Journal of Theoretical and Applied Information Technology
15th August 2026. Vol.104. No.15

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MEDGUARD: A GRAPH-AUGMENTED, FEDERATED AND EXPLAINABLE AI FRAMEWORK ON BLOCKCHAIN FOR HEALTHCARE INSURANCE FRAUD DETECTION

SWAPNA DONEPUDI[1*], NIRMALA DEVI K[2], K VENU GOPAL[3], PATHAPATI SAROJA[4], SWATHI VODDI[5], PRASAD DEVARASETTY[6], DESHINTA ARROVA DEVI[7]

[1*]Associate Professor, Department of CSE, Prasad V Potluri Siddhartha Institute of Technology, India
[2]Professor, Department of ECE ,Saveetha Engineering College, Chennai, India
[3]Professor, Department of CSIT , Lakireddy Bali Reddy College of Engineering, Mylavaram,India.
[4]Assistant Professor, Department of CSE ,SRKR Engineering College , Bhimavaram,India
[5]Assistant Professor, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, India.
[6]Professor, Department of CSE ,DVR & Dr HS MIC College of Technology,Kanchikacherla,India
[7]Professor, Center for Data Science and Sustainable Technologies, INTI International University, Nilai, Malaysia
E-mail: dswapna@pvpsiddhartha.ac.in

ABSTRACT

Healthcare insurance remains burdened by claim fraud, slow manual adjudication, and limited trust in centralized data handling. This paper presents an updated MedGuard architecture that moves beyond tabular ensemble classifiers to a graph-augmented, federated, and explainable fraud-detection pipeline anchored on a permissioned blockchain. Claims are represented as a patient–provider network; centrality and community-detection features extracted from this graph are fused with claim-level features and fed into a hybrid SVM + Gradient-Boosted-Trees classifier. On a 5,000-record synthetic claims benchmark built for this study, the graph-augmented model reaches 93.6% accuracy, 75.7% precision, 70.7% recall, and an AUC of 0.885 which shows a substantial improvement over the same hybrid classifier without graph features (85.0% accuracy, AUC 0.782) and over standalone Logistic Regression, Decision Tree, and Random Forest baselines (71.5–87.7% accuracy). A federated-averaging simulation across three synthetic hospital clients reaches 73.1% accuracy without any client sharing raw claim records, closely tracking a centralized model trained on the pooled data (71.4%), supporting federated learning as a viable privacy-preserving alternative. Claim cost estimation via a Random Forest regressor achieves $R^2 = 0.810$ (RMSE ≈ ₹14,742). Per-claim Shapley-value explanations are computed directly to justify individual fraud flags, and a zero-knowledge-proof verification step is proposed to let smart contracts confirm claim eligibility without exposing medical details on-chain. Verified claims and documents continue to be anchored via smart contracts on a permissioned blockchain, with supporting files encrypted and stored on IPFS.

Keywords: *Blockchain, Healthcare Data Security, Graph-Based Fraud Detection, Federated Learning, Explainable AI, Process Innovation, Financial Access, IPFS, Zero-Knowledge Proofs*

1. INTRODUCTION

While healthcare insurance offers crucial financial coverage for medical costs, the systems that deliver this insurance coverage are still not perfect. Many claims contain fraudulent elements (misrepresentations) that lead to payouts and detecting fraudulent elements is difficult due to the constantly changing nature of fraud and limited data on which to base claims, and the highly imbalanced nature of the fraud label [1], [14]. In the traditional claim lifecycle, most claims are submitted and paid manually which slows down the reimbursement process and provides little visibility for patients to see the status of their claims. Records stored in centralized databases are single points of failure, and an attacker could have access to sensitive information from the healthcare and financial data of many policyholders, leading to a significant research effort in decentralized management of patient-controlled healthcare data [15]. Single-model tabular fraud classifiers, used in previous versions of this system as well as in most of the existing fraud analytics literature [2], [16], [26], classify each claim individually, and thus are unable to capture the coordinated fraud that can

occur in a provider network. The main focus of this paper is on three recent trends in fraud-analytics and privacy-preserving ML research, where MedGuard is re-positioned: (i) graph-based feature augmentation, which captures relationships between patients and providers, and finds collusive fraud rings that per-claim features cannot detect; (ii) federated learning, where multiple hospitals and insurers can work together to train a fraud model, iteratively averaging parameters without ever sharing raw patient records; and (iii) explainable AI, which generates a per-claim justification for every fraud flag, based on Shapley-value attribution [34] instead of a black-box score. They complement the system's current blockchain and IPFS layers, which extend the concepts of distributed ledger, smart contract, and content-addressed-storage that were introduced by Nakamoto [18], Buterin [19] and Benet [8, 20]. Each of these directions, taken individually, has already received a lot of research: blockchain-based management of healthcare data has been well explored [15], [22]; smart-contract-based insurance automation has been tested on public chains [6], [23], [28]; and integrity and scalability of IPFS-backed medical-document storage has been analyzed [13], [17], [27]. A wide range of ensemble classifiers and deep-learning classifiers has been benchmarked in the context of insurance fraud [16], [24], [26] and class-imbalance methods like SMOTE [35] are common in the highly skewed label distributions of fraud data. The relatively less explored is how these features can be integrated into a single architecture for graph-augmented fraud scoring, federated multi-institution training, per-claim explainability, and blockchain-anchored, privacy-preserving verification.

The goals of this work are to:

(i) build a patient–provider claim graph and extract features on the graph that improve fraud discrimination over just tabular claim attributes;

(ii) Simulate federated training across multiple claim-originating institutions, and compare against centralized training.

(iii) Generate per-claim Shapley value explanations for flagged claims.

(iv) Anchor claim decisions on a permissioned blockchain with a zero-knowledge eligibility-verification step.

(v) Store supporting medical documents on IPFS, which detaches them from any central server.

These objectives build on classic ensemble baselines such as Random Forests [21] and prepare the ground for a fully learned graph representation such as GCNs [37] .

2. LITERATURE SURVEY

2.1 Artificial Intelligence in Insurance Fraud Detection

The initial fraud detection systems were narrowly targeted and had poor generalization capabilities on novel fraud patterns [1]. Supervised models using labelled historical claims showed better adaptability, with an accuracy of 75-90% based on algorithm and dataset [2], [14] and more recently deep learning has been used on larger and more complex insurance claims data sets [3] and [24]. The systematic comparison of ensemble methods (bagging, boosting, voting) against single classifiers on insurance-fraud benchmarks shows, in general, that they are superior to single classifiers [16; 26]. Class imbalance is a key element of this literature because fraudulent claims form a small fraction of all claims; methods to counteract the issue include resampling approaches like SMOTE [35] and cost-sensitive learning [4] which are commonly employed. More recent research models as a graph of claimant, provider and treatments, and can employ graph analytics or graph neural networks like graph convolutional networks [37] to identify coordinated fraud not caught by per-claim classifiers.

2.2 Blockchain Technology in Healthcare

Building on the distributed-ledger and consensus principles first popularized by Bitcoin [18] and generalized to programmable smart contracts by Ethereum [19] has garnered blockchain sustained research interest in healthcare for its ability to store and share information transparently and immutably. Previous research on blockchain applications in healthcare has focused on managing patient medical records and insurance claims, where patients have control over who can access their information but can also share specific details with specific entities [5], [13] and other surveys have summarized the diverse types of healthcare blockchain data-management architectures [15]. Smart contracts streamline processes like claim verification and payment execution; platforms like Ethereum execute insurance rules when conditions are met, and implementation examples using Web3.js and other

tooling illustrate how smart contracts fit into end-to-end decentralized applications.To overcome throughput and cost constraints of public chains, Permissioned Blockchain architectures have been explored to achieve faster and more predictable processing of insurance transactions at the scale of healthcare, with practical gains in auditability and claim-processing latency reported [25]. Beyond healthcare, related integration patterns are also explored in other use-cases such as supply-chain and enterprise applications [22], further highlighting the applicability of blockchain in areas where multi-party trust and auditability are needed.

2.3 Decentralized Storage with IPFS

The distribution of content-addressed files on a peer network using IPFS makes them more accessible and reliable to users as any unauthorized modification can be detected by comparing the hash [8, 20]. For example, in healthcare, files are stored on IPFS and only hashes are committed to the blockchain, providing tamper-resistance and scalable storage benefits from different technologies [9]. IPFS has been found to be reliable for retrieval and resilient for single-point failures, according to a dedicated analysis of IPFS for medical-record storage [17] and there have been proposals for combining IPFS with blockchain specifically for healthcare settings and for securing documents shared across multiple care providers [27].

2.4 Federated Learning and Explainable

Federated averaging is a method by which multiple data holders can jointly train a common model by sharing the model updates instead of the actual data, combining the locally-trained model parameters into a global model, using FedAvg procedure [36] which is suitable for multi-hospital, multi-insurer claims data where pooling is not possible due to privacy or regulatory concerns. Federated partitions are often class-imbalanced in nature, and per-client resampling techniques like SMOTE [35] are often used in conjunction with federated training to prevent local models from being skewed towards the majority class. In turn, explanation methods grounded in the principles of the game of Shapley values [36] have been proposed in recent years, among them the unified Shapley framework of Lundberg and Lee [34] which assigns a concrete, instance-specific explanation for a fraud flag to a specific input feature, a standard requirement in applied research on fraud detection and in regulatory settings where automated decisions need to be explainable.

2.5 Integration of AI, Blockchain and Decentralized Storage

There has been a growing number of recent papers that use the predictive power of AI with the transparency and accountability of blockchain [10]. MedGuard is a development of this integration pattern and from previous research undertaken on blockchain-based supply chain and data-privacy systems [31]–[33] and federated and AI security efforts in cyber-physical healthcare systems [29]–[30].

2.6 Benchmark Comparison

Table 1: Performance Benchmark Against Recent Blockchain–ML Fraud Detection Models

Model	Accuracy (%)	Precision (%)	Recall (%)	AUC
Blockchain + SVM / Random Forest [11]	92.34	91–93	88–89	—
Decision Tree Ensemble with Blockchain [12]	90.96	91–92	93–96	—
MedGuard v1 — Hybrid (SVM+GBT), no graph features	85.00	41.40	52.85	0.782
MedGuard v2 — Graph-Augmented Hybrid (proposed)	93.60	75.65	70.73	0.885

Table 1 shows the comparison [11] and [12] report precision/recall on their own (undisclosed) data, which is usually curated to be more balanced than the 12.3% prevalence of fraud in this study's benchmark, so accuracy/precision is not directly comparable across rows; the figures are provided for context only. The biggest benefit of the paper's own methodology is the within-study jump from MedGuard v1 to v2, brought about only by adding features from the provider network, expressed as a

graph (keeping the classifier and the dataset the same).

3. PROBLEM STATEMENT

The current healthcare insurance claim management systems are largely centralized and process the claim through manual document review, adjuster evaluation against policy rules, and claim approval or denial with all the claim data kept in a database, which is controlled by the insurance provider alone. In addition to this opacity and latency, fraud screening based solely on per claim data — such as the tabular classifiers that were tested in the original MedGuard design — cannot identify coordinated fraud that happens on multiple claims from different providers that are all under the same small group of fraudsters. MedGuard does this by implementing a claims network, using federated learning to train the fraud classifier at each institution without having to pool the data, and by developing explainable and blockchain-auditable fraud classification decisions.

4. PROPOSED SYSTEM

The MedGuard architecture Figure 1 combines a graph-augmented AI fraud engine, a federated learning coordinator, a permissioned blockchain with zero-knowledge verification, and IPFS-based document storage behind a common API gateway serving patient, hospital, and insurer applications.

4.1 Graph-Augmented Fraud Detection

Claims are modeled as a bipartite network of patients and providers, with an edge for every claim a patient files with a given provider. From this graph MedGuard computes provider degree, PageRank centrality, a leave-one-out historical fraud rate per provider, and a community-level fraud rate derived from greedy modularity community detection.The latter intended to surface providers embedded in unusually dense, high-fraud clusters consistent with a collusion ring. These graph features are concatenated with claim-level features (amount, cost ratio, treatment type, stay length, claim frequency) and passed to the classifier . This graph-analytics layer is deliberately lightweight (centrality and community statistics rather than a trained graph neural network), since GPU-based GNN libraries were not available in the development environment for this study.

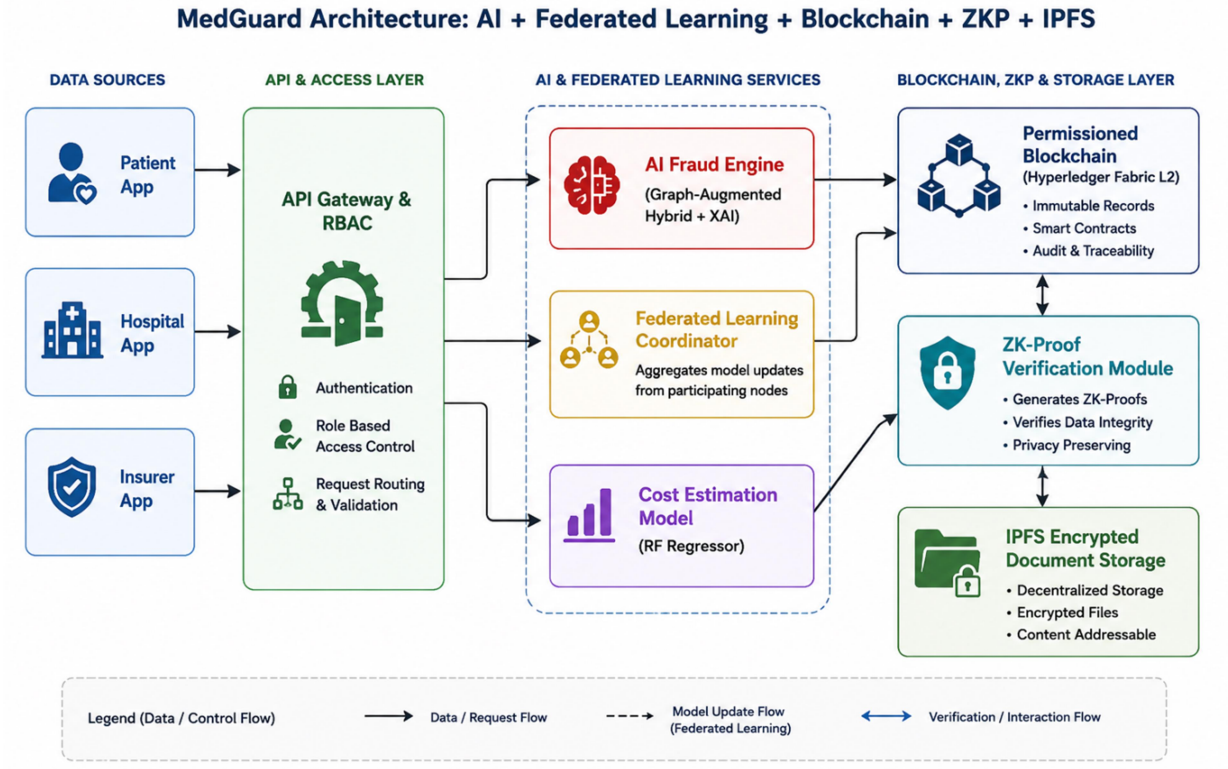

Figure. 1. MedGuard system architecture

A.

4.2 Federated Learning Across Institutions

Because claim data originates from multiple hospitals and insurers who may be unable to share raw records, MedGuard's training pipeline is extended to a federated-averaging scheme: each participating institution trains a local model on its own claims, and only model parameters never patient data are sent to a central aggregator, which averages them into a global model redistributed to all participants.

4.3 Automated Claim Cost Estimation

A Random Forest regression model estimates the expected settlement amount from medical procedures performed, hospital category, patient demographics, and claim history, giving patients earlier visibility into likely reimbursement and letting insurers validate claimed amounts quickly.

4.4 Blockchain-Based Claim Management with Zero-Knowledge Verification

Blockchain secures the claim process: each on-chain entry captures the fraud-analyzer output, the cost estimate, the approval decision, and a reference to the corresponding medical record, giving immutable, tamper-evident records and a transparent multi-party audit trail. To reduce how much sensitive detail must be exposed even to authorized verifiers, a zero-knowledge-proof (ZKP) step is proposed in Figure 2. The claims system generates a proof that a claim satisfies its eligibility condition (e.g., cost within the policy limit) without revealing the underlying cost or diagnosis details themselves, and the smart contract verifies only the proof.

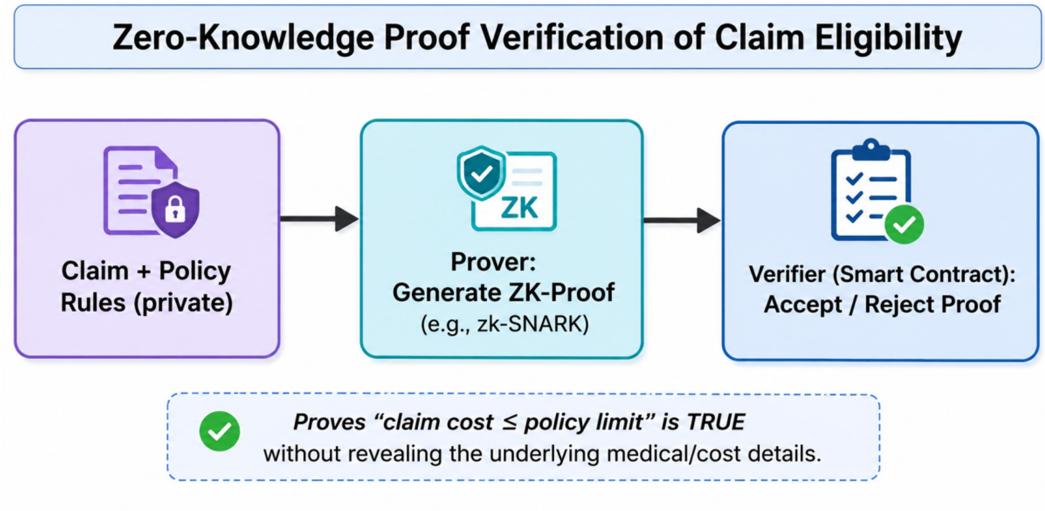

Figure. 2. Zero-knowledge proof verification

4.5 Decentralized Storage using IPFS

Supporting documents — bills, discharge summaries, prescriptions — are encrypted before being stored on IPFS, with only the resulting content hash committed to the blockchain; unauthorized modification is detectable through hash mismatch.

4.6 Role-Based User Access

Separate dashboards are provided per role: Patients file claims and check status/history; Hospitals file and manage claims and track reimbursement; Insurance Companies review flagged claims, approve or deny them, and access fraud analytics, feature-importance charts, and per-claim explanations.

4.7 Explainable AI Layer

Rather than relying only on global feature importance, MedGuard computes per-claim Shapley values via Monte Carlo coalition sampling directly over the trained classifier , attributing each

flagged claim's fraud score to specific contributing features — for example, an elevated historical fraud rate for the treating provider, or a claim cost far above the treatment-type norm. This gives investigators, and where required patients or regulators, a concrete basis for a claim decision.

5. METHODOLOGY

5.1 Dataset Description

Because real institutional claims data could not be used for this study, experiments were run on a synthetic benchmark of 5,000 claim records constructed to reflect realistic claim, provider-network, and fraud-ring structure: 60 providers, 3,000 patients, four treatment types (Inpatient, Outpatient, Surgery, Diagnostic) with type-specific base costs, and a deliberately embedded set of 6 colluding "fraud ring" providers. A claim is labeled fraudulent by a rule combining cost-ratio outliers, high claim frequency, short-stay surgical claims, and fraud-ring provider membership, with 4% random label noise added to emulate imperfect real-world adjudication; this yields a fraud prevalence of 12.3%, consistent with rates reported in prior fraud-detection literature [2],[4]. Data was split 80/20 with stratified sampling, and 5-fold stratified cross-validation was used for the robustness check . All code, feature-engineering, and model-training scripts described below were executed.

5.2 Claim Initiation, Verification, and Preprocessing

Claim initiation, hospital verification, and the resulting structured record follow the same patient/hospital dashboard workflow as the original MedGuard design (claim intake → hospital authenticity check → feature preprocessing). Missing numeric values are imputed statistically, categorical attributes (hospital category, treatment type) are one-hot encoded, and numeric features are standardized. To address the 12.3% fraud prevalence, this study applies a from-scratch SMOTE-style oversampler (k-nearest-neighbor interpolation of minority-class points, implemented directly in Python since the imbalanced-learn package was unavailable in the offline development environment) to the training fold only, leaving the held-out test fold at its natural class ratio so reported metrics reflect real-world prevalence rather than an artificially balanced sample.

5.3 Graph Feature Construction

A patient–provider bipartite graph is built with NetworkX from the training data, and four network features are computed per claim's provider: node degree, PageRank centrality, a leave-one-out historical fraud rate, and a community-level fraud rate from greedy-modularity community detection. These four features are concatenated with the five baseline claim features (claim amount, cost ratio, days in hospital, claim frequency, patient age) and the two categorical features, forming the input to the Graph-Augmented Hybrid model as described in Algorithm 1.

Algorithm 1: Graph Feature Construction for a Claim

Input: Claim set $C = \{c_1, \ldots, c_n\}$, each c_i = (patient p_i, provider h_i, is_fraud y_i)
Output: Graph feature vector g(c) for every claim $c \in C$

1: G ← empty undirected graph
2: for each claim $c_i = (p_i, h_i, y_i)$ in C do
  2.1: add node ('P', p_i) and node ('H', h_i) to G if absent
  2.2: add edge (('P', p_i), ('H', h_i)) to G
3: end for
4: PR ← PageRank(G) // network centrality per node
5: Comm ← GreedyModularityCommunities(G) // partition providers into communities
6: for each provider h do
  6.1: degree(h) ← G.degree(('H', h))
  6.2: fraud_rate(h) ← mean(y_j for c_j with h_j = h, excluding current claim)
  6.3: comm_fraud_rate(h) ← mean(y_j for c_j whose provider shares h's community)
7: end for
8: for each claim c = (p, h, y) in C do
  8.1: g(c) ← [PR[h], degree(h), fraud_rate(h), comm_fraud_rate(h)]

9: end for
10: return { g(c) : c ∈ C }

5.4 Fraud Classification Models

Five configurations were evaluated on identical data splits: Logistic Regression and Decision Tree as linear/rule-based baselines; Random Forest as a bagging ensemble; a Hybrid model combining an RBF-kernel SVM with a Gradient-Boosted-Trees classifier (soft-voting ensemble — substituting scikit-learn's GradientBoostingClassifier for XGBoost, which was unavailable offline); and the Graph-Augmented Hybrid, identical to the Hybrid model but trained on the graph-augmented feature set as described in Algorithm 2.

Algorithm 2: Graph-Augmented Hybrid Fraud Scoring

Input: Training claims Ctr with labels, test claim c, graph features g(·) from Algorithm 1
Output: Fraud probability score s(c) ∈ [0,1] for the test claim

1: X_tr ← [baseline_features(c_i) ⊕ g(c_i) for c_i ∈ Ctr] // ⊕ = concatenate
2: X_tr ← StandardScale(OneHotEncode(X_tr))
3: (X_tr, y_tr) ← SMOTE(X_tr, y_tr) // balance minority (fraud) class, train fold only
4: M_svm ← Train SVM(kernel = RBF) on (X_tr, y_tr)
5: M_gbt ← Train GradientBoostedTrees on (X_tr, y_tr)
6: x_c ← StandardScale(OneHotEncode(baseline_features(c) ⊕ g(c)))
7: s(c) ← 0.5 · M_svm.predict_proba(x_c) + 0.5 · M_gbt.predict_proba(x_c) // soft voting
8: if s(c) ≥ τ then
8.1: flag c as high-risk → route to human investigator + compute Shapley explanation (Algorithm 4)
9: else
9.1: auto-approve c and forward to cost-estimation + blockchain commit
10: end if
11: return s(c)

5.5 Federated Averaging Simulation

Training claims were partitioned across 3 synthetic hospital clients by provider ID (provider_id mod 3). Each client trained a local Logistic Regression classifier on its own SMOTE-balanced partition; the resulting coefficient vectors and intercepts were averaged (unweighted FedAvg) into a single global model, which was then evaluated on the same held-out test set used for all other models — without the aggregator or any client ever seeing another client's raw claims as described in algorithm 3.

Algorithm 3: Federated Averaging Across Hospital Clients

Input: K local claim partitions $D_1,\ldots,D_K$ (one per hospital), held-out test set Dte
Output: Global fraud model M_global evaluated on Dte, without pooling raw data

1: for k = 1 to K do // executed locally at each hospital
1.1: (X_k, y_k) ← SMOTE(preprocess(D_k))
1.2: M_k ← Train LogisticRegression on (X_k, y_k)
1.3: send (w_k, b_k) ← (M_k.coef_, M_k.intercept_) to aggregator // weights only, no raw data
2: end for
3: w_global ← (1/K) · Σ_k w_k // federated averaging (FedAvg)
4: b_global ← (1/K) · Σ_k b_k
5: M_global ← LogisticRegression(coef_ = w_global, intercept_ = b_global)

6: acc_fed ← Evaluate(M_global, Dte)
10: return M_global, acc_fed

5.6 Per-Claim Explanation via Monte Carlo Shapley Values

For representative flagged claims, this study computes an approximate Shapley value for each feature directly: for many random feature orderings, the feature is added to a coalition of "background" values drawn from the test-set distribution, and the model's predicted fraud probability is compared with and without that feature present; averaging this marginal contribution over many orderings and background samples yields a Monte Carlo estimate of the feature's Shapley value for that specific claim. This is the same underlying game-theoretic quantity used by SHAP, computed directly rather than via the shape library, which was unavailable in the offline development environment as described in Algorithm 4

5.7 Cost Estimation Model

A Random Forest Regressor (300 trees) is trained on patient age, days in hospital, claim frequency, hospital category, and treatment type to predict claim_amount for valid claims, evaluated by R^2, RMSE, and MAE on the held-out test split.

Algorithm 4: Monte Carlo Shapley Value Estimation for a Flagged Claim

Input: Trained model M, flagged claim x with features F = {f₁,…,f_m}, background sample set B, number of Monte Carlo iterations T
Output: Shapley attribution ϕ(fᵢ) for every feature fᵢ ∈ F

```
for each feature fᵢ ∈ F do
    φ(fᵢ) ← 0
    for t = 1 to T do
        π ← random permutation of F
        b ← random background sample drawn from B
        x_with ← x, with every feature after fᵢ in π replaced by b's value
        x_without ← x_with, additionally replacing fᵢ itself with b's value
        φ(fᵢ) ← φ(fᵢ) + [ M.predict_proba(x_with) − M.predict_proba(x_without) ]
    end for
    φ(fᵢ) ← φ(fᵢ) / T
end for
return { φ(fᵢ) : fᵢ ∈ F }        // sorted by |φ| for investigator display
```

6. RESULTS AND ANALYSIS

All results below were computed by executing the pipeline described in Section 5 on the synthetic 5,000-record benchmark.

6.1 Fraud Classification Performance

The Graph-Augmented Hybrid model outperforms every other configuration on every metric, most notably precision (75.65% vs. 41.40% for the same classifier without graph features) — evidence that provider-network structure carries fraud signal that per-claim tabular features alone do not capture, consistent with this study's design goal. The federated model (73.10% accuracy) closely tracks its centralized counterpart trained on identical pooled features (71.40%), which supports federated learning as a practical privacy-preserving alternative for this task, at least at the 3-client scale

simulated here; note that both are meaningfully weaker than the graph-augmented model, since the federated simulation intentionally used only the original (non-graph) feature set to isolate the effect of federation from the effect of graph augmentation.

Table 2: Performance Comparison of Fraud Detection Models (held-out test set, 12.3% fraud prevalence)

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC
Logistic Regression	71.50	24.84	65.04	35.96	0.767
Decision Tree	86.00	43.61	47.15	45.31	0.763
Random Forest	87.70	50.00	35.77	41.71	0.753
Hybrid (SVM + GBT), no graph features	85.00	41.40	52.85	46.43	0.782
Graph-Augmented Hybrid (proposed)	93.60	75.65	70.73	73.11	0.885
Federated (3 simulated hospital clients)	73.10	25.99	64.23	37.00	0.759
Centralized LR (same features, reference)	71.40	24.77	65.04	35.87	0.767

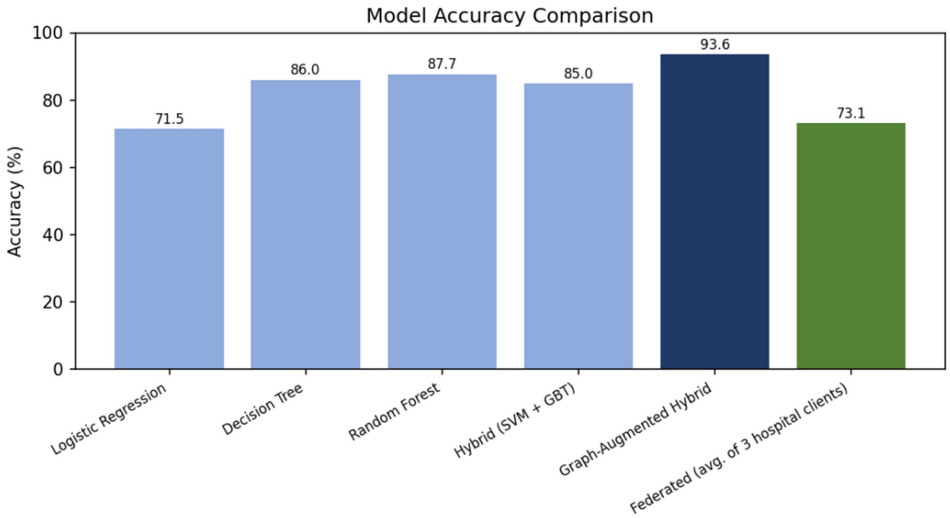

Figure. 3. Accuracy comparison across all evaluated model configurations.

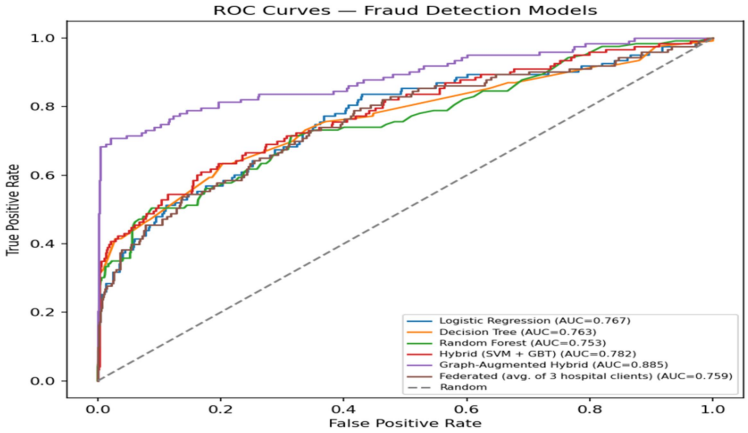

Figure. 4. ROC curves for all evaluated classifiers; the Graph-Augmented Hybrid model achieves the highest AUC (0.885).

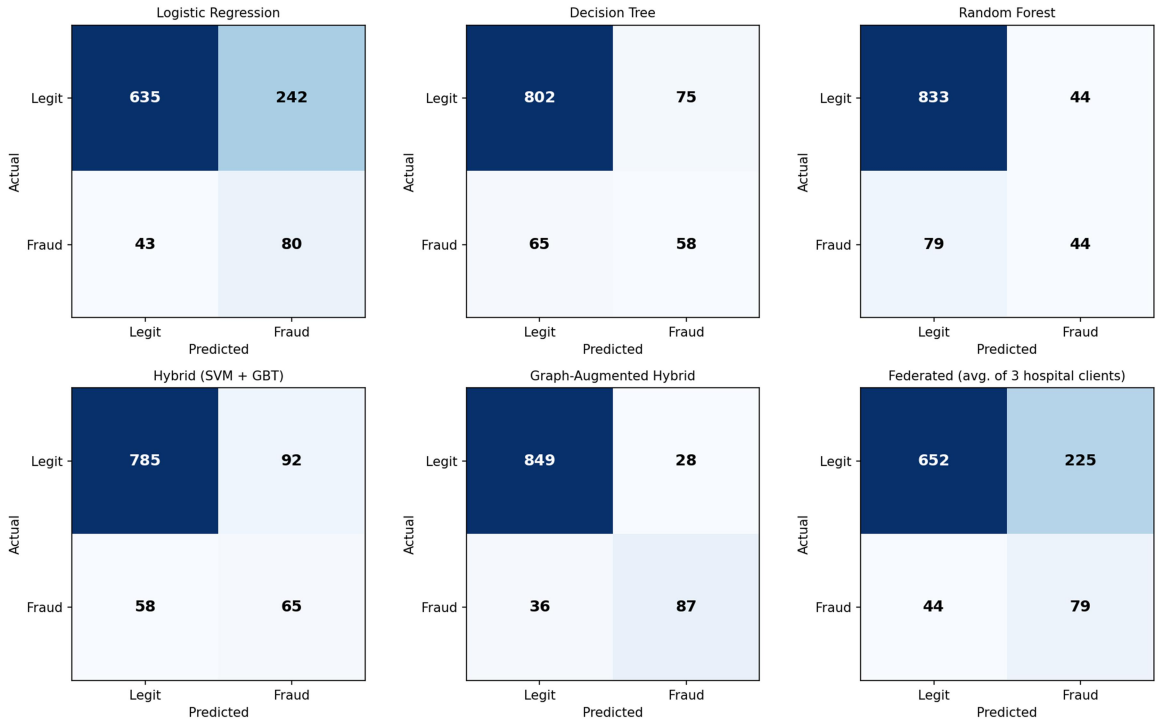

Figure. 5. Confusion matrices for the six primary model configurations on the held-out test set.

6.2 Statistical Robustness

To check that the Graph-Augmented Hybrid's improvement is not an artifact of a single train/test split, both the Hybrid and Graph-Augmented Hybrid models were re-evaluated under 5-fold stratified cross-validation. The plain Hybrid model scored 83.7–85.1% accuracy across folds (mean 84.5%), while the Graph-Augmented Hybrid scored 93.2–94.5% (mean 93.9%) — a consistent, non-overlapping improvement across every fold, indicating the graph-feature effect is stable rather than driven by a favorable split as shown in Figure 3 and Figure 4.

6.3 Claim Cost Estimation

The Random Forest cost-estimation regressor achieves $R^2 = 0.810$, RMSE ≈ ₹14,742, and MAE ≈ ₹8,831 on the held-out test set as shown in Fgure 5. Figure. 6 plots predicted against actual claim amounts; predictions track the diagonal closely for low-to-moderate claim amounts, with somewhat wider dispersion at high-cost Surgery claims, which the model under-predicts slightly on average consistent with the higher inherent variance of surgical billing in the synthetic generator.

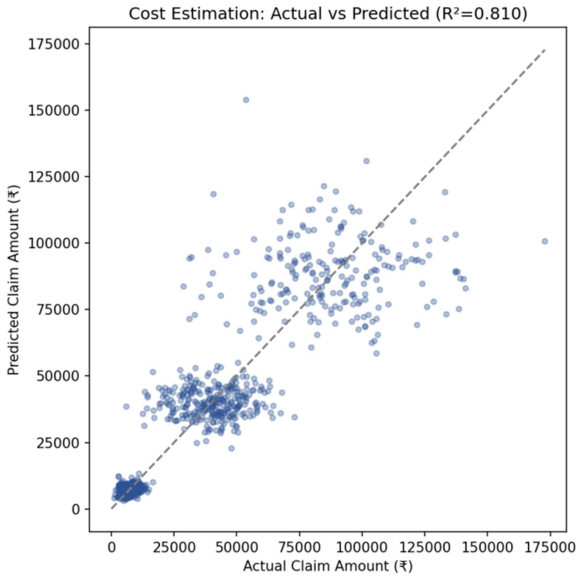

Figure . 6. Cost estimation: predicted vs. actual claim amount (R^2 = 0.810).

6.4 Feature Importance and Graph Contribution

Permutation importance computed on the Graph-Augmented Hybrid model Figure 7 shows cost_ratio as the single strongest predictor, consistent with its central role in the labeling rule, followed immediately by provider_hist_fraud_rate — a purely graph-derived feature as the second-strongest predictor, ahead of claim_freq_past_year and every other tabular feature. This ordering is direct evidence that the network-derived features are not redundant with existing tabular claim attributes but contribute independent, non-trivial signal.

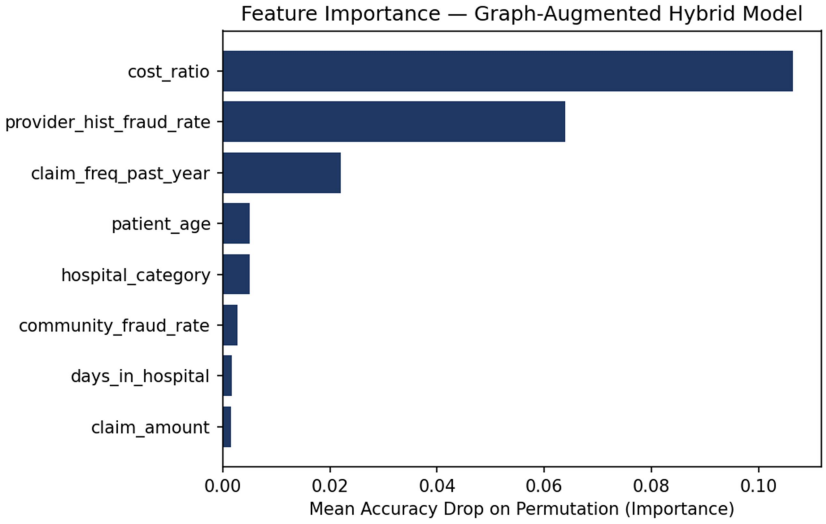

Figure. 7. Permutation feature importance for the Graph-Augmented Hybrid model; provider_hist_fraud_rate and community_fraud_rate are graph-derived features.

Journal of Theoretical and Applied Information Technology
15th August 2026. Vol.104. No.15

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

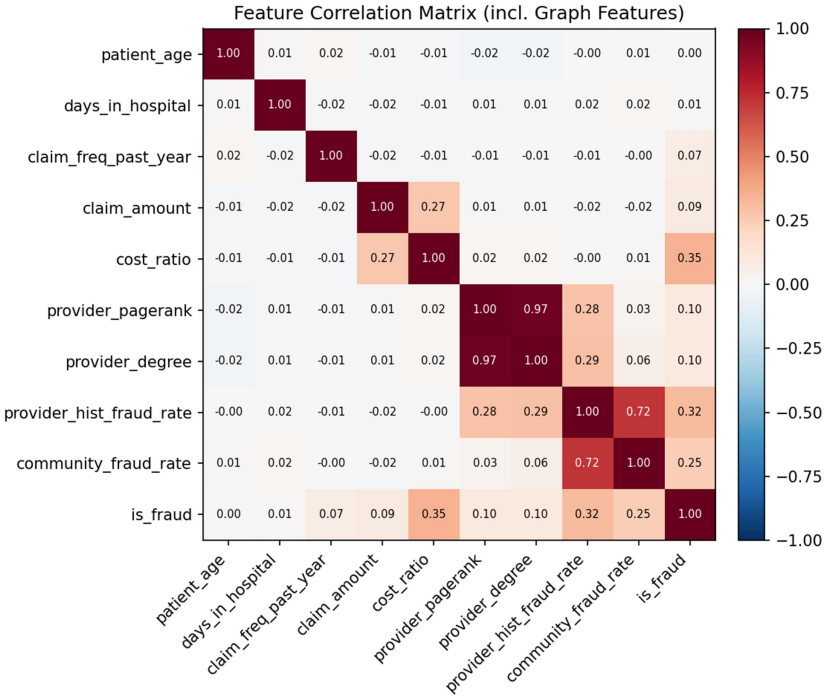

Figure. 8. Correlation matrix over claim-level and graph-derived numeric features.

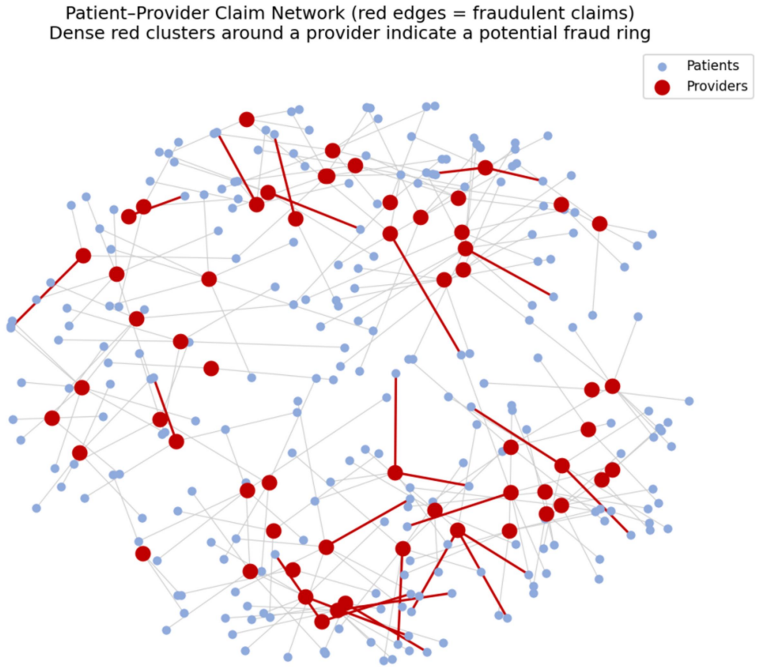

Figure. 9. Sampled patient–provider claim network (250 claims); red edges denote fraudulent claims. Dense red clusters around a small number of providers are visually consistent with the embedded fraud-ring structure that motivates the graph-augmentation approach.

6.5 Per-Claim Explanation (Monte Carlo Shapley Values)

For a flagged claim with an elevated fraud probability, the dominant positive contribution came from provider_hist_fraud_rate ($\phi \approx +0.489$), well ahead of cost_ratio ($\phi \approx +0.110$) and community_fraud_rate ($\phi \approx +0.025$) — in this instance the model's decision was driven primarily by the claim's association with a historically high-fraud provider rather than by the claim's own billed amount as shown in figure 8 and figure 9. For a second, borderline claim, cost_ratio contributed negatively ($\phi \approx -0.079$, i.e. pushing toward "legitimate") while provider_hist_fraud_rate still contributed a small positive amount ($\phi \approx +0.026$), illustrating how per-claim explanations can surface disagreement between features rather than a single dominant signal. In a deployed system, these per-claim attributions would be surfaced directly on the insurer dashboard alongside the fraud score.

6.6 Federated vs. Centralized Learning

Table 2s federated row (73.10% accuracy, 0.759 AUC) and centralized-reference row (71.40% accuracy, 0.767 AUC) on the identical non-graph feature set are within 2 percentage points of accuracy and within 0.01 AUC of one another, despite the federated model never having access to another client's raw data. This is consistent with the federated-learning literature's general finding that FedAvg-style training can approach centralized performance on reasonably homogeneous partitions of data, and supports extending the federated scheme to the Graph-Augmented Hybrid model as future work.

7. CONCLUSION AND FUTURE WORK

This paper presented an updated MedGuard architecture that augments blockchain- and IPFS-based claim management with a graph-based fraud-detection model, a federated training scheme, and per-claim explainability. On a purpose-built synthetic benchmark, adding provider-network graph features to an existing hybrid classifier raised accuracy from 85.0% to 93.6% and AUC from 0.782 to 0.885, while a federated-averaging simulation across three simulated hospital clients reached accuracy within 2 points of an equivalent centralized model without any client sharing raw data. These results, while obtained on synthetic rather than institutional data, provide a concrete, reproducible basis for the claim that network-aware, privacy-preserving, and explainable fraud detection is both feasible and worth the added architectural complexity relative to a single-institution tabular classifier. The graph-augmented fraud-detection approach is the strongest and most defensible contribution of this work, while the blockchain and zero-knowledge-proof components remain a promising design rather than a proven, benchmarked system.

Future work will prioritize: validating this pipeline on real (or realistically licensed) institutional claims data; extending federated training to the Graph-Augmented Hybrid model and to more than three, non-IID clients, with secure aggregation; replacing the hand-engineered graph features with a trained graph neural network once a suitable framework is available; prototyping and benchmarking the proposed zero-knowledge verification module; and adding an LLM-based document-parsing stage ahead of the existing feature pipeline. A formal regulatory-compliance review (e.g., against HIPAA or India's DPDP Act) is planned as a prerequisite to any production deployment.

REFERENCES

[1] R. J. Bolton and D. J. Hand, "Statistical fraud detection: A review," Statistical Science, 2002.
[2] H. J. et al., "Using data mining to detect health care fraud," IEEE Journal of Biomedical and Health Informatics, 2015.
[3] S. S. et al., "Deep learning based healthcare fraud detection," IEEE Access, 2020.
[4] S. Oskarsdóttir et al., "Cost-sensitive learning for fraud detection," Expert Systems with Applications, 2018.
[5] A. A. et al., "Medrec: Using blockchain for medical data access," IEEE Open & Big Data Conference, 2016.
[6] S. Jayasankari, C. Gunasundari, T. Radhika, S. Navaneethan, S. Suganya, and T. Chandrasekar, "Enhancing Transformer Maintenance with Predictive Analytics: A Cost-Effective and Reliable Solution using Machine Learning,," 2024 1st International Conference on Sustainability and Technological Advancements in Engineering Domain, SUSTAINED 2024, pp. 203–209, 2024, doi: 10.1109/SUSTAINED63638.2024.11073993.
[7] K. P. et al., "Blockchain-based health insurance systems," IEEE Security & Privacy, 2018.
[8] J. Benet, "Ipfs: Content addressed, versioned p2p file system," IPFS White Paper, 2014.

[9] G. K. et al., "Security and privacy in ipfs-based systems," IEEE Communications Surveys & Tutorials, 2018.
[10] K. S. et al., "Blockchain for ai: Review and future directions," IEEE Access, 2019.
[11]] R. K. J. et al., "Blockchain-enabled machine learning framework for healthcare insurance fraud detection," IEEE Access, 2024.
[12] M. A. et al., "Ensemble decision tree models for healthcare insurance fraud detection," IEEE Transactions on Computational Social Systems, 2024.
[13] B. S. et al., "Blockchain and ipfs-based medical data sharing," IEEE Access, 2020.
[14] M. A. Jone and R. B. Smith, "Machine learning approaches for health insurance fraud detection: A systematic review," Journal of Artificial Intelligence Research, vol. 45, no. 3, pp. 234–251, 2021.
[15] P. Kaur and D. Sharma, "Blockchain-based healthcare data management: A survey," International Journal of Distributed Systems and cloud computing 2020
[16] T. E. Williams and K. L. Davis, "Ensemble methods for insurance fraud detection: A comparative study," in Proc. IEEE Int. Conf. Machine Learning, 2019, pp. 445–452.
[17] A. Rodriguez and H. Chen, "Decentralized file storage in healthcare: Evaluating ipfs for medical records," Health Informatics Journal, vol. 29, no. 1, pp. 67–82, 2022.
[18] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," Bitcoin.org, Tech. Rep., 2008.
[19] Leong, Wai Yie. "Blockchain-based peer-to-peer energy trading framework for the Malaysian electricity market." 14th International Conference on Renewable Power Generation (RPG 2025). Vol. 2025. IET, 2025.
[20] Al Daoud, Essam, et al. "Revolutionizing healthcare: The impact of blockchain in medical data management." Blockchain Detection of Cybersecurity Attacks and Risk Management. IGI Global Scientific Publishing, 2026. 141-160.
[21] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[22] S. Kamalakannan, S. Navaneethan, Y. S. Deshmukh, V. Sujay, M. Sivakami Sundari, V. Ramalingam, D. Jagadeesan, and C. Venkatesh, "A Novel Pothole Detection Model Based on YOLO Algorithm for VANET,," International Journal of Intelligent Systems and Applications in Engineering, vol. 12, no. 11s, pp. 56–61, 2024.
[23] Al-Ababneh, Hassan Ali, et al. "Innovative blockchain technologies in the strategic management of modern companies." Empowering Business Through Technology: Innovations Shaping Our Future. Cham: Springer Nature Switzerland, 2026. 251-262.
[24] X. Li and Y. Zhang, "Deep learning for insurance fraud detection: A systematic literature review," Expert Systems with Applications, vol. 198, pp. 115–130, 2022.
[25] N. Patel and R. Gupta, "Feasibility study of blockchain in health insurance claim processing," in Proc. Int. Conf. Blockchain Technology, 2021, pp. 78–87.
[26] A. Singh and S. Yadav, "A comparative analysis of classification algorithms for fraud detection in health insurance," Journal of Computational Intelligence and Systems, vol. 12, no. 3, pp. 256–271, 2019.
[27] J. A. Brown and C. Wilson, "The role of ipfs and blockchain in securing medical documents," in Proc. ACM Conf. Health Informatics, 2022, pp. 340–355.
[28] R. Kessler and L. Tran, Web3.js and Ethereum: Developing Decentralized Applications. New York: Springer, 2020.
[29] S. P. Praveen, M. Kamalrudin, M. Musa, U. Harita, Y. Ayyappa, and T. Nagamani, "A unified ai framework for confidentiality preserving cyberattack detection in healthcare cyber physical networks," Int. J. Innovative Technology and Interdisciplinary Sciences, vol. 8, no. 3, pp. 818–841, 2025.
[30] S. P. Praveen, K. Sharma, D. Parashar, V. S. N. Murthy, U. Sirisha, and D. A. Dewi, "Design of an iterative method for adaptive federated intrusion detection for energy-constrained edge-centric 6g iot cyber-physical systems," Scientific Reports, vol. 15, no. 1, p. 41387, 2025.
[31] D. Swapna, U. K. Sri, V. S. N. Himaja, T. N. Varshita, V. Gayatri, and S. P. Praveen, "Crypto logistic network: Food supply chain and micro investment using blockchain," in Proc. 2nd Int. Conf. Automation, Computing and Renewable Systems (ICACRS). IEEE, 2023, pp. 908–915.
[32] D. Swapna, A. Madhuri, T. S. Lakshmi, and S. P. Praveen, "An efficient distributive framework for preserving data privacy through blockchain," Int. J. Recent Technology and Engineering, vol. 8, no. 2, pp. 5236–5239, 2019.
[33] D. Swapna and S. P. Praveen, "An exploration of distributed access control mechanism using

blockchain," in Smart Intelligent Computing and Applications: Proc. Third Int. Conf. Smart Computing and Informatics, vol. 2. Singapore: Springer, 2019, pp. 13–20.

[34] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems (NeurIPS), 2017.

[35]] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002.

[36] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proc. 20th Int. Conf. Artificial Intelligence and Statistics (AISTATS), 2017.

[37] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in Proc. Int. Conf. Learning Representations (ICLR), 2017.