IET Blockchain Journal

WILEY

IET Blockchain

ORIGINAL ARTICLE OPEN ACCESS

SHIELD: Secure Hybrid Insurance Ledgers & Ensemble Detection for Health Insurance Fraud Prevention

Mubasshir Ahmed[1] | Md. Mazharul Islam[1] | Rajesh Palit[1] | Mohammad Shahriar Rahman[2] | Salekul Islam[1]

[1]Department of Electrical and Computer Engineering, North South University, Dhaka, Bangladesh | [2]Department of Computer Science and Engineering, United International University, Dhaka, Bangladesh |

Correspondence: Salekul Islam (salekul.islam@northsouth.edu)

Received: **Revised:** **Accepted:**

Funding: This work was supported by the Institute for Advanced Research (IAR), United International University, and the Office of Research (OR), North South University, under Grant UIU-IAR-02-2023-SE-41.

Keywords: Blockchain | Health Insurance | Fraud Detection | Risk Scoring | Anomaly Detection | Privacy Preservation

ABSTRACT

Health insurance fraud remains a persistent challenge in modern healthcare ecosystems, increasing insurers' operational costs and, in turn, raising premiums for legitimate beneficiaries. Conventional fraud detection pipelines rely on centralized, rule-based validation, which offers limited auditability and struggles to detect high-dimensional, previously unseen fraud patterns. To address these limitations, this work presents SHIELD, a hybrid privacy-preserving framework that integrates permissioned blockchain technology with machine learning–based fraud risk analytics for claim adjudication. The blockchain layer, implemented using Hyperledger Fabric and IPFS, performs deterministic verification of policy eligibility, timestamp consistency, duplicate submissions, and document integrity, while SHA-256-based anonymization safeguards patient identity. An event-driven off-chain ML engine complements these checks by applying supervised classification for labeled provider-submitted claims and unsupervised anomaly detection for unlabeled patient-submitted claims. Experiments on Medicare fraud datasets show that supervised ensemble models achieve high performance, with accuracy and F1-scores in the 93–95% range. In contrast, the unsupervised ensemble captures additional anomalous patterns beyond rule-based violations. The results demonstrate that combining blockchain assurance with machine learning–based risk scoring broadens fraud coverage, enhances transparency, and preserves privacy during claim processing, offering a practical fraud-aware workflow for digital health insurance infrastructures.

1 | Introduction

Health insurance fraud has become one of the most persistent and financially damaging challenges within modern healthcare ecosystems [1]. Fraudulent activities—including provider abuse, claim inflation, fictitious treatments, and beneficiary misrepresentation—result in substantial financial losses each year and increase premiums for legitimate patients while burdening insurers with escalating administrative costs [2]. As healthcare systems continue to digitize, the volume, diversity, and velocity of claim data have increased considerably, making manual, deterministic fraud-detection strategies increasingly insufficient for identifying sophisticated, coordinated fraudulent schemes [3].

Traditional fraud detection mechanisms in health insurance rely primarily on rule-based verification performed by centralized authorities. These systems are effective at identifying straightforward inconsistencies such as invalid procedure codes, policy mismatches, or duplicate billing [4]. However, they struggle to generalize to complex, multi-entity fraud behavior or to detect evolving patterns engineered specifically to evade deterministic validation logic. Furthermore, centralized infrastructures create challenges related to trust, auditability, tamper resistance, and privacy. A single compromised authority may modify claim records, suppress fraudulent findings, or manipulate decision outcomes, motivating the need for secure, distributed, and privacy-preserving verification mechanisms [5].

Blockchain technology provides structural advantages for modernizing claim processing pipelines [6]. Its decentralized and immutable ledger enables tamper-proof claim lifecycle management, ensures verifiable traceability, and supports privacy-preserving identity management. Smart contracts further enable automated enforcement of deterministic validation rules for service codes, document consistency, policy eligibility, and timestamp correctness [7]. Prior blockchain-based privacy-preserving frameworks for claim adjudication demonstrated that cryptographic credentialing, off-chain document storage, and on-chain rule validation significantly improve transparency, data integrity, and auditability. However, rule-based verification remains inherently limited in detecting emergent or high-dimensional fraud patterns that cannot be expressed by deterministic logic [8].

Machine learning (ML) has strong potential to enhance fraud analytics by learning from historical claim data to identify behavioral patterns and statistical anomalies [9]. ML techniques can model correlations between claim attributes, provider practices, reimbursement structures, and demographic variables to identify suspicious anomalies that are difficult to encode as deterministic rules. Ensemble models and unsupervised outlier detection methods further improve detection coverage by uncovering previously unseen or subtle behaviors [10]. Importantly, in the context of this work, ML is not used to maximize dataset benchmarking performance or to improve prediction accuracy as an academic objective. Instead, ML modules are integrated solely as operational fraud-detection components, designed to generate probabilistic risk assessments that enhance claim adjudication and reduce undetected fraud [11].

To address the limitations of purely rule-based blockchain verification and solely data-driven ML detection, this paper introduces a unified hybrid architecture termed SHIELD—a privacy-preserving health insurance fraud prevention system that integrates blockchain-based claim verification with ensemble-based fraud risk analytics. SHIELD maintains all core functionalities of the prior blockchain framework, including decentralized execution, off-chain encrypted storage, credential anonymization, and tamper-proof auditing, while introducing an adaptive ML risk engine that supports supervised provider-level fraud detection and unsupervised anomaly detection for unlabeled patient-submitted claims. The ML modules operate entirely off-chain and are triggered through blockchain events, ensuring scalability, modularity, and strict privacy boundaries. A decision service bridges the blockchain network and the ML engine by forwarding validated claim metadata for off-chain scoring and returning fraud risk scores for downstream adjudication.

The resulting architecture implements a three-tier fraud detection pipeline consisting of: (i) deterministic rule-based validation on the blockchain, (ii) supervised ensemble classification for labeled fraud data, and (iii) unsupervised anomaly detection for unlabeled claim submissions. Sensitive claim documents remain stored off-chain via IPFS, and only cryptographic hashes and compact ML-derived risk summaries are committed to the ledger. This hybrid design enables transparent, auditable, and privacy-preserving claim processing while providing stronger fraud resilience than either rule-based or ML-only methods.

The above discussion highlights a gap between privacy-preserving blockchain verification systems—focused primarily on deterministic rule enforcement—and machine learning-based systems—focused on probabilistic fraud analysis. Existing blockchain approaches improve auditability and storage integrity but lack adaptivity to evolving fraud behaviors, while ML systems lack trust guarantees, verifiable execution, and privacy protection. Few existing frameworks combine both paradigms into a cohesive pipeline capable of detecting both known and emerging fraud patterns across heterogeneous claim submission scenarios. This work closes the gap by integrating blockchain-based verification with ensemble-based fraud analytics into a unified, operational framework for health insurance fraud detection. The main contributions of this work are summarized as follows:

- A trust-aware hybrid fraud detection architecture is introduced that integrates blockchain-based rule validation with supervised and unsupervised machine learning, enabling secure and auditable insurance claim assessment.
- Two realistic operational claim submission scenarios—patient-initiated and healthcare provider–initiated—are modeled, and scenario-specific learning strategies are designed based on data availability and fraud characteristics.
- A stacking-based supervised hybrid classifier is developed, which combines Random Forest, XGBoost, and CatBoost models using Logistic Regression as a meta-learner, achieving improved classification performance on labeled patient claims.
- An unsupervised anomaly ensemble framework is proposed that fuses multiple complementary anomaly detectors through voting and score aggregation, facilitating the discovery of emerging or previously unseen fraud patterns.
- A unified decision-fusion mechanism is designed that integrates blockchain rule outputs, supervised fraud probabilities, and unsupervised anomaly scores for scalable deployment in real-world insurance workflows.

The remainder of this paper is organized as follows. Section 2 introduces the operational context of health insurance claims and outlines the technical foundations that motivate a combined blockchain–machine learning approach. Section 3 surveys the state of the art in blockchain-based verification, supervised fraud prediction, anomaly detection, and hybrid strategies, positioning the proposed system within current research trends. Section 4 presents the SHIELD framework, describing its architectural components, security assumptions, and claim submission model. Section 5 explains the implementation of the machine learning risk engine and its integration with the blockchain execution environment. Section 6 analyzes the security and privacy properties of the system and discusses how the architectural design mitigates fraudulent behaviors and adversarial manipulation. Section 7 reports supervised classification performance and situates the results within recent fraud detection literature. Finally, Section 8 concludes the paper and outlines prospective research directions.

2 | Preliminaries

This section introduces the foundational concepts underpinning the proposed hybrid blockchain–machine learning framework. We describe how health insurance claims are structured, how fraud manifests operationally, and how blockchain, anonymous

credentialing, IPFS storage, and machine learning jointly contribute to secure and intelligent fraud screening.

2.1 | Health Insurance Claims and Fraud Landscape

A health insurance claim represents a structured request for financial reimbursement submitted either by a patient or by a healthcare provider [12]. Claims contain demographic information, diagnosis and procedure codes (ICD, CPT/HCPCS), timestamps, reimbursement amounts, deductible allocations, and policy metadata. Fraud in this domain corresponds to deliberate misrepresentation intended to obtain unauthorized financial benefits and manifests in multiple operational forms, including billing irregularities (e.g., inflated charges, unbundled procedures, or services not rendered), identity and provider misuse (e.g., duplicate submissions or falsified provider credentials), and beneficiary-level misconduct [13].

Behavioral and temporal inconsistencies also constitute important fraud signals, including abnormal visit frequencies, overlapping treatment intervals, geographic implausibility between service locations, or claims submitted for deceased policyholders. These characteristics motivate the dual use of deterministic rule validation and statistical machine learning in later sections.

2.2 | Blockchain-Based Claim Integrity and Credentialing

A permissioned blockchain (Hyperledger Fabric) is utilized to enforce verifiable claim integrity and tamper-resistant credential management. Blockchain provides immutability, decentralized validation, and auditable transaction history, enabling claim metadata and credential events to be stored in a manner that cannot be retroactively altered [14]. Smart contracts encode deterministic validation rules and credential checks, while privacy-preserving anonymous identifiers derived via one-way SHA-256 hashing allow claims to be validated without exposing real-world identities. An event-driven execution model allows smart contracts to emit claim validation events (e.g., `ClaimValidated`) that trigger off-chain machine learning pipelines responsible for fraud risk assessment and anomaly scoring. This division of labor preserves privacy and ensures that model updates do not require blockchain consensus modifications [15].

2.3 | Off-Chain Document Management via IPFS

Insurance claims often include supporting medical evidence such as laboratory reports, hospital invoices, radiology images, and discharge summaries. Storing such documents directly on-chain would impose excessive storage overhead and confidentiality risks. Instead, sensitive documents are stored in the InterPlanetary File System (IPFS), while the blockchain stores only cryptographic file hashes [16]. This architecture prevents blockchain state bloat, maintains patient privacy, and enables verifiable integrity checks, since any modification to the off-chain file would alter its hash and be rejected by the validation logic.

2.4 | Machine Learning Foundations for Fraud Detection

Machine learning serves as the analytical layer that identifies suspicious claim behavior beyond deterministic rule enforcement [17]. Both supervised and unsupervised paradigms are required due to the heterogeneous nature of claim submissions and the scarcity of labeled fraud data. Supervised models leverage historical labeled Medicare datasets and exhibit strong effectiveness in provider-level classification tasks, particularly with ensemble models such as Random Forest, XGBoost, and CatBoost that capture nonlinear utilization patterns. In contrast, unsupervised learning detects emerging fraud behaviors in unlabeled datasets by identifying statistical deviations from normal claim trajectories using techniques such as Isolation Forest, One-Class SVM, or neural Autoencoder reconstruction scoring [18].

Because fraud often emerges through behavioral patterns across multiple claims, feature aggregation at both provider and beneficiary levels is employed to expose institutional and temporal anomalies that would not be observable from single claims in isolation.

3 | Literature Review

Health insurance fraud detection has become an increasingly important research area as healthcare systems worldwide digitize their claim processing infrastructures. The literature spans blockchain-based solutions, supervised machine learning, unsupervised anomaly detection, and hybrid computational models, all aimed at enhancing detection accuracy, preserving privacy, and improving claim integrity. This section critically reviews the most relevant existing works, beginning with our baseline system, followed by additional studies covering complementary techniques that directly motivate the proposed hybrid blockchain–machine learning framework.

3.1 | Baseline Privacy-Preserving Blockchain System

Islam et al. [19] introduced a privacy-preserving health insurance framework that employed a blockchain-based architecture using Hyperledger Fabric, the InterPlanetary File System (IPFS), and SHA-256–based anonymous credentials to secure patient identity, ensure document integrity, and support deterministic rule-based fraud checks through smart contracts. The model effectively prevented common fraud types—duplicate claims, falsified timestamps, invalid procedure codes, and document tampering—while maintaining strict privacy guarantees.

However, because the system relied exclusively on static rule-based logic, it was unable to identify evolving or previously unseen fraud behaviors such as unusual utilization patterns, coordinated provider–patient collusion, or subtle anomalies across aggregated claim histories. This limitation motivates the integration of machine learning (ML)–driven analytics, which can adapt to complex and high-dimensional behavioral patterns beyond the expressive capacity of handcrafted rules.

3.2 | Supervised Machine Learning Approaches

A substantial body of research has explored supervised learning for fraud detection using labeled healthcare claim datasets. These works demonstrate the predictive strength of ML algorithms but often lack privacy features and structural integration with secure claim-processing frameworks.

Zhang et al. [21] presented a medical fraud and abuse detection system using Support Vector Machines, Decision Trees, and Random Forests. Their study demonstrated that tree-based models outperform linear baselines and highlighted the importance of comparative model evaluation and feature selection. However, the approach focused primarily on claim-level features and did not fully exploit behavioral aggregation across providers and beneficiaries.

Bayerstadler et al. [20] proposed a Bayesian multinomial latent variable model for detecting fraud and abuse in health insurance data. Their probabilistic formulation captured hidden relationships in claim counts and cost structures, offering interpretable latent factors and uncertainty-aware predictions. Nonetheless, the model faces scalability challenges in modern, high-dimensional claim environments and does not address issues such as temporal dynamics or large-scale deployment.

Permai and Herdianto [37] applied Logistic Regression and XGBoost to predict health insurance claim outcomes. Their empirical results emphasized the superiority of gradient boosting techniques over linear models when dealing with complex claim features. However, the work focused on claim acceptance prediction rather than explicit fraud labeling and did not consider adversarial or collusive fraud behaviors.

Veena et al. [35] investigated multiple supervised ML algorithms, including Logistic Regression, K-Nearest Neighbors, Support Vector Machines, Decision Trees, and Random Forests, for predicting fraudulent health insurance claims. Their findings reinforced the empirical dominance of tree-based ensembles in fraud classification tasks, though the study relied on relatively small datasets and limited feature engineering.

Chitteti et al. [36] designed an ML-based healthcare insurance fraud detection system and highlighted several practical challenges, such as handling missing values, noisy attributes, and heterogeneous claim formats. While demonstrating the feasibility of deploying ML classifiers in real settings, their framework did not integrate with a secure or decentralized claim infrastructure.

Majumder [38] proposed an intelligent fraud detection system using a modular ML pipeline that compares the performance of multiple classifiers. The system architecture is flexible and generalizable to different datasets; however, it operates entirely off-chain, does not incorporate privacy-preserving mechanisms, and lacks unsupervised modules for detecting unknown fraud patterns.

Ashok and Durge [39] explored regression-based approaches for fraud detection and prevention in healthcare insurance claims. By modeling fraud risk as a continuous score, their system supports graded decision-making. Nevertheless, regression models are less suited to highly imbalanced fraud classification problems and may struggle to capture nonlinear interactions that characterize complex fraudulent behavior.

Taken together, these supervised learning approaches achieve high predictive performance on labeled datasets and underscore the importance of tree-based ensembles. However, they generally do not address privacy preservation, are not integrated with blockchain-based claim workflows, and rarely incorporate anomaly detection mechanisms for emerging fraud patterns.

3.3 | Unsupervised and Anomaly Detection Methods

Since fraud cases are rare and labels may be incomplete or unavailable, unsupervised anomaly detection plays a crucial role in identifying emerging or previously unknown fraud patterns.

Winarso et al. [27] evaluated ensemble-based anomaly detection techniques—including Isolation Forest, Local Outlier Factor, and One-Class Support Vector Machines—for validating health insurance costs. Their results showed that aggregating anomaly scores from multiple detectors leads to more stable rankings of suspicious claims, highlighting the value of ensemble anomaly detection for complex datasets.

He et al. [28] developed a data-driven intelligent supervision system to generate high-risk organized fraud clues in medical insurance funds. Their framework combines clustering, correlation analysis, and rule mining to discover suspicious groups of providers and beneficiaries. This work emphasizes the importance of group-level and network-level fraud analysis, an insight that complements provider- and beneficiary-level aggregation in our proposed framework.

Feng et al. [30] proposed a medical insurance fraud risk monitoring and identification model that integrates feature selection with machine learning classifiers. By employing methods such as LASSO and ReliefF, they reduced dimensionality and improved model generalization in noisy, high-dimensional claim spaces. Their results demonstrate that careful feature selection is an essential precursor to both supervised and unsupervised fraud detection.

Aljufri et al. [29] presented a domain-knowledge-based feature engineering approach for fraud detection using health administrative claims. Encoding expert knowledge into engineered features and constraints, they showed significant performance improvements in downstream fraud detection models. This reinforces the importance of combining data-driven methods with domain expertise, particularly for designing aggregated financial and utilization indicators.

Du and Yu [26] applied the Isolation Forest algorithm to large-scale medical insurance big data for fraud detection. They reported that Isolation Forest is effective in identifying rare, extreme claim patterns in high-dimensional spaces, especially when combined with appropriate feature scaling and preprocessing. Their findings support the use of Isolation Forest as a core unsupervised component in fraud analytics pipelines.

These unsupervised studies collectively demonstrate that anomaly detection is essential for capturing fraud behaviors that do not appear in labeled training datasets. However, current works rarely incorporate secure data infrastructures, nor do they integrate anomaly scores with deterministic on-chain verification mechanisms or supervised ML predictions in a unified system.

3.4 | Hybrid and Rule-Augmented Approaches

Several studies attempt to bridge supervised and unsupervised learning or incorporate domain rules into data-driven models.

For example, Bayerstadler et al. [20] implicitly combined statistical structure and domain knowledge through their latent variable framework, while Winarso et al. [27] advocated ensemble anomaly detection as a way to improve robustness. Feng et al. [30] integrated feature selection with supervised classification to balance complexity and interpretability.

Despite these advances, none of the existing works provide an end-to-end, privacy-preserving, blockchain-enabled, hybrid ML system that simultaneously supports deterministic rule verification, supervised fraud classification, unsupervised anomaly detection, secure identity protection, and tamper-proof document validation. In particular, there is a lack of event-driven integration between blockchain smart contracts and off-chain ML services, and little attention is given to combining rule-based checks with probabilistic risk scoring in a unified decision pipeline.

3.5 | Critical Gap Analysis

Across the reviewed literature, several limitations consistently emerge:

- **Lack of privacy preservation:** Most ML-based fraud detection systems process raw patient and provider data directly, without anonymization or secure off-chain storage. Sensitive information is exposed during model training and inference, posing privacy and regulatory concerns.
- **Absence of blockchain-integrated ML workflows:** While blockchain offers strong guarantees of immutability and auditability, most prior studies treat fraud detection as a separate, off-chain analytical process. There is minimal work on event-driven integration between on-chain claim verification and off-chain ML inference.
- **Limited hybrid frameworks:** Existing systems typically rely on either deterministic rules, supervised learning, or unsupervised anomaly detection in isolation. Few, if any, combine these techniques into a coherent hybrid framework that can capture both known and emerging fraud patterns.
- **Insufficient modeling of complex fraud behavior:** Frauds involving temporal inconsistencies, provider–patient collusion, or abnormal utilization patterns require temporal modeling and multi-level aggregation. Many works focus solely on individual claim records, ignoring provider- and beneficiary-level behavior.
- **Lack of explainable, auditable end-to-end systems:** Although some works investigate interpretability (e.g., through SHAP or LIME), these explanations are not embedded into tamper-proof ledgers or combined with rule-based evidence in a unified decision trail.

3.6 | Summary and Comparative Analysis

The literature demonstrates strong progress in machine learning–based fraud detection and meaningful advances in blockchain-based claim verification. However, no prior work effectively integrates privacy-preserving identity generation, IPFS-based document integrity, smart-contract rule validation, supervised ML fraud classification, unsupervised anomaly detection, and event-driven blockchain-to-ML communication into a single unified architecture. The proposed system addresses these gaps through a hybrid blockchain–ML framework that ensures strong privacy, adaptive fraud detection, and transparent, verifiable claim handling across the entire health insurance claim lifecycle.

The reviewed literature reveals several common trends. First, tree-based ensemble methods (Random Forest, XGBoost, CatBoost) consistently outperform linear models due to their ability to capture nonlinear interactions and heterogeneous claim behaviors. Second, feature engineering—particularly provider-level and beneficiary-level aggregation—is critical for effective fraud detection, as it captures behavioral patterns that individual claims cannot reveal. Third, unsupervised methods such as Isolation Forest and clustering-based techniques are essential for detecting novel fraud patterns in unlabeled or weakly labeled datasets, thereby complementing supervised learning. Finally, only a limited number of works combine these analytics with privacy-preserving, blockchain-based infrastructures, highlighting the uniqueness and necessity of the proposed hybrid framework.

To position our work relative to closely related studies not discussed in detail above, Table 1 summarizes a set of recent contributions that address healthcare insurance fraud detection using machine learning and/or blockchain. Each entry highlights the main idea, advantages, limitations, learning algorithms, datasets, and reported outcomes. Importantly, none of the works in the table are analyzed in the preceding subsections, thereby avoiding repetition while providing a comprehensive comparative view.

4 | Proposed Model

The proposed framework extends our prior blockchain-based privacy-preserving health insurance architecture by incorporating a machine learning (ML)–driven fraud risk assessment layer. While the blockchain ensures deterministic verification, immutability, and trustless coordination among participants, the ML engine identifies high-dimensional fraud patterns that traditional systems are unable to capture. The combined system establishes a hybrid fraud detection pipeline, where rule-based smart contract checks are performed on-chain, and probabilistic ML-based anomaly scoring is executed off-chain via an event-driven interface.

The overall architecture consists of five major components: (i) patients and healthcare provider interfaces, (ii) blockchain network with smart contracts, (iii) IPFS-based off-chain storage, (iv) ML risk scoring engine, and (v) decision service. Claims are first validated through on-chain rules, after which valid claims are forwarded to the ML engine for risk scoring. The result is then relayed to the insurer, enabling automated approval, rejection, or escalation for manual review.

In this section, we outline the threat assumptions, security properties, and detailed submission workflows for two operational scenarios: claim submission by a patient and claim submission by a healthcare provider.

TABLE 1 | Summary of related works on healthcare insurance fraud detection.

Author	Main Idea	Pros	Cons	ML	Dataset	Result
Kumaraswamy et al. [22] (2022)	Domain-specific feature engineering for healthcare claim fraud	Shows that enriched claim-level and patient-level features improve fraud detection	Focuses on feature design; lacks deployment considerations and privacy support	LR, RF, XGBoost, etc.	US commercial claim data	AUC and F1 improved over baseline feature sets
Nalluri et al. [23] (2023)	Builds predictive models and identifies key fraud factors	Systematic model comparison and feature relevance analysis	No blockchain, privacy, or operational integration with insurer workflows	RF, XGBoost, SVM, KNN, etc.	Taiwan insurance data	High accuracy and AUC; key predictive variables reported
Hancock et al. [24] (2023)	Explainable ML for Medicare fraud detection	Provides interpretable outputs for auditor review via SHAP	Focuses on explainability; no hybrid or privacy-preserving design	Tree ensembles, XGBoost	US Medicare (CMS)	Strong predictive performance with feature attributions
Nabrawi and Alanazi [34] (2023)	Supervised ML framework for healthcare fraud detection	Compares several classifiers and evaluates PR trade-offs	Lacks blockchain integration and unsupervised fraud detection	RF, DT, LR, SVM, etc.	Real-world insurance claim data	Competitive accuracy and F1-score
Settipalli and Gangadharan [25] (2023)	Unsupervised multivariate fraud detection (WMTDBC)	Detects anomalies without labeled data	Model interpretability and production integration limited	Unsupervised multivariate analysis	Indian health insurance data	High anomaly detection rate
Mohammed et al. [31] (2023)	Blockchain-based healthcare network with ML scoring	Combines blockchain transparency with ML detection	Evaluated on simulated data; limited real claim evidence	RF, SVM, KNN, NB	Synthetic network data	Improved accuracy/F1 over baseline
Kaafarani et al. [32] (2024)	Platform recommender for blockchain-based insurance solutions	Provides comparative guidance on smart-contract platforms	No fraud models or complete detection workflow implemented	No ML	Multiple blockchain testbeds	Qualitative recommendations for developers
Wang et al. [3] (2025)	Robust interpretable ensemble for insurance fraud	Balances predictive performance with interpretability	No blockchain or privacy-preserving identity support	GBDT, RF, XGBoost	Chinese medical insurance data	Higher AUC and F1 than baselines
Kapadiya et al. [33] (2025)	Blockchain-assisted ensemble ML for insurance fraud	Improves traceability and decision transparency	Privacy and off-chain analytics not fully specified	RF, XGBoost, SVM (ensemble)	Synthetic + real datasets	Higher accuracy and security than ML-only frameworks
SHIELD (this paper)	**Hybrid privacy-preserving blockchain and ML fraud detection framework**	**IPFS, smart-contract, supervised and unsupervised ML, and anonymous credentials**	**Prototype evaluated offline; real-time deployment**	**RF, XGBoost, CatBoost, Isolation Forest, LOF, OC-SVM, Autoencoder**	**Medicare fraud dataset, CMS provider data, synthetic claim data**	**High AUC/F1 for supervised models; improved anomaly coverage**

4.1 | Threat Model

The system considers adversaries who may attempt to manipulate claim data, impersonate identities, falsify medical services, or collude with other actors to obtain unauthorized reimbursements. We categorize potential threats as follows:

- **Malicious Patient:** A patient may submit false claims, alter diagnosis information, misuse expired or inactive policies, or collude with providers to fabricate treatments.
- **Malicious Healthcare Provider:** A provider may engage in upcoding, unbundling, billing for unprovided services, duplicate submissions, or altering timestamps to hide inconsistencies.
- **Compromised Insurance Agent:** An internal actor may unlawfully modify claim records, deny legitimate claims, or manipulate policy metadata.
- **Network Adversary:** An external attacker may attempt replay attacks, data tampering, document substitution, or unauthorized access.
- **Model Evasion Adversary:** An attacker may attempt to engineer adversarial inputs to evade ML-based fraud detection.

We assume standard cryptographic primitives remain secure and that the blockchain consensus mechanism cannot be subverted

by a single entity. Hyperledger Fabric's permissioned model prevents Sybil attacks, while IPFS integrity checks remove the risk of external document tampering.

4.2 | Security and Privacy

Security and privacy are enforced across multiple layers of the architecture:

- **Anonymous Credential Generation:** Patient and provider identities are anonymized using SHA-256–derived credentials, ensuring unlinkability and preventing identity disclosure.
- **Permissioned Blockchain Security:** Hyperledger Fabric restricts ledger access to authenticated entities only. Endorsement policies and MSP (Membership Service Provider) guarantee integrity and trust among participants.
- **Smart Contract Verification:** Deterministic on-chain rule-based checks detect invalid service codes, policy mismatch, duplicate claims, fraudulent timelines, and fake providers before ML analysis is triggered.
- **IPFS-Based Storage Integrity:** Sensitive medical documents stored off-chain are protected via content hashing. Any alteration results in a mismatched hash, enabling tamper detection.
- **Event-Driven ML Isolation:** The ML engine operates off-chain and receives only necessary anonymized claim summaries. Raw data is never exposed to the model or external services.
- **Privacy Preservation:** No personally identifiable information (PII) is stored on-chain. All identifiers are hashed; only verifiable proofs and ML risk scores are recorded on the ledger.
- **Adversarial Robustness:** Ensemble ML models reduce the risk of adversarial evasion by leveraging heterogeneous decision boundaries (CatBoost, XGBoost, Random Forest). Isolation Forest adds robustness against abnormal unseen patterns.

Together, these properties provide confidentiality, integrity, auditability, and strong fraud resistance throughout the entire claim life cycle.

4.3 | System Overview

This section presents the proposed blockchain-enabled and machine learning–assisted framework for secure and privacy-preserving health insurance claim processing. The model integrates smart contract–based rule enforcement with off-chain fraud risk assessment to improve the reliability and transparency of claim verification. Blockchain technology is used to ensure integrity, traceability, and tamper resistance of claim records, while machine learning provides probabilistic fraud risk scores to support informed decision-making by the insurance provider. An overview of the system architecture is shown in Fig. 1. The proposed system supports two claim submission scenarios:

i. Claim submission by the patient

ii. Claim submission by the healthcare provider

Although the initiating entity differs between these scenarios, all subsequent processing stages—including on-chain rule-based validation, off-chain fraud risk scoring, and final decision-making—remain identical. Separate figures are provided to explicitly illustrate the different claim initiation paths while preserving a unified validation and decision pipeline.

The proposed architecture consists of the following core components:

- Patient and healthcare provider interfaces for claim preparation and submission,
- A permissioned blockchain network hosting smart contracts for rule-based claim verification,
- Off-chain storage using the InterPlanetary File System (IPFS) for encrypted claim documents,
- An event-driven decision service for orchestration and feature preparation,
- An off-chain machine learning fraud risk scoring engine, and
- The health insurance provider as the final decision authority.

Sensitive medical data and personally identifiable information are never stored on the blockchain. Instead, encrypted claim documents reside off-chain in IPFS, while only cryptographic hashes, validation metadata, and verification outcomes are recorded on-chain.

4.4 | Claim Submission

This subsection describes the two operational paths for submitting claims supported by the proposed system. In real-world workflows, a claim may be submitted either by the patient after service delivery or directly by the healthcare provider. Although the initiating entity differs, both paths ultimately trigger the same verification sequence, comprising on-chain rule validation, off-chain fraud risk scoring, and final adjudication by the insurer. The following discussion presents each scenario in detail.

4.4.1 | Scenario 1: Patient-Initiated Claim Submission

In this scenario, the patient submits the insurance claim after receiving medical treatment. The workflow begins with claim preparation and proceeds as follows. The patient shares a valid identifier with the healthcare provider during service delivery, upon which the provider generates a service token corresponding to the medical service performed. Using the patient identifier, provider identifier, service token, and timestamp, an anonymized claim credential is constructed. A nonce or salt is included to prevent linkability across multiple claims.

The encrypted claim documents and associated metadata are stored off-chain in IPFS. The resulting content hash serves as an immutable reference to the stored data. The patient submits the claim reference, including the IPFS hash and credential commitment, to the health insurance provider through the designated submission interface. The patient-initiated claim submission workflow is illustrated in Fig. 2.

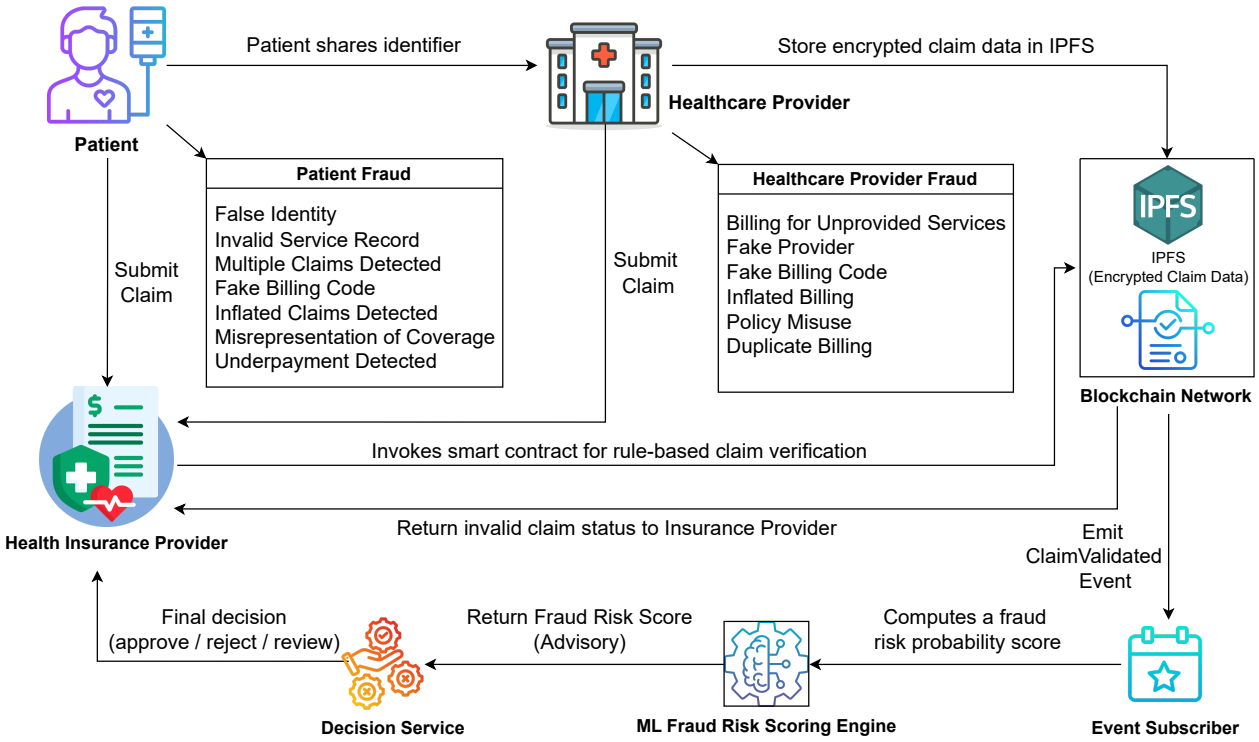


FIGURE 1 | System overview of the SHIELD framework for blockchain-based claim validation and ML-assisted fraud scoring.

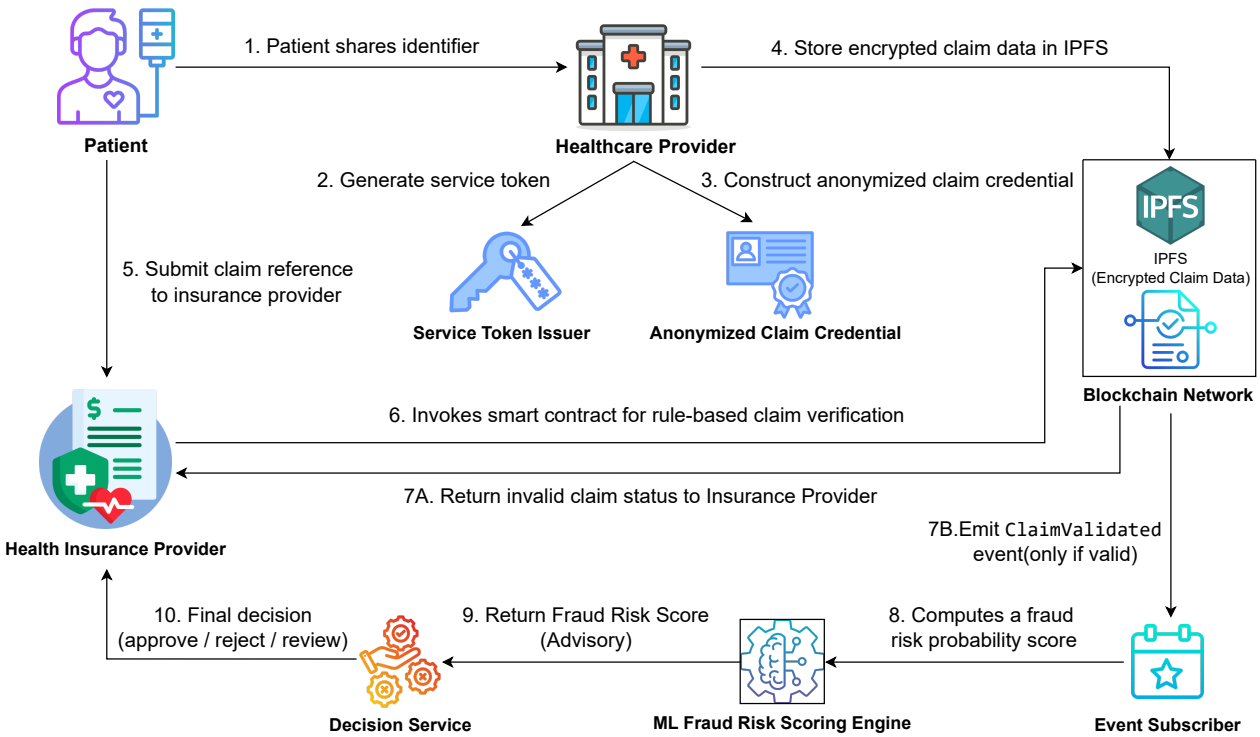


FIGURE 2 | Patient-initiated claim submission workflow.

4.4.2 | Scenario 2: Healthcare Provider-Initiated Claim Submission

In this scenario, the healthcare provider submits the claim on behalf of the patient. The initial claim preparation steps closely mirror those of the patient-initiated workflow, with the distinction that the provider acts as the submitting entity. The patient provides a valid identifier to the provider, who generates a service token for the delivered medical service and constructs an anonymized claim credential using the patient identifier, provider identifier, service token, and timestamp. The encrypted claim data are stored off-chain in IPFS, producing a content hash that is subsequently included in the claim reference.

The healthcare provider submits the claim reference directly to the insurance provider through the submission interface. The healthcare provider–initiated claim submission workflow is illustrated in Fig. 3.

Although the entry points differ, both submission scenarios result in a standardized claim representation that encapsulates credential metadata and a verifiable off-chain document reference. Subsequent execution follows a unified processing path enforced within the system.

4.4.3 | Common Validation and Decision Pipeline

For both submission scenarios, the health insurance provider initiates claim verification by invoking the smart contract deployed on the permissioned blockchain network. The smart contract performs rule-based checks, including policy validity, provider registration, service code consistency, duplicate claim detection, and temporal coherence. If the claim fails any verification rule, the smart contract returns an invalid claim status to the insurance provider. The rejection outcome is recorded on-chain, and the process terminates without invoking machine learning.

If the claim satisfies all verification rules, the smart contract emits a `ClaimValidated` event. This event indicates that the claim has passed on-chain validation and is eligible for further off-chain analysis. An off-chain event subscriber listens for the `ClaimValidated` event and extracts anonymized, non-sensitive claim features required for fraud analysis. These features are forwarded to the machine learning fraud risk scoring engine, which computes a probabilistic fraud risk score using either supervised or unsupervised models, depending on label availability.

The fraud risk score is returned to the insurance provider as a supporting signal. Based on the verification outcome and the ML-generated risk score, the insurance provider makes the final decision to approve the claim, reject it, or escalate it for manual review. Based on the above workflow and system design, the following section describes the implementation of the supervised and unsupervised machine learning models used for off-chain fraud risk scoring.

5 | Implementation

The proposed system integrates machine learning-based fraud analytics with blockchain-enabled verification and secure off-chain computation. This section presents the datasets used, the model implementation pipeline, and the integration between the

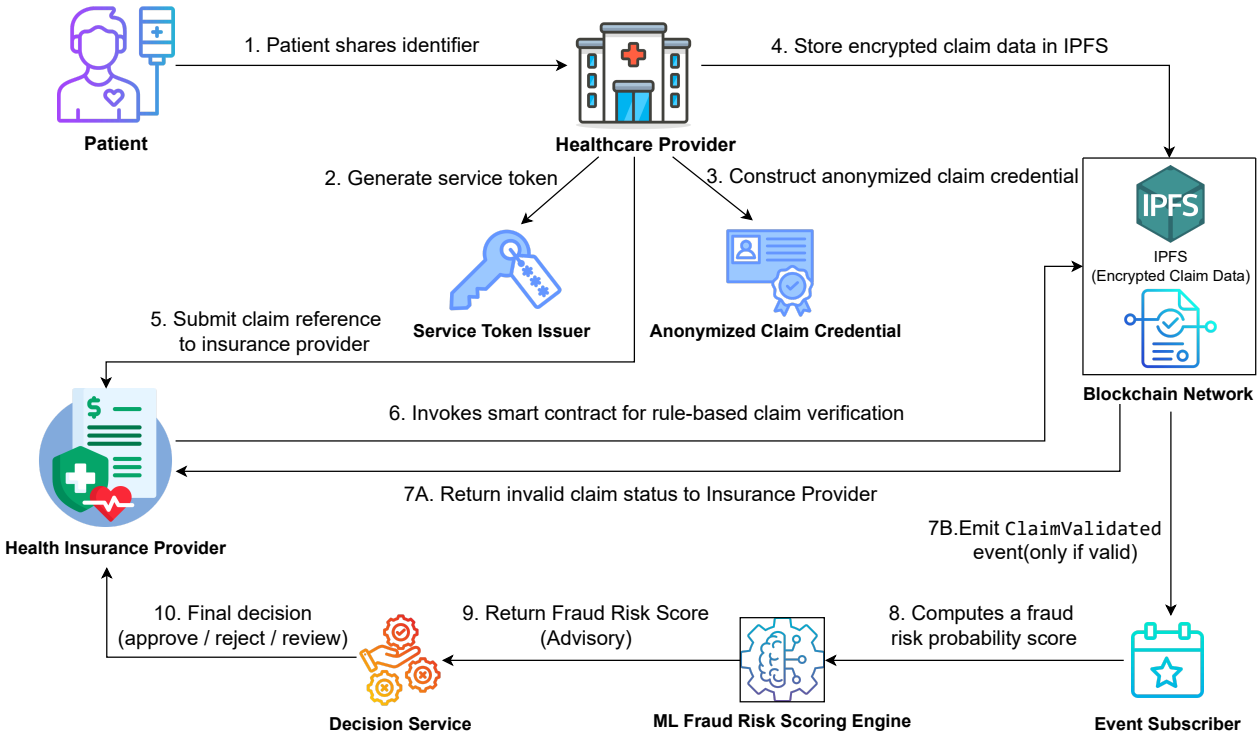


FIGURE 3 | Healthcare provider-initiated claim submission workflow.

machine learning components and the blockchain verification layer.

5.1 | Data Sources and Preparation

The implementation of the fraud detection framework relies on two major datasets: a *supervised dataset* for training classification models for fraud prediction and an *unsupervised dataset* for anomaly detection. This subsection describes the supervised dataset and provides an exploratory analysis of its statistical properties, which informs feature engineering and model development.

5.1.1 | Supervised Dataset

Healthcare fraud in the Medicare ecosystem often involves billing for non-rendered services, misrepresenting procedure codes, upcoding, duplicate submissions, and inflating reimbursements. These behaviors introduce atypical statistical patterns into claim histories that supervised machine learning models can capture. The primary supervised dataset used in this study is obtained from [40], which focuses on detecting potentially fraudulent Medicare providers based on historical billing and utilization patterns. The dataset includes detailed inpatient, outpatient, and beneficiary-level records spanning multiple medical procedures, diagnoses, demographic attributes, and reimbursement values. It contains binary fraud labels at the provider-level, allowing models to learn discriminative patterns that separate fraudulent from legitimate providers.
The dataset consists of three major relational tables:

- **Inpatient claims** containing admissions, diagnosis and procedure codes, lengths of stay, and reimbursement values.
- **Outpatient claims** containing visit-level diagnoses, procedure codes, and billing details without inpatient admission.
- **Beneficiary information** capturing demographics, chronic conditions, and mortality status.

These tables are linked using the attributes `ProviderID` and `BeneID`, enabling the construction of an integrated relational view of provider–beneficiary interactions. Provider-level aggregation is then performed to summarize utilization statistics, reimbursement amounts, and beneficiary condition profiles across all claims associated with a given provider. The resulting merged dataset maps each provider to a binary fraud label (*fraudulent* or *legitimate*), yielding a supervised learning formulation in which the objective is to predict whether a provider exhibits fraudulent billing behavior based on aggregated claim features.

5.1.2 | Unsupervised Dataset

In addition to the supervised dataset, two large-scale healthcare datasets were employed to support unsupervised anomaly detection experiments. Unlike supervised fraud classification, these datasets do not contain fraud labels, making them suitable for detecting unusual behavioral patterns among providers and claims without requiring predefined categories.
The first unsupervised dataset is the *Medicare Physician Provider and Service Resampled* dataset published by Drasco [41]. This dataset is derived from the publicly available Medicare Provider Utilization and Payment Data released by the U.S. Centers for

Medicare & Medicaid Services (CMS). It contains provider-level billing, service utilization, and payment summaries across multiple HCPCS procedure codes. Key attributes include:

- Provider identifiers (e.g., NPI, entity type, specialty),
- Geographical attributes (state, ZIP code),
- Utilization metrics (number of beneficiaries, number of services, service days),
- Reimbursement statistics (submitted charges, allowed amounts, standardized payments).

These features allow unsupervised models to identify providers exhibiting unusually high utilization or billing patterns relative to national averages.

The second dataset is the *Enhanced Health Insurance Claims Dataset* published by Nash [42]. This synthetic dataset contains 4,500 individual health insurance claims generated using the *Faker* library to mimic realistic healthcare billing behavior. The dataset includes:

- Claim-level identifiers (ClaimID, PatientID, ProviderID),
- Demographic fields (age, gender, marital status),
- Medical fields (diagnosis codes, procedure codes, provider specialty),
- Economic attributes (claim amount, patient income),
- Administrative fields (claim type, submission method, claim status).

Although synthetic, the dataset introduces irregularities such as inconsistent reimbursement ratios and heterogeneous submission patterns, making it a valuable source for anomaly detection experiments.

Together, these two datasets support unsupervised analysis at both the provider level and the claim level. The Medicare dataset captures aggregate provider billing behavior in real-world settings, whereas the synthetic claims dataset provides diverse fine-grained claim records suitable for modeling contextual irregularities. Their complementary structures enable unsupervised algorithms to identify outliers that may correspond to potential fraud, erroneous data entry, or rare claim patterns not represented in supervised labels.

5.1.3 | Exploratory Data Analysis and Interpretations

An exploratory data analysis (EDA) was performed on the supervised dataset to understand demographic distributions, healthcare utilization trends, and potential signals relevant for fraud classification. The Key observations include:

- **Fraud Label Distribution:** The dataset exhibits a pronounced class imbalance, with fraudulent providers constituting only a small minority of the population. This skewed distribution reflects real-world healthcare fraud dynamics, where illegitimate billing activity occurs far less frequently than legitimate claims. Class imbalance poses challenges for supervised learning, as conventional classifiers tend to favor the majority class, motivating the use of stratified sampling, cost-sensitive learning, or anomaly detection techniques to improve minority-class detection. The distribution of fraud labels in the dataset is shown in Fig. 4.

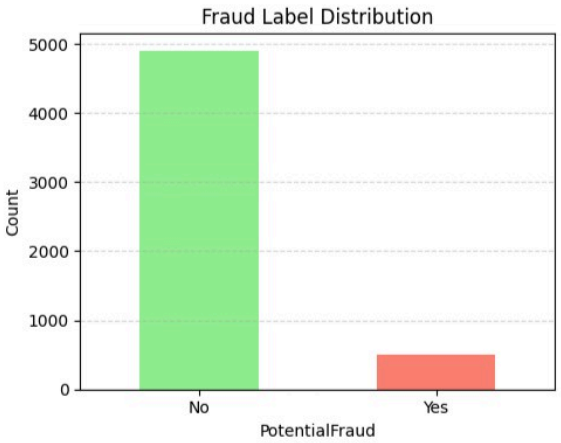


FIGURE 4 | Distribution of fraud labels.

- **Race Distribution:** The majority of beneficiaries belong to race category "1", whereas the remaining categories appear with substantially lower frequency. This demographic skew reflects the composition of the Medicare population and carries modeling implications, since epidemiological patterns, procedure utilization, and reimbursement behavior may differ across demographic groups. Moreover, ensuring that predictive models do not learn spurious correlations associated with demographic attributes is essential to avoid fairness concerns in fraud classification. The distribution of race categories in the dataset is illustrated in Fig. 5.

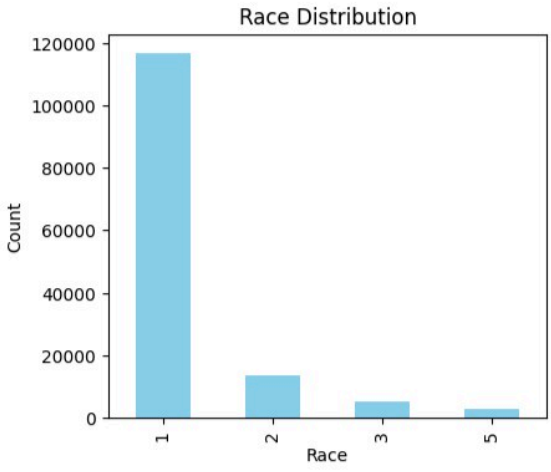


FIGURE 5 | Race distribution of beneficiaries.

- **Mortality Distribution:** The dataset indicates that the vast majority of beneficiaries are alive at the time of service, with deceased individuals representing only a very small proportion of the population. Mortality status is operationally relevant for fraud detection because providers may submit claims for beneficiaries after their recorded date of death, generating temporal inconsistencies between claim dates, service delivery, and patient status. Detecting such inconsistencies provides a strong signal for potential fraudulent billing behavior. The distribution of beneficiary mortality status is illustrated in Fig. 6.

FIGURE 6 | Mortality status distribution.

- **Age Distribution:** The age distribution is heavily concentrated in the 65–85 year range, consistent with Medicare eligibility criteria and expected demographic composition. Age is an important covariate in healthcare utilization modeling, as procedure frequency, chronic disease prevalence, and reimbursement levels tend to increase with age. Moreover, unusually young beneficiaries exhibiting high claim volumes may indicate miscoding or potentially fraudulent submissions. The corresponding age distribution is shown in Fig. 7.
- **Chronic Condition Prevalence:** Chronic illnesses such as ischemic heart disease, diabetes, and heart failure occur with high frequency among beneficiaries, reflecting the clinical characteristics of the Medicare population. These conditions are strongly associated with increased healthcare utilization and higher reimbursement levels, which may be exploited by fraudulent providers through practices such as upcoding or billing for medically unnecessary services. Understanding the prevalence of chronic conditions is therefore essential for distinguishing legitimate high-utilization patterns from anomalous billing behavior. The distribution of chronic conditions in the dataset is shown in Fig. 8.

Overall, the exploratory analysis highlights meaningful demographic and clinical characteristics of the supervised dataset that directly influence fraud modeling. The presence of chronic comorbidities and age-related utilization trends explains the high

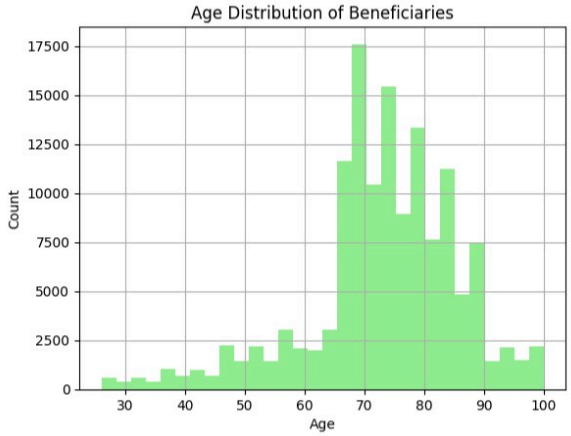


FIGURE 7 | Age distribution of beneficiaries.

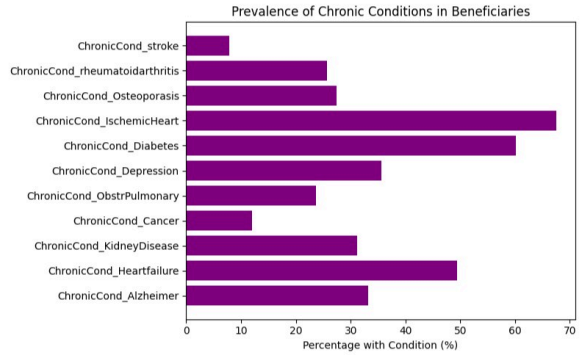


FIGURE 8 | Prevalence of chronic conditions among beneficiaries.

claim volumes. At the same time, mortality status, rare-event frequencies, and class imbalances provide strong indicators of potential fraudulent activity. These observations motivate the use of hybrid modeling strategies that combine supervised learning with auxiliary techniques for handling imbalanced data, demographic fairness considerations, and anomalous claim behavior, forming the basis for the machine learning implementation described in the subsequent subsection.

5.1.4 | Feature Preparation

Following the exploratory analysis, the raw Medicare claims data underwent several transformations to construct a unified, machine-learning-ready feature space. Inpatient, outpatient, and beneficiary tables were merged using `BeneID` and `ProviderID` to produce a consolidated provider-level dataset. Statistical aggregation was performed across claim histories to derive utilization indicators such as average reimbursement, total payment, claim frequency, and beneficiary counts, enabling models to capture both financial and behavioral patterns at the provider level. Categorical variables such as diagnosis codes, procedure codes, and demographic attributes were encoded to ensure compatibility

with downstream learning algorithms. At the same time, monetary and numerical fields were normalized to mitigate scale disparities and improve model stability. Missing values were imputed using distribution-preserving methods to minimize information loss while preserving key signals.

Additional preprocessing steps were applied to improve learning quality and reduce the risk of overfitting. Outlier-handling techniques were used to limit extreme reimbursement values that could disproportionately influence model gradients, while redundant and low-variance features were removed using correlation analysis and variance thresholds. Given the strong class imbalance in the supervised dataset, stratified sampling was used to maintain label proportions across the training and test splits.

5.2 | Fraud Risk Analytics Module

The objective of this module is not to advance the state of machine learning model accuracy or to propose novel classification architectures. Instead, machine learning is integrated into the SHIELD framework as an operational fraud detection mechanism that supports real-world insurance claim processing. The supervised and unsupervised models serve as risk-scoring components that generate fraud likelihood and anomaly indicators, which are then consumed by the blockchain layer for transparent enforcement of decisions. In this sense, machine learning functions as a fraud analytics engine embedded within a larger verification pipeline, rather than as a standalone predictive benchmarking task.

Accordingly, the evaluation of supervised and unsupervised models is framed in terms of fraud-detection effectiveness and operational suitability rather than in terms of pure predictive accuracy benchmarks. Metrics such as false-negative reduction, anomaly coverage, and risk signal interpretability are emphasized because they directly affect insurance operations, audit workload, and financial loss mitigation. This orientation distinguishes the proposed framework from conventional machine learning studies that primarily focus on improving model accuracy on isolated datasets.

5.2.1 | Notation

The notation used in the supervised and unsupervised fraud detection modules is summarized in Table 2. The supervised component operates on labeled provider-level data to produce fraud risk scores, while the unsupervised component operates on unlabeled claim-level data to produce anomaly indicators suitable for patient-side submissions.

5.2.2 | Supervised Fraud Detection

In the proposed system, supervised learning is employed to support insurance-initiated claim submissions by assessing whether a provider exhibits fraudulent billing behavior. Since provider-level fraud labels are available in the supervised dataset, the problem is formulated as a binary classification task in which the objective is to distinguish fraudulent from legitimate providers based on aggregated claim features. To achieve robust predictive performance, a stacking-based hybrid learning architecture is adopted, combining multiple heterogeneous classifiers whose outputs are integrated through a meta-learning layer.

5.2.2.1 | Model Architecture. The supervised model consists of three base learners — Random Forest (RF), Extreme Gradient Boosting (XGBoost), and CatBoost — chosen for their strong performance on structured tabular data and their complementary inductive biases. RF employs bootstrap aggregation of decision trees to reduce variance; XGBoost uses gradient boosting to correct residual errors iteratively; and CatBoost performs ordered boosting with efficient categorical encoding to mitigate target leakage. To aggregate the outputs of these models, a Logistic Regression classifier is used as a meta-learner. Logistic Regression is selected due to its regularization properties and its ability to integrate probability estimates from multiple base learners without introducing significant overfitting.

5.2.2.2 | Training Procedure. To prevent information leakage between the base learners and the meta-learner, an out-of-fold (OOF) stacking strategy with stratified k-fold cross-validation is employed. The training data is partitioned into k folds; for each fold and each base model, training is performed on $k-1$ folds, and probability estimates for the held-out fold are recorded as OOF predictions. Once all folds are processed, the OOF predictions are concatenated to form a stacked training matrix for fitting the meta-learner. After OOF training, each base model is retrained on the full dataset to prepare for deployment. During inference, the meta-learner receives averaged probability estimates from the fully trained base models and outputs a final fraud risk score for the provider. The training logic is summarized in Algorithm 1.

Algorithm 1 Provider-side stacking-based fraud risk classification.

Require: Training data (X, y), base classifiers $\{M_1, M_2, M_3\}$, meta-classifier M_{meta}, number of folds k
Ensure: Provider risk scores $\hat{y} \in [0,1]^n$
Split X into k stratified folds
for each base classifier M_j do
  for each fold $i \in \{1, \ldots, k\}$ do
    Train $M_j^{(i)}$ on $k-1$ folds
    Obtain out-of-fold predictions for held-out fold
  end for
  Concatenate OOF predictions to form column Z_j
end for
Form stacked matrix $Z = [Z_1, Z_2, Z_3]$
Train meta-classifier M_{meta} using (Z, y)
Retrain each M_j on full X and obtain predicted probabilities to form Z^{test}
return $\hat{y} = M_{\text{meta}}(Z^{\text{test}})$

5.2.2.3 | Model Architecture Diagram. The stacking workflow described above is illustrated in Fig. 9, which shows the OOF prediction generation, meta-feature construction, and final inference procedure.

5.2.2.4 | Experimental Results. To evaluate the predictive performance of the proposed model, the supervised dataset was

TABLE 2 | Notation used in the supervised and unsupervised fraud detection components.

Symbol	Description	Appears in
$X \in \mathbb{R}^{n \times d}$	Labeled provider feature matrix	Algorithm 1
$y \in \{0,1\}^n$	Provider fraud labels (1 = fraudulent)	Algorithm 1
$\{M_1, M_2, M_3\}$	Base classifiers in the supervised ensemble	Algorithm 1
M_{meta}	Meta-classifier for stacking integration	Algorithm 1
$Z \in \mathbb{R}^{n \times 3}$	Stacked OOF prediction matrix	Algorithm 1
$\hat{y} \in [0,1]^n$	Provider risk scores (supervised output)	Algorithm 1
$X_u \in \mathbb{R}^{m \times d}$	Unlabeled claim dataset	Algorithm 2
$\{A_1, \ldots, A_k\}$	Unsupervised anomaly detectors	Algorithm 2
$p_j(x) \in \{0,1\}$	Binary anomaly decision by detector A_j	Algorithm 2
$S(x)$	Ensemble anomaly score $S(x) = \frac{1}{k}\sum_{j=1}^{k} p_j(x)$	Algorithm 2
$\hat{y}(x)$	Final anomaly label for sample x	Algorithm 2
τ	Score threshold for anomaly decision	Algorithm 2
k	Number of anomaly detectors ($k = 4$)	Algorithm 2

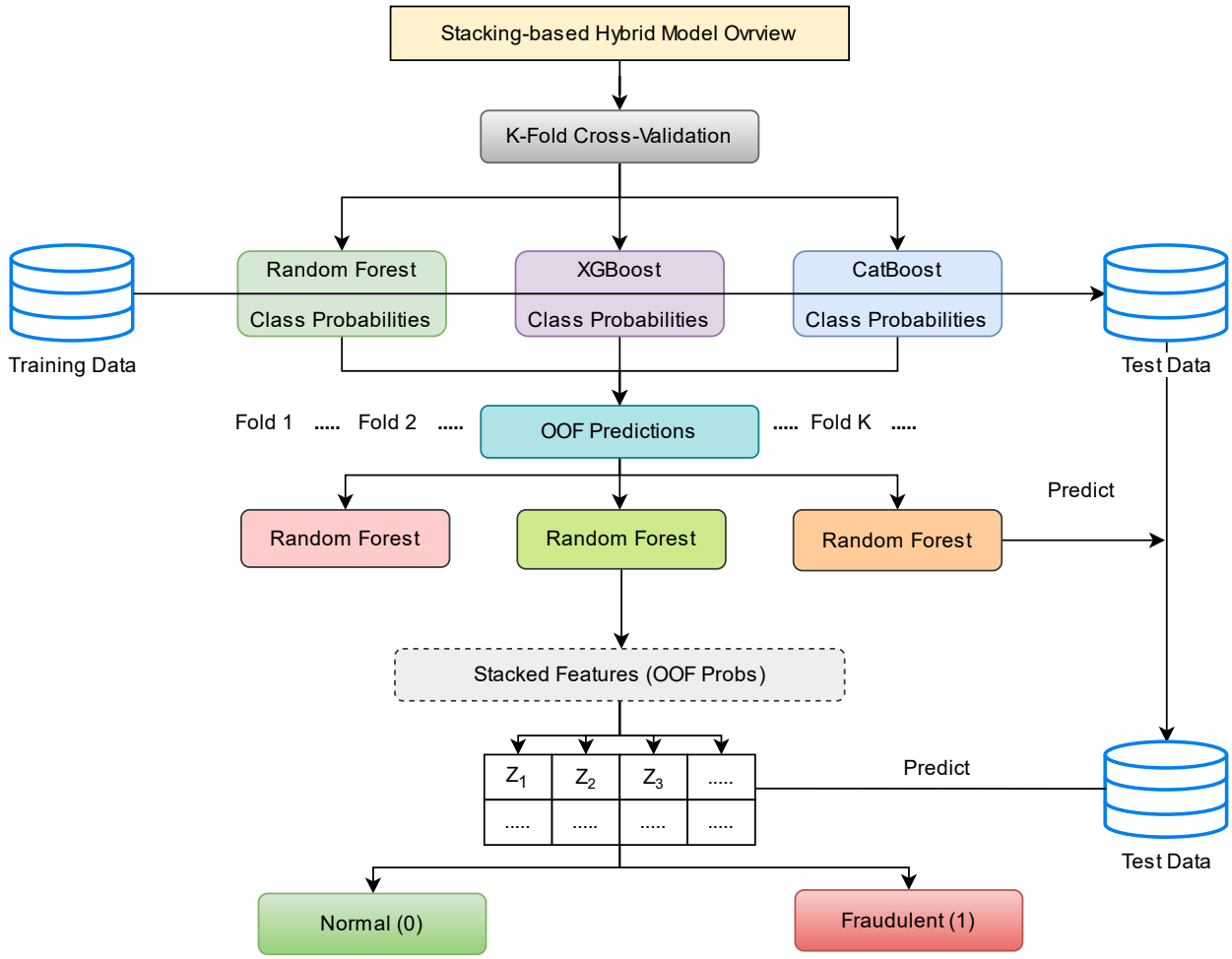


FIGURE 9 | Stacking-based hybrid model architecture for supervised provider fraud classification.

partitioned into training and testing subsets using stratified sampling to preserve class proportions. Table 3 reports the performance of the three base learners, while Table 4 reports the performance of the final stacked model. The stacked model achieves superior performance across accuracy, F1-score, and AUC metrics, confirming that meta-learning improves discrimination between fraudulent and legitimate providers.

5.2.2.5 | Confusion Matrix and ROC Analysis. Fig. 10 presents binary confusion matrices for the four models. The

TABLE 3 | Performance of individual supervised learning models.

Model	Accuracy	F1-Score	ROC-AUC
Random Forest	0.93	0.93	0.9475
XGBoost	0.94	0.94	0.9400
CatBoost	0.93	0.93	0.9385

stacked model exhibits fewer false negatives compared to individual classifiers, which is desirable in fraud detection, where missed fraudulent cases carry higher operational risk than false alarms. The receiver operating characteristic (ROC) curves in Fig. 11 illustrate the discriminative performance of the supervised learning models across varying decision thresholds. All four models achieve high area-under-the-curve (AUC) values, indicating strong capability in separating fraudulent and non-fraudulent providers. Random Forest and CatBoost exhibit stable performance with steep initial rises in true positive rate, reflecting effective early detection of fraudulent cases. XGBoost achieves competitive recall for the fraud class while maintaining a controlled false positive rate, making it suitable for prioritizing suspicious claims. The Logistic Regression meta-learner attains the highest AUC among the evaluated models, demonstrating that the stacking strategy effectively integrates complementary strengths of the base learners. Overall, the ROC analysis confirms that the proposed ensemble approach provides improved discrimination compared to individual classifiers and supports its use for fraud risk assessment in insurance claim verification.

5.2.2.6 | **Feature Interpretation.** To gain insight into the decision process of the ensemble, feature importance scores were extracted from the Random Forest model. As shown in Fig. 12, reimbursement-related variables and claim volume indicators form strong predictors of fraud risk, aligning with known fraud tactics such as inflating reimbursement amounts or increasing procedure frequency.

5.2.2.7 | **Discussion.** Overall, the supervised model demonstrates high accuracy and robust discriminatory performance, confirming that provider-level fraud signals can be effectively captured through stacked ensemble learning. Importantly, the model reduces false negatives relative to single-model baselines, which is advantageous for insurance operations where undetected fraud incurs substantial financial loss. These supervised outputs are subsequently integrated with anomaly signals generated by patient submissions

5.2.3 | Unsupervised Fraud Detection

In scenarios where fraud labels are unavailable, the system uses unsupervised anomaly detection to identify claims that deviate from the dataset's dominant statistical distribution. Rather than predicting predefined categories, the objective is to isolate records exhibiting unusual patterns in reimbursement, frequency, or beneficiary behavior that may correspond to fraudulent activity, data inconsistencies, or rare clinical contexts. Four anomaly detection algorithms are utilized: Isolation Forest, Local Outlier Factor (LOF), One-Class Support Vector Machine (OC-SVM), and a deep Autoencoder. These models are selected because they capture abnormality under different assumptions — distance-based, density-based, margin-based, and reconstruction-based — providing complementary perspectives on what constitutes an outlier.

The ensemble architecture is depicted in Fig. 13, which illustrates the flow from dataset ingestion through anomaly-level decision-making.

5.2.3.1 | **Algorithm: Ensemble Voting and Score-Based Detection.** To improve robustness, outputs from the four detectors were fused via majority voting and anomaly-score aggregation. For each claim x, the binary anomaly decisions $p_j(x) \in \{0, 1\}$ from detector M_j were averaged to form an anomaly score,

$$S(x) = \frac{1}{k} \sum_{j=1}^{k} p_j(x),$$

where k denotes the number of detectors. Claims with $S(x) \geq \tau$ were labelled as anomalous, and those with $S(x) < \tau$ were labelled as normal. The threshold was set to $\tau = 0.5$, consistent with majority voting.

Algorithm 2 Patient-side ensemble anomaly detection using majority voting.

```
Require: Unlabeled dataset X_u, detectors {A_1, ..., A_k},
    threshold τ
Ensure: Anomaly score S(x) and binary label ŷ(x) for
    each x ∈ X_u
for each detector A_j do
    Train A_j on X_u
end for
for each sample x ∈ X_u do
    for each detector A_j do
        Obtain binary decision p_j(x) ∈ {0, 1}
    end for
    Compute ensemble score S(x) = (1/k) Σ_{j=1}^{k} p_j(x)
    if S(x) ≥ τ then
        ŷ(x) = 1                                  ▷ anomaly
    else
        ŷ(x) = 0                                  ▷ normal
    end if
end for
```

5.2.3.2 | **Results and Analysis.** All four detectors identified a comparable number of anomalous claims, as shown in Table 5. Despite the similar counts, the degree of overlap across anomaly sets differs substantially. Table 6 shows pairwise Jaccard similarity values, revealing low inter-model agreement and confirming that different detectors capture distinct anomaly structures.

Anomaly Statistics and Detector Agreement. All four detectors discovered a comparable number of anomalies, as illustrated in Fig. 14, indicating that each method identifies a similar proportion of the dataset as statistically unusual.

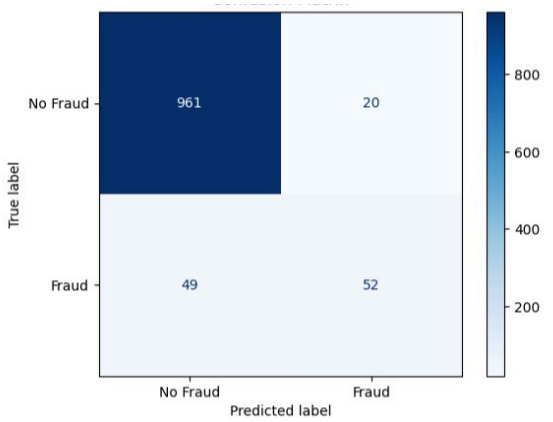
To further study agreement between detectors, the Jaccard similarity between binary anomaly sets was computed:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}.$$

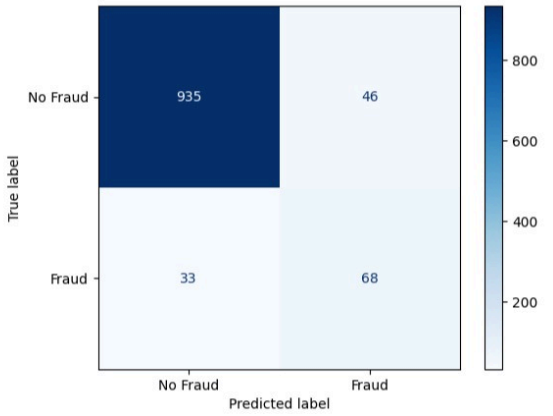
TABLE 4 | Performance of the proposed stacking-based hybrid model.

Model	Accuracy	F1-Score	ROC-AUC
Stacking-based hybrid model (Random Forest, XGBoost, CatBoost, Logistic Regression)	0.95	0.95	0.9557

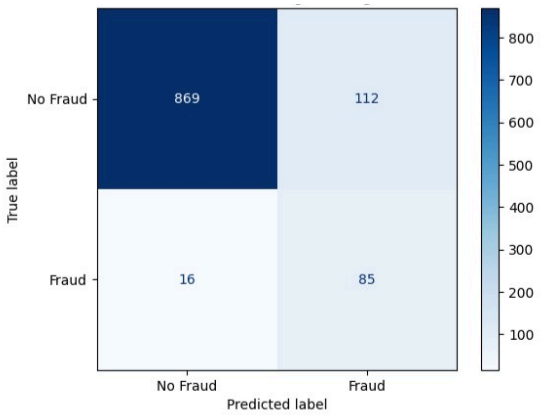
Binary confusion matrix for Random Forest.



Binary confusion matrix for XGBoost.



Binary confusion matrix for CatBoost.



Binary confusion matrix for Logistic Regression.

FIGURE 10 | Binary confusion matrices for supervised learning models: (a) Random Forest, (b) XGBoost, (c) CatBoost, and (d) Logistic Regression.

TABLE 5 | Number of anomalies detected by each unsupervised model.

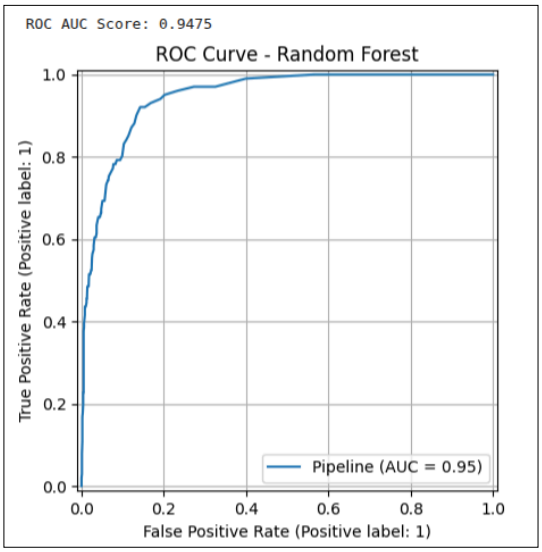
Model	Anomalies Detected
Isolation Forest	2927
Local Outlier Factor	2927
One-Class SVM	2926
Autoencoder	2927

TABLE 6 | Jaccard similarity between anomaly sets across detectors.

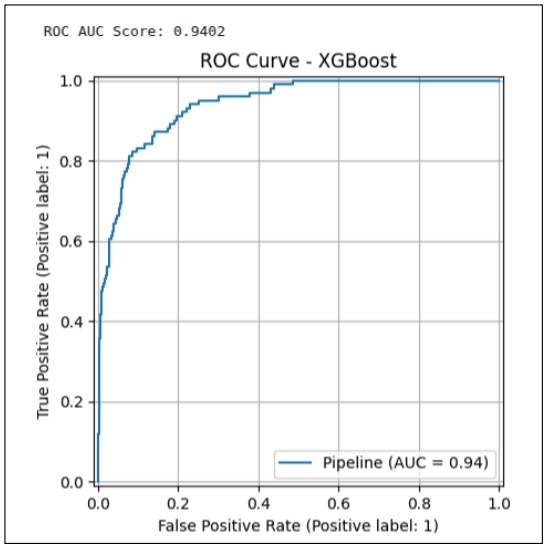
	IF	LOF	OC-SVM	AE
IF	1.000	0.109	0.366	0.493
LOF	0.109	1.000	0.065	0.110
OC-SVM	0.366	0.065	1.000	0.427
AE	0.493	0.110	0.427	1.000

The results indicate modest similarity scores across detectors, suggesting low overlap and implying that different models capture distinct anomaly structures. Notably, Autoencoder exhibits higher agreement with Isolation Forest than with LOF, which reflects their shared tendency to emphasize reconstruction and density-based deviations respectively. As shown in Fig. 15, these patterns highlight complementary behavior among the unsupervised detectors.

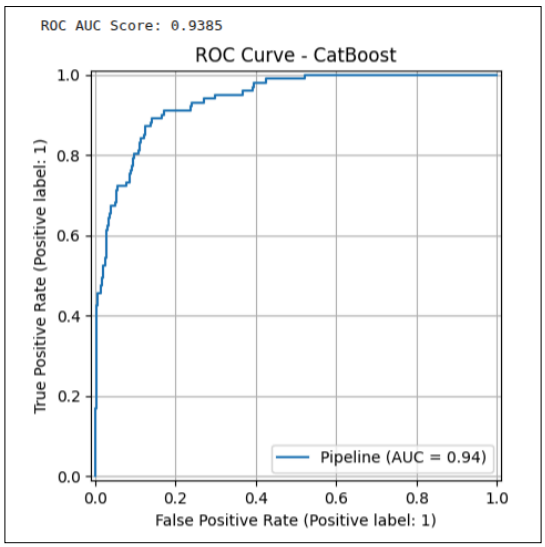
A 2D PCA projection of Isolation Forest anomalies reveals that anomalous claims occupy multiple sparse regions rather than forming a single coherent cluster, which is consistent with heterogeneous fraud typologies observed in insurance domains, as shown in Fig. 16.



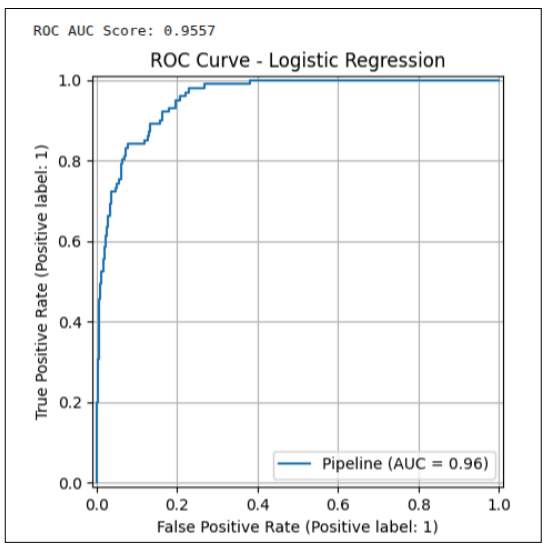
ROC curve for Random Forest.



ROC curve for XGBoost.



ROC curve for CatBoost.



ROC curve for Logistic Regression.

FIGURE 11 | ROC curves of supervised learning models: (a) Random Forest, (b) XGBoost, (c) CatBoost, and (d) Logistic Regression.

5.2.3.3 | **Interpretation..** The unsupervised analysis yields three key observations: (i) anomalies exist and are consistently detected across models, indicating meaningful irregularities in patient-level claim behavior; (ii) low Jaccard similarity confirms that individual detectors identify different subsets of anomalies, motivating ensemble aggregation; and (iii) PCA dispersion suggests multiple underlying anomaly modalities rather than a single fraud mechanism. These properties support the deployment of unsupervised ensemble detection methods for real-world insurance auditing, where labeled fraud data are scarce and expert review capacity is limited.

The unsupervised results complement the supervised insurance submission pipeline by providing additional signals for claim prioritization, expert investigation, and cross-verification. These anomaly-based insights enable the system to flag statistically rare or behaviorally abnormal claim patterns, offering an effective mechanism for ranking suspicious records for downstream manual review or fraud adjudication. This joint architecture allows anomaly signals to be securely propagated, logged, and cross-validated, resulting in a fraud-aware claim processing system that remains transparent, explainable, and resistant to manipulation. In the subsequent section, we describe how these machine learning outputs are integrated with the blockchain-enabled verification layer to establish a secure, auditable, and tamper-resistant decision pathway.

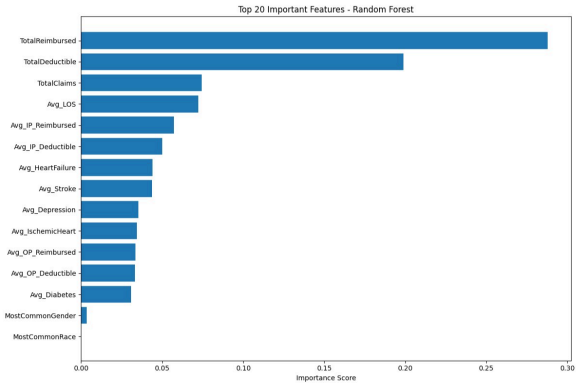


FIGURE 12 | Top twenty most important features identified by Random Forest.

5.3 | Integration with Blockchain System

The supervised and unsupervised machine learning modules operate as an off-chain risk assessment engine that evaluates insurance claims before they are validated on the blockchain network. The integration between the ML components and the blockchain system ensures that fraudulent or anomalous claims are intercepted early, while legitimate claims proceed through the standard verification pipeline without delay.

In the proposed architecture, claim submission is initiated either by the healthcare provider (supervised path) or by the patient (unsupervised path). For provider-initiated submissions, the stacking-based supervised classifier outputs a fraud likelihood score that quantifies the probability of fraudulent billing behavior. For patient-submitted claims without prior fraud labels, the ensemble anomaly detector computes an anomaly score reflecting the degree of deviation from normal claim behavior.

Both outputs are aggregated into a unified *risk score* that is forwarded to the smart contract layer for on-chain enforcement. The blockchain acts as a trust-preserving coordination mechanism that records claims, verifies risk-assessment outcomes, and ensures that no party can alter or repudiate decision traces. Smart contracts encode policy rules such as:

- **Auto-approval** for low-risk claims,
- **Manual review** for medium-risk or borderline claims,
- **Auto-rejection or audit** for high-risk claims.

Storing ML inference results off-chain and only committing decisions and metadata on-chain reduces transaction overhead and preserves privacy, while still maintaining auditability. The blockchain ledger thus serves as the final decision authority that ensures transparent, tamper-proof, and non-repudiable handling of insurance claims. This integration bridges data-driven risk prediction with decentralized enforcement and provides a scalable pathway toward secure and fraud-resistant insurance processing.

6 | Security Analysis

This subsection analyzes how the proposed hybrid blockchain–machine learning framework achieves security and privacy objectives in the context of health insurance claim processing. The system incorporates multiple mutually reinforcing mechanisms, including anonymous credentialing, permissioned blockchain execution, smart contract rule enforcement, off-chain encrypted document storage, and adaptive machine learning risk modeling. Together, these components mitigate a wide range of threats posed by malicious users, compromised providers, and external adversaries attempting to manipulate claim workflows for financial gain.

6.0.0.1 | Integrity and Tamper Resistance. Hyperledger Fabric ensures that all claim submissions, policy verification outcomes, and fraud risk decisions are stored on an immutable ledger. Because Fabric employs endorsement policies and replicated consensus across multiple authorized nodes, no participant can alter or delete existing records without being detected. Moreover, off-chain documents stored in IPFS are content-addressed, meaning that any modification to a medical receipt, invoice, or diagnostic report results in a different hash and fails validation during on-chain credential verification. These properties collectively prevent post-hoc manipulation of claim data, document substitution, and retrospective fraud attempts.

6.0.0.2 | Authentication, Authorization, and Access Control. The security of the proposed model is strengthened by Fabric's permissioned trust architecture. Node enrollment, certificate issuance, and role-based access control are managed by the Membership Service Provider (MSP), ensuring that only authenticated entities can submit claims, invoke smart contracts, or read ledger updates. Fabric's channels further restrict visibility of transactions to authorized participants, preventing unauthorized disclosure or eavesdropping. This eliminates common attack vectors found in public blockchains, such as Sybil attacks or anonymous write operations.

6.0.0.3 | Identity Anonymization and Privacy Preservation. The framework prevents the exposure of personally identifiable information (PII) by substituting real identities with anonymous credentials derived through one-way SHA-256 hashing. Since only hashed claims and document references are written to the ledger, no patient or provider identifiers appear in plaintext on-chain. Sensitive documents are stored off-chain and encrypted before being pushed to IPFS, preventing both network adversaries and blockchain peers from accessing medical content. Furthermore, the machine learning engine receives only anonymized feature vectors and does not require raw clinical documentation, thereby minimizing the risk of reconstruction or model inversion attacks. These design choices ensure compliance with privacy requirements common in healthcare contexts.

6.0.0.4 | Rule-Based Fraud Prevention and Deterministic Verification. Smart contracts serve as the first line of defense against trivial or highly structured fraud attempts. Prior to invoking machine learning models, smart contracts validate temporal consistency (e.g., service-date ordering), policy eligibility, service code correctness, and duplicate claim submissions. Claims containing reused credentials, expired policy tokens, or inconsistent

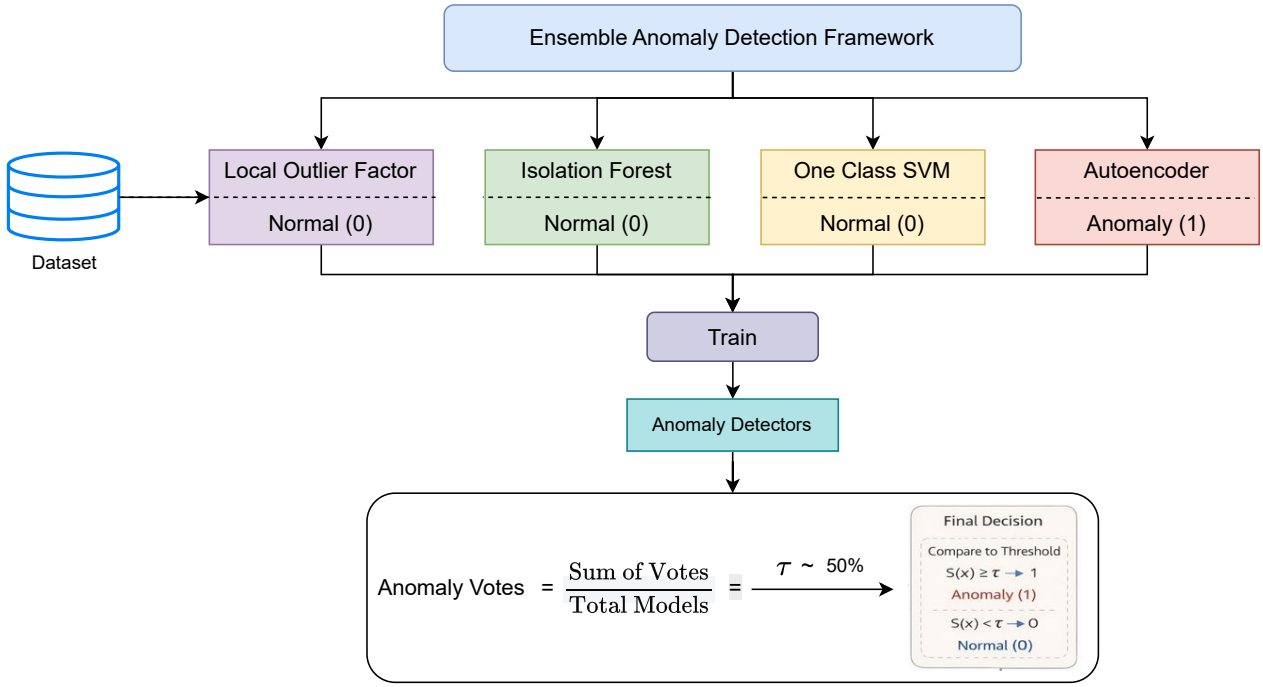


FIGURE 13 | Ensemble anomaly detection framework integrating Isolation Forest, LOF, OC-SVM, and Autoencoder models.

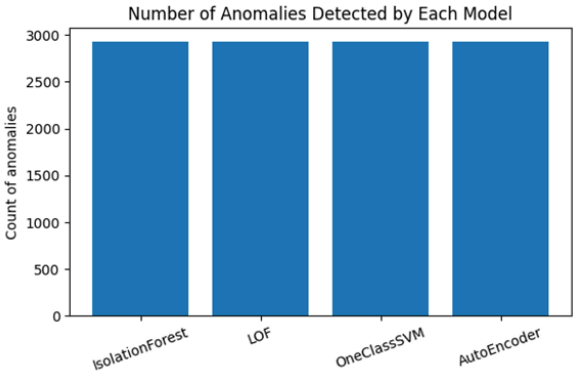


FIGURE 14 | Number of anomalies detected by each unsupervised model.

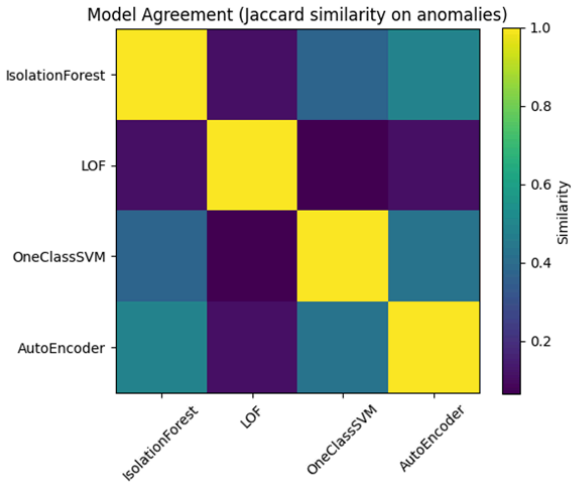


FIGURE 15 | Pairwise Jaccard similarity of anomaly sets across detectors.

timestamps are rejected deterministically on-chain. This prevents fraudulent claims from reaching later processing stages and reduces the ML workload by filtering out predictable rule-violating cases.

6.0.0.5 | Resilience Against Behavioral and Statistical Manipulation. In many insurance systems, adversaries exploit the limitations of static rule-based validation by gradually adapting their billing behaviors. The proposed architecture mitigates these attempts by integrating supervised and unsupervised ML models that analyze behavioral and statistical irregularities. Supervised models detect provider-level fraud patterns already observed in labeled datasets (e.g., abnormal reimbursement rates), while unsupervised anomaly detectors identify emerging fraud tactics or rare deviations in patient-level submissions that lack labels. This layered analytics pipeline increases adversarial effort by requiring both policy compliance and statistical plausibility to evade detection.

6.0.0.6 | Adversarial Robustness in Machine Learning. Machine learning components introduce the possibility of evasion attacks, wherein adversaries intentionally craft claims to mislead a classifier. The use of model ensembles reduces vulnerability to such attacks by combining heterogeneous

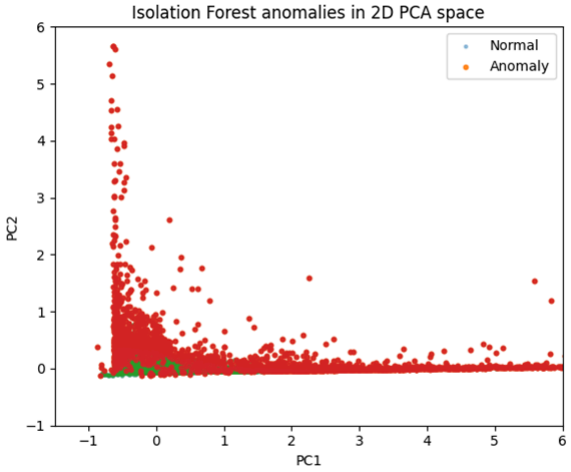


FIGURE 16 | Isolation Forest anomalies projected into 2D PCA space.

decision boundaries. Tree-based ensembles (Random Forest, XGBoost, CatBoost) are complemented by Isolation Forest anomaly scoring, making it difficult for an attacker to manipulate feature distributions so that all models simultaneously output benign predictions. Additionally, the off-chain deployment of ML models allows periodic retraining and defense updates without modifying blockchain consensus, enabling continuous adaptation to fraud evolution.

6.0.0.7 | **Auditability and Non-Repudiation.** All claim processing events—including rule validation outcomes, ML-based decisions, and IPFS document hash references—are recorded in append-only ledger entries. Since every write operation is cryptographically signed, participants cannot repudiate submitted claims, policy decisions, or fraud assessments. This auditable trail is particularly valuable for forensic analysis and regulatory compliance, enabling insurers and auditors to investigate disputes and uncover coordinated fraud schemes involving multiple actors.

6.0.0.8 | **Discussion.** Overall, the proposed system achieves a defense-in-depth security profile through the interplay of blockchain immutability, controlled participation, privacy-preserving credential management, and adaptive machine learning analytics. Rule-based verification prevents basic inconsistencies and document manipulation, while ML models capture evolving behavioral fraud strategies that exceed the expressive capacity of deterministic smart contracts. Privacy is preserved throughout by limiting on-chain disclosures and decoupling document storage from ledger state. Although adversarial ML evasion remains an inherent challenge, the ensemble architecture and off-chain retraining capability provide flexible mechanisms for improving resilience over time.

7 | Performance Analysis

This section evaluates the supervised fraud-detection module of the proposed SHIELD framework and compares its performance with that of recent studies on health insurance fraud detection. As noted, the objective of this work is not to maximize classifier benchmarking accuracy as an isolated academic goal, but to operationalize machine learning as a fraud risk scoring component that complements blockchain-based verification. Accordingly, performance metrics are assessed based on their practical implications for reducing undetected fraud, improving detection sensitivity, and supporting downstream claim adjudication workflows.

To contextualize performance, Table 7 presents a comparative analysis between SHIELD and four recent studies spanning 2023–2025 that employed supervised learning for fraud detection. The comparison focuses on accuracy and F1-score, which are commonly reported metrics for fraud detection tasks given class imbalance and the operational costs of misclassification.

Here, two notable observations emerge. First, several recent works, such as [43], achieve strong performance using traditional supervised models, indicating that labeled fraud datasets can be effectively modeled using tree-based classifiers. Second, the work of [45] highlights the comparative advantage of gradient boosting methods over linear models for capturing non-linear feature interactions in healthcare claim data.

Compared to existing approaches, the supervised component of SHIELD exhibits competitive performance across all evaluated metrics, achieving 93–94% accuracy and 93–94% F1-score across Random Forest, XGBoost, and CatBoost models. The improvement in F1-score over several earlier models is particularly relevant, as F1-score penalizes both false negatives (missed fraud) and false positives (false alerts)-both of which carry operational costs for insurance providers.

It is important to clarify that the objective of this work is not to advance benchmark performance or to claim state-of-the-art accuracy from a machine learning research perspective. Instead, the supervised model serves as an operational fraud scoring module that integrates into a larger blockchain-based insurance workflow. The classifier provides probabilistic advisory risk scores that complement deterministic on-chain rule-based checks, enabling insurers to prioritize claims for manual review better and reduce undetected fraudulent submissions.

The utility of the supervised classifier is further strengthened by the presence of an unsupervised anomaly detection module within the SHIELD framework. Whereas supervised learning operates on labeled fraud data—typically representing known patterns—unsupervised models detect statistically irregular behaviors that may indicate emerging or previously unseen fraudulent schemes. This dual mechanism increases overall fraud coverage, particularly in realistic settings where fraud labels are sparse, delayed, or unavailable.

In summary, the performance evaluation indicates that the proposed supervised ensemble classifier achieves competitive accuracy and F1 performance relative to recent literature while satisfying the operational requirements of a real-world fraud detection system. When combined with rule-based verification and unsupervised anomaly detection, the architecture provides stronger fraud resilience than purely rule-based or purely ML-based solutions, enabling more reliable and privacy-preserving health insurance claim adjudication.

TABLE 7 | Comparison of supervised fraud detection model performance.

Reference Paper	Model Used	Accuracy	F1-Score
Nalluri et al. [23] (2023)	Decision Trees	0.7683	0.78
	Random Forest (RF)	0.75	0.78
	Support Vector Machines	0.75	0.77
	Multilayer Perceptron	0.77	0.81
Al-Ghazi et al. [43] (2024)	Random Forest	0.87	0.89
	Logistic Regression	0.90	0.90
Jena et al. [44] (2024)	Support Vector Machines	0.9334	0.5170
Fourkiotis et al. [45] (2025)	CatBoost	0.93	0.60
	SVM	0.91	0.59
	XGBoost	0.92	0.59
	Random Forest	0.92	0.58
	Logistic Regression	0.92	0.59
	LightGBM	0.92	0.57
Proposed Model	**Random Forest**	**0.93**	**0.93**
	XGBoost	**0.94**	**0.94**
	CatBoost	**0.93**	**0.93**

8 | Conclusion

Health insurance fraud continues to impose significant financial and operational burdens on healthcare systems worldwide, particularly as digital claim submission processes increase the volume and complexity of transactions. Traditional centralized and rule-based fraud detection mechanisms are effective at identifying straightforward inconsistencies but remain insufficient for detecting sophisticated, multi-entity, and statistically rare fraud behaviors. To address these limitations, this work introduced SHIELD, a hybrid privacy-preserving fraud detection framework that integrates permissioned blockchain technology with supervised and unsupervised machine learning–based risk analytics.

The proposed system maintains all the security and auditability guarantees of blockchain-enabled claim management while extending its detection capabilities through a machine learning risk engine that supports fraud scoring for both labeled and unlabeled claim submissions. Deterministic smart contract validation ensures policy eligibility, credential integrity, temporal consistency, and duplicate claim detection, while the supervised ensemble classifier and anomaly detection modules enhance detection sensitivity against complex and emerging fraud behaviors. Experimental results on Medicare datasets demonstrated that supervised models such as Random Forest, XGBoost, and CatBoost achieve strong predictive performance, while the unsupervised ensemble effectively identifies anomalous patterns that lack prior fraud labels. Together, these components form a multi-layered defense pipeline that improves fraud coverage and reduces reliance on manual auditing.

Beyond its technical contribution, SHIELD offers several operational benefits for insurers. First, the event-driven integration between blockchain and machine learning allows the fraud scoring engine to operate off-chain without exposing sensitive data or incurring excessive ledger overhead. Second, the cryptographic credentialing and IPFS-based storage architecture ensure that medical documents and identities remain protected throughout the claim lifecycle. Third, the unified risk scoring mechanism enables claims to be prioritized for approval, rejection, or manual review, allowing insurers to allocate investigative resources more efficiently and reduce false-negative fraud leakage.

Although the prototype implementation demonstrates the feasibility and advantages of the hybrid approach, several opportunities for further enhancement remain. Future work will investigate online and real-time deployment scenarios, including streaming-based claim ingestion and continuous retraining of supervised and unsupervised models. Adversarial robustness techniques will be explored to mitigate model evasion and poisoning attacks, which represent growing threats in ML-assisted decision systems. In addition, federated and privacy-enhancing learning paradigms may enable cross-institutional fraud intelligence sharing without revealing sensitive data, thereby improving detection coverage across regional or national insurance networks. Integrating explainable machine learning mechanisms into the decision pipeline also represents a valuable direction to support regulatory compliance and human-in-the-loop adjudication.

In summary, SHIELD demonstrates that combining blockchain-based deterministic validation with machine learning–based probabilistic risk assessment provides a practical, scalable, and privacy-preserving solution for modern health insurance fraud detection. The hybrid design improves transparency, strengthens security guarantees, and enhances fraud detection coverage, making it suitable for adoption in emerging digital health insurance infrastructures.

Acknowledgments

This research was supported by the Institute for Advanced Research (IAR), United International University, and the Office of Research (OR), North South University, under Grant UIU-IAR-02-2023-SE-41. The authors would like to thank both institutions for providing research facilities, computational resources, and technical support necessary for conducting this study. The authors also acknowledge the publicly available Medicare datasets and supporting research resources that enabled the experimental evaluation of the proposed framework.

Financial Disclosure

None reported.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Anli du Preez, Sanmitra Bhattacharya, Peter Beling,, and Edward Bowen. "Fraud detection in healthcare claims using machine learning: A systematic review." *Artificial Intelligence in Medicine*160 (2025): 103061.
2. Zain Hamid, Fatima Khalique, Saba Mahmood, Ali Daud, Amal Bukhari,, and Bader Alshemaimri. "Healthcare insurance fraud detection using data mining." *BMC Medical Informatics and Decision Making* 24, no. 1 (2024): 112.
3. Zeyu Wang, Xiaofang Chen, Yiwei Wu, Linke Jiang, Shiming Lin,, and Gang Qiu. "A robust and interpretable ensemble machine learning model for predicting healthcare insurance fraud." *Scientific Reports* 15, no. 1 (2025): 218.
4. Ophelie Lavoie-Gagne, Joshua J Woo, Riley J Williams III, Benedict U Nwachukwu, Kyle N Kunze,, and Prem N Ramkumar. "Artificial intelligence as a tool to mitigate administrative burden, optimize billing, reduce insurance-and credentialing-related expenses, and improve quality assurance within health care systems." . *Arthroscopy: The Journal of Arthroscopic & Related Surgery*.
5. Sunayana Das, Tushar Kanta Samal, Bhabendu Kumar Mohanta,, and N Nasurudeen Ahamed. "Blockchain-Driven Solutions for Healthcare Insurance Fraud Detection and Prevention." *Security and Privacy* 9, no. 1 (2026): e70149.
6. Kelly Anderson. "Blockchain-Assisted Secure Data Conversion for Medicare and Medicaid Claims Processing." .
7. Tsung-Chih Hsiao, Tzer-Long Chen, Shu-Chen Chang,, and Tsui-Ping Chang. "Research on smart contracts for Omnipresent AI in commercial health insurance." *Enterprise Information Systems* 19, no. 3-4 (2025): 2442405.
8. Minh Quang Nguyen. "Developing an Expert System for Healthcare Claims Validation Using Knowledge Representation Techniques." *Transactions on Computational Science, Mathematical Modeling, and Simulation Techniques* 15, no. 4 (2025): 1–13.
9. Kamran Razzaq, and Mahmood Shah. "Next-generation machine learning in healthcare fraud detection: Current trends, challenges, and future research directions." *Information* 16, no. 9 (2025): 730.
10. Hannes De Meulemeester, Frank De Smet, Johan van Dorst, Elise Derroitte,, and Bart De Moor. "Explainable unsupervised anomaly detection for healthcare insurance data." *BMC Medical Informatics and Decision Making* 25, no. 1 (2025): 14.
11. Irum Matloob, Shoab Khan, Rukaiya Rukaiya, Hessa Alfraihi,, and Javed Ali Khan. "Healthcare fraud detection using adaptive learning and deep learning techniques." *Evolving Systems* 16, no. 2 (2025): 72.
12. Jay P Bae, David R Nelson, Kristina S Boye,, and Kieren J Mather. "Prevalence of complications and comorbidities associated with obesity: a health insurance claims analysis." *BMC Public Health* 25, no. 1 (2025): 273.
13. Shahram Attarian, Jean-Philippe Camdessanché, Andoni Echaniz-Laguna, Mariana Ciumas, Cécile Blein, Benjamin Grenier,, and Guilhem Solé. "Tracking myasthenia gravis severity over time: Insights from the French health insurance claims database." *European Journal of Neurology* 32, no. 1 (2025): e16518.
14. Shruti Jadon, HS Kumar, Bhuvan Vijay Kumar, Shaik Adeeba Saher, SM Chandana,, and Prasad B Honnavalli. "A comprehensive survey on record management system using blockchain." *Cluster Computing* 29, no. 2 (2026): 98.
15. Ruba Islayem, Senay Gebreab, Walaa AlKhader, Ahmad Musamih, Khaled Salah, Raja Jayaraman,, and Muhammad Khurram Khan. "Using large language models for enhanced fraud analysis and detection in blockchain based health insurance claims." *Scientific Reports* 15, no. 1 (2025): 29763.
16. Narendra Dewangan, and Preeti Chandrakar. "Patient-centric information management in blockchain and interplanetary storage." *Journal of Ambient Intelligence and Humanized Computing* 16, no. 1 (2025): 85–96.
17. Ahmad Al Doulat, Oluwayemisi Elizabeth Ayo-Bali,, and Shehenaz Shaik. 2025. "Fraud Detection in Insurance Claims Using Supervised Machine Learning Models." In *2025 International Conference on Smart Applications, Communications and Networking (SmartNets)*, 1–7. IEEE.
18. Ethan D Curtis, Preston Billion-Polak, Taghi M Khoshgoftaar,, and Borko Furht. "A review of distinct machine learning classifiers for healthcare fraud detection." *Journal of Big Data* 12, no. 1 (2025): 1–23.
19. Md Mazharul Islam, Mubasshir Ahmed, Rajesh Palit, Mohammad Shahriar Rahman,, and Salekul Islam. "Fraud Detection in Privacy Preserving Health Insurance System Using Blockchain Technology." *Engineering Reports* 7, no. 8 (2025): e70315.
20. Andreas Bayerstadler, Linda van Dijk,, and Fabian Winter. "Bayesian multinomial latent variable modeling for fraud and abuse detection in health insurance." *Insurance: Mathematics and Economics*71 (2016): 244–252.
21. Conghai Zhang, Xinyao Xiao,, and Chao Wu. "Medical Fraud and Abuse Detection System Based on Machine Learning." *International Journal of Environmental Research and Public Health* 17, no. 19 (2020): 7265.
22. Nishamathi Kumaraswamy, Mia K Markey, Jamie C Barner,, and Karen Rascati. "Feature engineering to detect fraud using healthcare claims data." *Expert Systems with Applications*210 (2022): 118433.
23. Venkateswarlu Nalluri, Jing-Rong Chang, Long-Sheng Chen,, and Jia-Chuan Chen. "Building prediction models and discovering important factors of health insurance fraud using machine learning methods." *Journal of Ambient Intelligence and Humanized Computing* 14, no. 7 (2023): 9607–9619.
24. John T Hancock, Richard A Bauder, Huanjing Wang,, and Taghi M Khoshgoftaar. "Explainable machine learning models for Medicare fraud detection." *Journal of Big Data* 10, no. 1 (2023): 154.
25. Lavanya Settipalli, and GR Gangadharan. "WMTDBC: An unsupervised multivariate analysis model for fraud detection in health insurance claims." *Expert Systems with Applications*215 (2023): 119259.
26. Jinlong Du, and Benhai Yu. 2023. "Application of Isolation Forest Algorithm in Fraud Detection of Medical Insurance Big Data." In *2023 8th International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*, 504–509. IEEE.
27. Raihan Adam Handoyo Winarso, Yusril Falih Izzaddien, Hartawan Bahari Mulyadi,, and Diana Purwitasari. 2025. "Anomaly Detection in Health Insurance Cost using Ensemble Models for Claim Validation." In *2025 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, 248–253. IEEE.
28. Qingyang He, Qi Ding, Conghui Zheng, Li Pan, Ning Liu,, and Wensheng Li. "A Data-Driven Intelligent Supervision System for Generating High-Risk Organized Fraud Clues in Medical Insurance Funds." *Electronics* 14, no. 16 (2025): 3268.
29. Muhammad Arifuddin Aljufri, Shabira Widyadhari, Anani Asmani, Ach Muhyil Umam,, and Diana Purwitasari. 2025. "Domain-Knowledge Based Feature Engineering in Fraud Detection Using Health Administrative Claims." In *2025 International Conference on Smart Computing, IoT and Machine Learning (SIML)*, 1–6. IEEE.
30. Junwei Feng, Mingze Li, Yiming Feng, Yiran Xin, Yuhang Zhao, Yongxin Jiang, Yixiao Zhang,, and Fayuan Li. 2024. "Medical Insurance Fraud Risk Monitoring and Identification Model Based on Feature Selection and Machine Learning." In *2024 3rd International Conference on Cloud Computing, Big Data Application and Software Engineering (CBASE)*, 571–574. IEEE.

31. Mohammed A Mohammed, Manel Boujelben,, and Mohamed Abid. "A Novel Approach for Fraud Detection in Blockchain-Based Healthcare Networks Using Machine Learning." *Future Internet* 15, no. 8 (2023): 250.
32. Rima Kaafarani, Leila Ismail,, and Oussama Zahwe. "Automatic Recommender System of Development Platforms for Smart Contract–Based Health Care Insurance Fraud Detection Solutions: Taxonomy and Performance Evaluation." *Journal of Medical Internet Research*26 (2024): e50730.
33. Khyati Kapadiya, Fenil Ramoliya, Keyaba Gohil, Usha Patel, Rajesh Gupta, Sudeep Tanwar, Joel JPC Rodrigues, Fayez Alqahtani,, and Amr Tolba. "Blockchain-assisted healthcare insurance fraud detection framework using ensemble learning." *Computers and Electrical Engineering*122 (2025): 109898.
34. Eman Nabrawi, and Abdullah Alanazi. "Fraud Detection in Healthcare Insurance Claims Using Machine Learning." *Risks* 11, no. 9 (2023): 160.
35. K Veena, S Supriya, GM Karpura Dheepan, et al. 2023. "Predicting Health Insurance Claim Frauds Using Supervised Machine Learning Technique." In *2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)*, 1–7. IEEE.
36. Chengamma Chitteti, Mopuri Yamuna, Mattam Srinath, Chakali Govardhan,, and Alavalapati Vignatha. 2025. "Healthcare Insurance Fraud Detection Using Machine Learning." In *2025 8th International Conference on Trends in Electronics and Informatics (ICOEI)*, 968–973. IEEE.
37. Syarifah Diana Permai, and Kevin Herdianto. "Prediction of Health Insurance Claims Using Logistic Regression and XGBoost Methods." *Procedia Computer Science*227 (2023): 1012–1019.
38. Ruhul Quddus Majumder 2025. "Designing an Intelligent Fraud Detection System for Healthcare Insurance Claims Using a Machine Learning Approach." In *2025 Global Conference in Emerging Technology (GINOTECH)*, 1–6. IEEE.
39. P Ashok, and Abhijit Sambhaji Durge. 2025. "Fraud Detection and Prevention in Healthcare Insurance Claims Using Machine Learning Regression Models." In *2025 International Conference on Data Science and Business Systems (ICDSBS)*, 1–7. IEEE.
40. Rohit Anand Gupta 2018. "Healthcare Provider Fraud Detection Analysis." , Version 1. Retrieved November 19, 2025 from Kaggle, https://www.kaggle.com/datasets/rohitrox/healthcare-provider-fraud-detection-analysis.
41. Steve Drasco 2024, August). "Medicare Physician Provider and Service Resampled." , Version 1. Retrieved November 19, 2025 from Kaggle, https://www.kaggle.com/datasets/sdrasco/medicare-physician-provider-and-service-resampled.
42. Leandre Nash 2024, July). "Enhanced Health Insurance Claims Dataset." , Version 1. Retrieved November 19, 2025 from Kaggle, https://www.kaggle.com/datasets/leandrenash/enhanced-health-insurance-claims-dataset.
43. Muhammad Kent Al-Ghazi, Ryan Bertrand, Muhammad Dzul Qarrnayn Destra, Alexander Agung Santoso Gunawan,, and Karli Eka Setiawan. 2024. "Classification of Health Insurance Fraud Risk with Machine Learning." In *2024 International Conference on Information Technology Research and Innovation (ICITRI)*, 24–29. IEEE.
44. Sanjay Kumar Jena, Brajesh Kumar, Barunaditya Mohanty, Ayush Singhal,, and Ram Chandra Barik. "An advanced blockchain-based hyperledger fabric solution for tracing fraudulent claims in the healthcare industry." *Decision Analytics Journal*10 (2024): 100411.
45. Konstantinos P Fourkiotis, and Athanasios Tsadiras. "Future internet applications in healthcare: Big data-driven fraud detection with machine learning." *Future Internet* 17, no. 10 (2025): 460.