monica G, This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited.

- Discuss practical deployment aspects, including interpretability, human–machine collaboration and model monitoring.

III. WORKINGS

1. Health Insurance Fraud And Detection Approaches

Health insurance fraud spans a range of behaviours, from opportunistic exaggerations of legitimate claims to organized schemes that systematically exploit reimbursement rules.

Traditional detection approaches include

- Manual Claims Review
- Audit Sampling
- Whistleblower Reports
- Rule-based Systems.

Rule-based systems encode expert knowledge as if-then rules or thresholds (for example, flagging claims above a certain amount or identifying certain combinations of diagnosis and procedure codes). These systems are interpretable but tend to be rigid and require frequent manual updates.

- Data-driven techniques arose to address these limitations.
- Classical statistical models such as logistic regression provide interpretable relationships between explanatory variables and fraud likelihood.
- More recent machine learning methods can capture nonlinear interactions and high-dimensional feature spaces that are typical of detailed claims data.

2. Supervised Learning For Fraud Detection

Supervised learning has been widely applied when labelled datasets of fraudulent and genuine claims are available.

The commonly used algorithms include:

- Logistic Regression: a baseline linear model, often used as a benchmark due to its simplicity and interpretability.
- Decision Trees: Non-linear models that produce human-readable rules.
- Random Forests: Ensembles of decision trees that reduce variance and improve generalization.
- Gradient Boosting (e.g., XGBoost, LightGBM): Boosted ensembles that typically perform strongly on structured tabular data.
- Support Vector Machines (SVM): SVM models that use kernel functions to separate classes in high-dimensional feature spaces

3. Unsupervised And Semi-Supervised Methods

In operational environments, labels for fraudulent claims are often incomplete or delayed.

Unsupervised anomaly detection methods, such as clustering, density estimation, and isolation-based models, attempt to identify claims that deviate markedly from normal patterns.

Semi-supervised approaches leverage both labelled and unlabelled claims, allowing models to learn the structure of normal behaviour while being guided by limited fraud labels.

Autoencoders and other deep learning models have also been explored for representation learning and anomaly detection. However, these models can be more difficult to interpret, which is a concern in regulated healthcare contexts.

4. Graph-Based Approaches

Fraud often involves interactions between multiple entities, including patients, providers, pharmacies, intermediaries, and even devices.

Graph-based approaches represent these entities as nodes and their relationships (such as shared addresses, bank accounts, or repeated collaborations) as edges.

Graph algorithms and graph neural networks can then detect suspicious communities, abnormal subgraphs, or unusual connectivity patterns that might indicate collusive fraud.

5. Research Gaps

Several gaps remain in the literature:

- **Operational integration:** Many research works stop at algorithmic performance without discussing

how models can be integrated into existing claim processing systems.

- **Interpretability:** Complex models may be accurate but difficult to explain to auditors, regulators, or affected parties.
- **Model governance**: long-term monitoring, retraining schedules, and concept drift are often not addressed in detail.

IV. PROPOSED METHODOLOGY

1. Problem Formulation

Let each claim be represented by a feature vector Eq.(1)

$$Xi \in Rdxi \in Rd,$$

Where,

dd is the number of features, and let

$$yi \in \{0,1\} yi \in \{0,1\}$$

denote the label (0 = genuine, 1 = fraudulent).

Given a dataset of nn labelled claims

$$\{(xi,yi)\}i=1n\{(xi,yi)\}i=1n,$$

the task is to learn a classification function f(x)f(x) that estimates the probability that a new claim is fraudulent.

The key challenge is that

$$\sum iyi \ll n \sum iyi \ll n;$$

that is, the fraud class is rare.

3. System Architecture

The proposed system is composed of the following modules:

1. Data Ingestion

- Extracts claims from core insurance systems and data warehouses.
- Ensures standardized formats, consistent identifiers, and basic validations.

2. Pre Processing and Feature Engineering:

- Cleans the data, handles missing values, encodes categorical features, and constructs derived attributes.

3.Model Training and Validation:

- Trains multiple supervised algorithms with appropriate hyperparameter tuning and imbalance handling.

3.Scoring and Integration:

- Produces fraud risk scores for new claims and integrates with claim processing workflows to prioritize investigations.

4. Monitoring and Feedback:

- Tracks model performance over time and incorporates investigator feedback into future retraining.

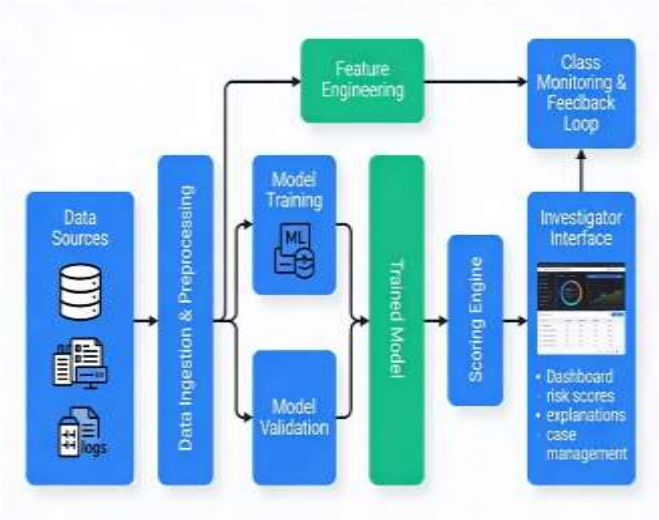


Figure 1**:** Overall architecture diagram showing data sources, preprocessing, model training, scoring engine, and investigator interface.

5. Dataset Structure

A typical health insurance claims dataset includes:

- **Policyholder Attributes:** Age, gender, policy type, policy tenure, number of dependents, historical claim count, and any prior fraud flag.
- **Provider Attributes:** Provider ID, provider type (hospital, clinic, practitioner), specialty, location, historical volume, and historical rejection rate.
- **Claim Attributes:** Claim ID, admission and discharge dates, length of stay, diagnosis codes, procedure codes, claimed amount, approved amount, co-payment amount, and payment mode.
- **Temporal and Behavioural Attributes:** Time since policy inception, time since previous claim, claim submission delay, frequency of claims in recent windows.

Label: A binary fraud indicator derived from internal investigations or audits.

Attribute Group	Sample Features	Data Type	Example Values
Policyholder	Age, Gender, Policy Type, Policy Tenure, Dependents	Numeric, Categorical	45 yrs, Male, Family Floater
Provider	Provider ID, Type, Specialty, Location, Rejection Rate	Categorical, Numeric	Hosp123, Hospital, Cardiology
Claim Details	Claim ID, LOS*, Diagnosis Codes, Procedure Codes	Numeric, Categorical	ICD10-E11, CPT99214
Financial	Claim Amount, Approved Amount, Co-pay, Payment Mode	Numeric, Categorical	$2500, $2200, Cashless
Temporal	Admission Date, Submission Delay, Weekend Admission	Date, Numeric, Binary	7 days delay, Weekend: Yes
Behavioral	Claim Freq. (3/6/12 mo), Hist. Fraud Flag	Numeric, Binary	5 claims/3mo, Prior Fraud: No
Target	Fraud Label	Binary	0=Legitimate, 1=Fraudulent

6. Data Preprocessing

The preprocessing pipeline consists of:

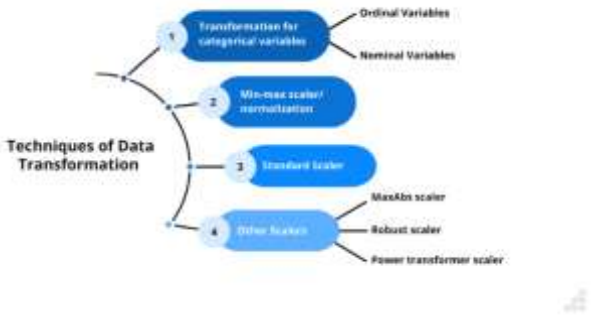
1. **Cleaning and Validation:**
- Removing duplicate claims and records with inconsistent or impossible values (e.g., negative amounts, discharge before admission).
- Basic rule checks to ensure that mandatory fields (IDs, dates, amounts) are present.

2. **Handling Missing Values:**
- For numerical attributes, using median or domain-specific imputation.
- For categorical attributes, introducing an "Unknown" category to represent missing values.

3. **Categorical Encoding:**
- One-hot encoding for low-cardinality variables (e.g., gender, provider type).
- Target or frequency encoding for high-cardinality variables (e.g., provider ID, certain medical codes), possibly with grouping of rare categories.

4. **Feature Scaling**:
- Standardising numerical features such as amounts and lengths of stay to have zero mean and unit variance, using statistics from the training set only.

5. **Outlier Treatment:**
- Identifying extreme values in amounts or length of stay and either capping (winsorization) or maintaining them as potential anomalies, depending on design choices.



7. Feature Engineering

Feature engineering is crucial for capturing fraud patterns that may not be apparent from raw attributes. Important groups of engineered features include:

1.Aggregated Policyholder Features:

Average claim amount per policyholder, total amount claimed in the last 6 or 12 months, and number of claims per time window.

Ratio of current claim amount to the policyholder's typical claim amount.

2.Aggregated Provider Features:

- Average claim amount per provider, distribution of procedures and diagnoses, and composition of patient demographics.
- Deviation of a provider from peers in the same specialty (for example, significantly higher average claim amounts or unusual procedure mixes).

3. Temporal and Behavioral Features:

- Time since policy inception or renewal, to detect concentration of claims near high-risk periods.
- Inter-claim intervals for the same policyholder or provider, capturing bursts of activity.
- Indicators for weekend or holiday admissions.

4.Medical Code Groupings:

- Grouping diagnosis and procedure codes into clinical categories, as fine-grained codes may be too sparse.
- Counting rare combinations of diagnoses and procedures that may signal inflated or fabricated claims.
- Interaction Features:
- Interactions between policyholder demographics and certain procedures, such as unlikely procedures for certain age groups.
- Ratios between billed and approved amounts.

Well-designed features can substantially improve model performance even when relatively simple algorithms are used.

8. Addressing Class Imbalance

The fraud class is typically very small compared to the genuine class. Training a classifier directly on such data often leads to models that predict the majority class for almost all claims. To mitigate this, the framework uses:

1.Resampling Methods:

- Random undersampling of genuine claims to reduce their dominance.
- Oversampling of fraudulent claims, either by replicating minority instances or by synthetic methods such as SMOTE.

2.Cost-Sensitive Learning:

- Assigning higher misclassification cost or class weight to fraudulent claims so that the learning algorithm penalizes missed fraud more heavily.

3. Threshold Tuning:

- Adjusting the decision threshold on predicted probabilities to achieve a desired trade-off between recall and precision for the fraud class.

9. Machine Learning Models

The framework compares several supervised algorithms:

- **Logistic Regression:**
  Provides a transparent baseline. Coefficients indicate the direction and approximate strength of feature effects.
  Decision Tree:
- Captures nonlinear relationships and generates rule paths that can be inspected by auditors.
  Random Forest:
- An ensemble of decision trees that averages predictions to reduce variance. It is robust and handles mixed-type features well.
  Gradient Boosting:
- Builds trees sequentially, each one improving upon the errors of the previous ensemble. Common implementations are highly effective on tabular data.
- **Support Vector Machine:**
  Uses kernel functions (e.g., radial basis function) to separate classes in transformed feature spaces, which can capture complex patterns.
  Hyperparameters (such as tree depth, number of trees, learning rate, and regularization terms) are tuned using stratified k-fold cross-validation on the training set.

10. Evaluation Metrics

- Given the strong class imbalance, accuracy is not an adequate metric. This framework emphasizes:
- Precision (fraud class): proportion of claims predicted as fraudulent that are actually fraudulent.
- Recall (fraud class): proportion of actual fraudulent claims that are correctly identified.
- F1-score: Harmonic mean of precision and recall, balancing both aspects.

- AUC-ROC: area under the Receiver Operating Characteristic curve, measuring the ranking capability of the model across thresholds.
- Confusion Matrix: distribution of true positives, false positives, true negatives, and false negatives, useful for operational impact analysis.
- A Precision–Recall curve may also be examined, as it is more informative than ROC when the positive class is rare.

V. EXPERIMENTAL SETUP AND TYPICAL RESULTS

1. Experimental Design

The dataset is split into training, validation, and test sets using stratified sampling to preserve fraud prevalence. Preprocessing, feature engineering, and resampling are performed on the training data within a cross-validation procedure.The validation set guides hyperparameter tuning and model selection. The final chosen model is evaluated on the test set to estimate generalization performance. Implementation can be done in Python using libraries such as pandas for data manipulation, scikit-learn for models and metrics, gradient boosting libraries (e.g., XGBoost or LightGBM), and specialized packages for resampling.

2. Baseline Models

To quantify the benefit of the advanced models, simple baselines are considered:

- A rule-based model using a small set of manually designed thresholds (for example, flagging claims above certain amounts or unusual combinations of codes).
- Logistic regression with a limited feature set (e.g., using only raw claim amounts and lengths of stay).
- These baselines typically achieve moderate accuracy but low recall for the fraud class, highlighting the need for more sophisticated approaches.

3. Illustrative Quantitative Results

The following is a representative example of performance patterns (to be replaced with actual experimental results in your implementation).

Model	Precision	Recall	F1-score	AUC-ROC
Rule-based	0.40	0.25	0.31	0.70
Logistic Reg.	0.60	0.50	0.55	0.84
Decision Tree	0.64	0.58	0.61	0.86
Random Forest	0.75	0.70	0.72	0.92
Gradient Boosting	0.78	0.73	0.75	0.94
SVM	0.71	0.65	0.68	0.90

Table II Example performance of models on the test set (fraud class).

In many real datasets, tree-based ensembles such as random forests and gradient boosting achieve higher F1-scores and AUC-ROC than simpler models, indicating better discrimination and more effective fraud detection.

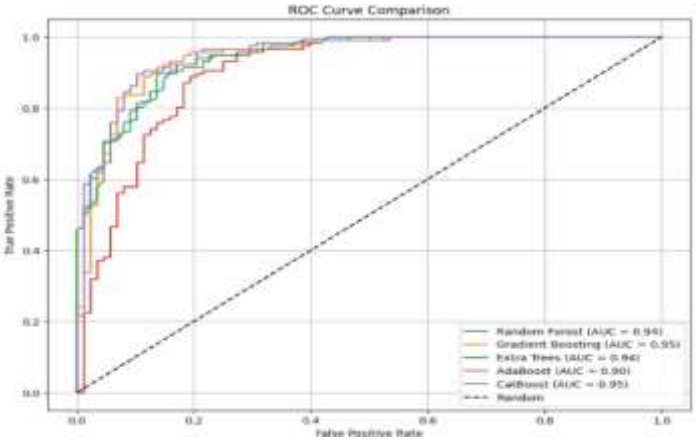


Figure 3: ROC Curves Comparison for Fraud Detection Models

4. Threshold And Confusion Matrix Analysis

By examining the confusion matrix for the best-performing model at different decision thresholds, insurers can choose operating points that match their investigation capacity. For example:

A lower threshold increases recall (detects more fraud) but triggers more false positives, increasing investigator workload.

A higher threshold reduces false positives but may miss more fraudulent claims. The choice of threshold can be guided by financial cost–benefit analyses, balancing the cost of investigations against potential fraud losses.

5. Feature Importance And Interpretability

Tree-based models provide measures of feature importance, indicating the relative contribution of each feature to the prediction. High-importance features may include:

- Ratios of current claim amount to historical averages for the same policyholder or provider.
- Length of stay anomalies for particular diagnoses.
- High frequency of claims in recent time windows.
- Deviations of provider behaviour from peer groups.

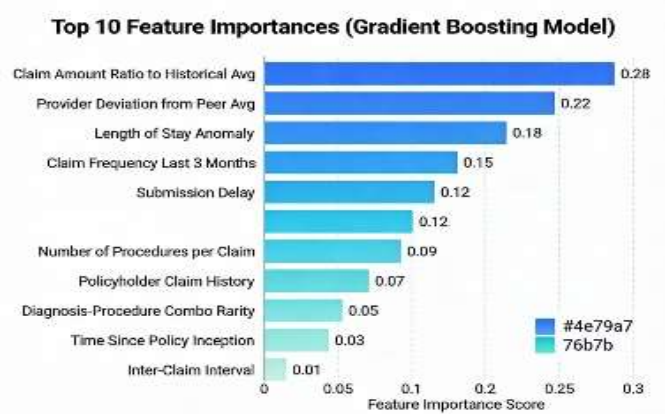


Figure 4: Bar chart of top-k feature importances from the best model.

6. Error Analysis

Error analysis examines false positives and false negatives:

False Positives:

Genuine claims incorrectly flagged as suspicious. These may involve rare but legitimate procedures or unusual combinations of attributes. Understanding these cases can help refine features or incorporate specialized rules to reduce unnecessary alerts.

False Negatives:

Fraudulent claims not detected by the model. These may represent emerging fraud schemes or gaps in the feature set. Investigating these cases guides future data collection and feature design.

VI. DISCUSSION

1. Integration With Existing Processes

For practical deployment, the fraud detection system must integrate with existing claim processing software and workflows. A typical strategy is to use the model output as a fraud risk score that:

Automatically routes low-risk claims to fast-track processing.

Sends medium-risk claims to random sampling or light review.

- Flags high-risk claims for detailed investigation.
- Investigators can access a dashboard showing risk scores, key features, explanations, and historical context for each flagged claim.

2. Human–Machine Collaboration

Machine learning models are intended to assist, not replace, human experts. Investigators bring domain knowledge, contextual understanding, and judgement that models cannot fully replicate. A collaborative loop can be established:

The model suggests suspicious claims and provides explanations.

- Investigators confirm or reject fraud labels and provide qualitative feedback.
- This feedback is logged and used as additional training data in future model iterations.
- Such a loop continuously improves the model while keeping humans in control of high-stakes decisions.

3. Model Governance And Monitoring

Once deployed, the model must be monitored for:

- Data Drift: changes in claim distributions, new procedure codes, or shifts in provider behavior.
- Performance Drift: degradation in metrics such as recall and precision over time.
- Fairness and Bias: ensuring that specific demographic or provider groups are not disproportionately flagged without justification.

Periodic retraining may be scheduled, for example annually or when significant drift is detected. Governance policies should define responsibilities, documentation requirements, and approval processes for major model changes.

4. Ethical And Legal Considerations

Health insurance claim data is sensitive and subject to data protection regulations. The fraud detection system must ensure:

- Secure storage and transmission of data.
- Access control and audit trails to prevent misuse.
- Data anonymization or pseudonymization where appropriate.

Moreover, decisions informed by the model should be explainable to policyholders and providers. Clear communication channels and appeal processes are important to maintain trust and ensure compliance with regulatory requirements.

5. Limitations

The proposed framework has limitations:

- It assumes the availability of labeled data, which may be limited or biased.
- It focuses on supervised models; unsupervised and graph-based approaches may capture additional patterns.
- It abstracts from some domain-specific complexities, such as international coding standards and country-specific regulations.
- Future work can expand the framework to address these limitations.

VII. CONCLUSION AND FUTURE WORK

This paper presented a comprehensive machine learning-based framework for detecting fraudulent health insurance claims. The framework covers data preprocessing, feature engineering, class imbalance handling, algorithm selection, evaluation, and deployment considerations.

It emphasizes not only predictive performance but also interpretability, integration with human investigators, and ongoing model governance.

Machine learning can substantially enhance an insurer's ability to detect suspicious claims, prioritize investigative resources, and reduce financial losses due to fraud. However, success depends on high-quality data, appropriate model design, and careful integration into operational workflows.

Future research directions include:

- Incorporating unstructured data such as clinical notes and scanned documents through natural language processing and image analysis.
- Applying graph-based and deep learning models to capture relationships among patients, providers, and services.
- Developing active and semi-supervised learning strategies to better exploit unlabelled claims.
- Evaluating the impact of such systems in real-world deployments over longer time horizons.

REFERENCES

1. K. Veena and S. Subramanian, "Predicting health insurance claim frauds using supervised machine learning technique," in Proc. 8th Int. Conf. Sci. Technol. Eng. Math. (ICONSTEM), Chennai, India, Apr. 2023, pp. 1-6, doi: 10.1109/ICONSTEM56934.2023.10142604.
2. M. K. Al-Ghazi and M. B. Abbood, "Classification of health insurance fraud risk with machine learning," in Proc. IEEE Int. Conf. Innov. Intell. Syst. Appl. (INISTA), 2024, pp. 1-6, doi: 10.1109/INISTA.2024.10699052.
3. R. Bounab, A. Naamane, and N. Hmina, "Optimizing machine learning for healthcare fraud detection using hybrid feature selection and hyperparameter tuning," in Proc. IEEE Int. Conf. Knowl. Based Eng. Innov. (KBEI), 2024, pp. 45-52, doi: 10.1109/KBEI.2024.10851054.
4. B. K. Sethi and S. Kumar, "Medical insurance fraud detection based on blockchain and machine learning," in Proc. IEEE Int. Conf. Comput. Intell. Comput. Res. (ICCIC), Coimbatore, India, Dec. 2022, pp. 1-6, doi: 10.1109/ICCIC55457.2022.10060811.

5. K. Shruthi, R. Kumar, and S. Gupta, "Healthcare insurance fraud detection powered by blockchain and machine learning: An analysis and framework," in Proc. IEEE Int. Conf. Adv. Comput. Commun. Paradigms (ICACCP), Gangtok, India, 2024, pp. 1-7, doi: 10.1109/ICACCP.2024.10675307.
6. B. Hong et al., "Health insurance fraud detection based on multi-channel heterogeneous graph structured learning," Heliyon, vol. 10, no. 8, Aug. 2024, Art. no. e29456, doi: 10.1016/j.heliyon.2024.e29456.