Journal of Theoretical and Applied Information Technology
28th February 2026. Vol.104. No.4

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

REAL-TIME INTELLIGENT CLAIM VERIFICATION USING BLOCKCHAIN-BASED RPA AND XGBOOST FOR SCALABLE HEALTHCARE FRAUD DETECTION

DR.V. BALASANKAR[1], DR. V SRINADH[2], DESIDI NARSIMHA REDDY[3], JALA PRASADARAO[4], K SWETHA[5], ELANGOVAN MUNIYANDY[6]

[1]Department of Computer Science and Engineering (AIML and CS), Godavari Global University (GGU), Rajahmundry, India.
[2]Associate Professor, Department of CSE-AIML, GMR Institute of Technology, Rajam, India.
[3]Data Consultant,Soniks consulting LLC, 101 E park blvd, suite no: 410, Plano,TX, 75074, USA.
[4]Department of Computer Applications, Aditya University, Surampalem, India.
[5]Assistant Professor, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, India.
[6]Department of Biosciences, Saveetha School of Engineering. Saveetha Institute of Medical and Technical Sciences, Chennai, India.
Email id: [1]balasankar.v@gmail.com, [2]srinadh.v@gmrit.edu.in, [3]dn.narsimha@gmail.com, [4]john211286@gmail.com, [5]swetha.k@kluniversity.in, [6]muniyandy.e@gmail.com

ABSTRACT

Healthcare insurance fraud has been a major problem and requires smart, scalable, and auditable detection systems. The study presents a framework of Robotic Process Automation (RPA) with Integrated Blockchain technology and integrates the methods of detecting fraudulent claims in real-time with Machine Learning (ML) and RPA. The system is developed in five layers, including data input, automation, intelligence, smart contracts, and blockchain ledger. It works on a structured dataset of 4,000 healthcare claims having 83 attributes. Claim intake and rule-based screening are automated by RPA bots, and fraud classification against predefined risk thresholds is done by an ML engine driven by extreme gradient boosting using smart contracts, with final results being impartially stored by Hyperledger Fabric. The proposed framework significantly outperformed the traditional models of predicting data (Convolutional Neural Network, Recurrent Neural Network, Decision Trees, and Logistic Regression) with a result of 92.4 percent of classification accuracy. It guarantees scalability and traceability without compromising the latency by making claims in under 2.5 seconds. Measurements of feature importance, e.g. KullbackLeibler Divergence, Entropy, and Gini, are more explainable and guide automation logic. What this hybrid framework will achieve is a new standard of real-time, audible, and intelligent fraud detection in healthcare insurance systems that will provide a technically sound and legally compliant basis of foundation to next-generation claims processing.

Keywords: *Healthcare Insurance, Blockchain, Fraudulent Claims, Hyperledger Fabric, Robotic Process Automation.*

1. INTRODUCTION:

The healthcare insurance sector is increasingly experiencing issues due to the increased complexity and number of false claims that are costly in monetary terms and require a significant amount of institutional strain [1]. False or exaggerated diagnoses, over-billing on treatment procedures that were not done, and deliberate falsification of medical records are the common dishonest practices in insurance claims. Human inspections or various ML systems are the common methods of locating scams [2], [3], [4], [5]. Nonetheless, the approaches are not highly scalable, understandable, and responsive in real-time, particularly in large national health care systems [6].

The increasing volume of claims in the health insurance system has made it a very important

requirement that we have intelligent, automated, and secure systems that would be able to efficiently filter, classify, and document the claims without compromising the effectiveness and compliance with the regulations [7], [8], [9], [10]. The multi-faceted nature of the contemporary claims data, which involves high-dimensional characteristics, such as diagnostic codes, billing amounts, patient history, and procedures, cannot be easily met by the use of the legacy systems [11], [12]. Besides that, the traditional deep learning (DL) model is versatile but lacks the ability to be interpreted as well as may be described as having a high latency, which is not favorable to operational environments that require explainability and must be audited legally [13], [14], [15].

This study proposes a solution to these issues that can be achieved by offering a Blockchain-Integrated RPA architecture, providing a combination of ML-based and deterministic rule-based automation with an audit that is enabled by a blockchain. The architecture brings in a five-layered architecture with RPA in smart claim intake, XGBoost-based ML in fraud classification, and Hyperledger Fabric in recording immutable transactions. Smart contracts provide automated enforcement of rules, which means that decisions can be made in real-time based on the acquired risk scores [16], [17], [18], [19], [20].

The key objective of the research will be to develop and evaluate a scalable, understandable, and secure fraud detection system that is structured healthcare claim information specific. This research analyzes the predictive power of the model, processing speed, and the capacity to safely log in a real-world dataset comprising of 4000 applications and 83 features. To prove the effectiveness, transparency, and regulatory compliance of the proposed hybrid model, the authors compare them with the models of the state-of-the-art (SOTA), such as Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Decision Trees (DT), and Logistic Regression (LR). The research is critical in the sense that it is laying a groundbreaking proposal on detection of healthcare insurance fraud in future. Proposed framework, along with AI-based intelligence, automation, and blockchain technologies, is capable of not only high classification rates (92.4%), as well as low latency (less than 2.5 seconds), but also compliance, achieved by means of an explainable logic and a tamper-proof record. The system suggests a standard of how the institutions can future-proof their claim handling processes using intelligent auditable technologies.

By combining RPA, explainable ML, and permissioned blockchain, this research contributes to the current knowledge base by demonstrating that the process of detecting healthcare insurance fraud can be accurate, in real-time, explainable, and auditable simultaneously. Although past studies have concentrated on these areas individually, the findings of this paper demonstrate that with a combination of these areas, the efficiency of operations, regulatory openness, as well as decision trust, is significantly improved- all without reducing accuracy. Sealing an important gap between theory and actual healthcare insurance functioning, the findings provide a realistic design outline of deployable fraud detection systems that are regulation-capable.

1.1 Research Motivation and Literature Screening

Current healthcare insurance fraud detection systems lack scalability, real-time processing, interpretability, and auditability, causing huge financial losses. Slow, error-prone manual audits and black box ML or deep learning models are unsuited for regulatory situations. Thus, an automatic, explainable, and safe fraud detection framework that processes claims in real time and ensures transparency and compliance is needed**.**

This study targets inefficient, non-auditable healthcare insurance fraud detection systems that lack accuracy, low latency, and explainability. Due to rising claim volumes and the lack of integrated solutions combining automation, intelligent decision-making, and tamper-proof record keeping, the topic was chosen.

Academic journals and conferences on healthcare fraud detection, machine learning, RPA, and blockchain were consulted. Scalability, interpretability, real-time processing, and security studies were included, but theoretical or non-healthcare works were excluded.

This work introduces a comprehensive framework for real-time fraud detection that combines RPA for automated claim processing, explainable XGBoost-based intelligence, and blockchain for immutable auditability. It differs from previous studies that depend on standalone ML/DL models or isolated blockchain solutions. In contrast to previous research, the suggested solution simultaneously ensures transparency, scalability, and regulatory compliance, while achieving a high accuracy rate of 92.4% and a low latency rate of less than 2.5 seconds.

1.2 RESEARCH GAP

Healthcare insurance fraud detection literature mostly employs standalone machine learning or deep learning models, which lack interpretability and real-time applicability, or blockchain-based systems that prioritize data security over intelligent decision-making. The majority of RPA studies focus on administrative automation, not predictive fraud intelligence. Thus, a verified, end-to-end system that provides automation, explainable fraud detection, real-time processing, and immutable auditability is lacking in the literature. This study proposes and validates a unified RPA–ML–Blockchain architecture for healthcare insurance fraud detection, filling this gap.

2. METHODOLOGY

2.1. System Architecture

The proposed system embraces a modular and scalable system that incorporates RPA, ML and Blockchain technology to identify fraudulent healthcare insurance claims in an automated, transparent, and secure system. The architecture consists of five primary layers, including the Data intake layer, the Automation layer, the Intelligence layer, the Smart contract layer and blockchain ledger layer.

2.1.1. Architecture Components and Data Flow

Data Intake Layer

Information in structured digital forms (e.g. a hospital billing system and electronic health records) and scanned documents are inputted into this layer and it handles the processing of raw healthcare claim data. The data will consist of 4000 insurance claims with 83 features such as billing amounts, procedure codes, hospital names, patient history, and past fraud flags.

- **Automation Layer (RPA)**
  Robotic process automation (RPA) bots developed with platforms such as UIPath or Automation Anywhere can currently perform basic eligibility checks, retrieve pertinent information from forms, verify required inputs, and forward claims to the fraud detection pipeline without human intervention. RPA is like human agents, except it's far faster and more accurate at what it performs.

- **Preprocessing & Feature Engineering Module**
  Extracted data is passed to a preprocessing unit that performs the missing value imputation, one-hot encoding for categorical attributes, and Z-score normalization or Min-Max scaling for numerical attributes. Feature selection techniques such as recursive Feature Elimination (FRE) are used to reduce dimensionality while preserving fraud-related signals.

- **Intelligence Layer (ML Engine)**
  Claims are defined as either fake or real by learnt ML models (like XGBoost and DNN) in this layer. Binary Cross-Entropy Loss, as well as Adam Optimiser, are used to make the models smarter after they are trained on labelled claim data from the past. The model provides each claim a risk score based on its chance result. To improve the accuracy of predictions, activation functions such as ReLU and Sigmoid are used, along with entropy-based measures such as Gini and Shannon Entropy.

- **Smart Contract Layer**
  Based on the model's output, a smart contract evaluates whether the risk score exceeds the predefined fraud threshold (for instance, 0.85). The claim is marked and directed for manual audit if such is the case. The smart contract initiates the proper Blockchain transactions while applying

deterministic rules to express branching logic.

- **Blockchain Ledger Layer**
  The Blockchain, which is built on Hyperledger Fabric, permanently stores the hashed version of the final opinion, and whether it was automated or reviewed. Included in this are the decision date, fraud label, risk score, and claim information. Every claim that goes through the pipeline can now be fully audited and recorded in an unchangeable way.

2.2. System Overview

The proposed hybrid model uses a well-combined pipeline of Robotic Process Automation (RPA), Machine Learning (ML), and Blockchain to provide a high-scale, secure, and smart fraud detection framework of healthcare insurance claims. The end-to-end architecture includes the initial stage of data extraction and rule-based screening, which is provided through RPA and then, the fraud is classified using ML, and finally, the immutable decision-making is made with the help of a permissioned blockchain infrastructure. This pipeline ensures transparent, real-time fraud detection, reducing manual intervention while increasing accuracy and auditability. Figure 1 depicts the schematic flow of the proposed hybrid framework. This structured pipeline serves as the backbone of the proposed solution and was built to operate in real-time while processing sensitive medical as well as financial data securely.

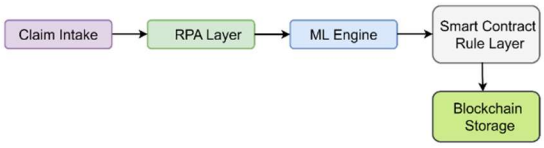

Figure 1. Schematic Flow of Proposed Hybrid Framework

2.3. Dataset Description

The dataset utilised in the research is made up of 4000 healthcare insurance forms, and each provides 83 features. Among these characteristics are claim information, billing summaries, audit logs, diagnostic and procedure codes, patient demographics, and past claim behaviour.

Missing value management, categorical variable encoding, and normalisation are all part of the initial preprocessing. Z-score normalisation is used to standardise feature values and remove scale discrepancies; it is calculated as follows:

$$z = \frac{x - \mu}{\sigma}$$

Where x represents the feature value, μ indicates the mean, and σ signifies the standard deviation of the feature.

Parallelly, Min-Max scaling is also employed for the bounded attributes like monetary values to compress them within the range [0, 1]:

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Such a two-fold normalization method enhances convergence of the model and helps in alleviating the problem of gradient vanishing or explosion. After that to preserve the most predictive capacity and to avoid unwanted features, dimensionality reduction methods like recursive feature elimination (RFE) are used.

2.4. Robot Process Automation Layer

RPA is the intake engine of the system. Using systems such as UIPath and Automation Anywhere, bots are coded to find structured and unstructured data in claim forms, verify them according to rule-based limitations, and send them to the right decisioning route.

RPA uses fixed regulations in implementing deterministic screening, making it consistent and faster in nature. As an example, when there are pregnancy-related treatment codes in a claim and a patient is identified as a male, the bot automatically marks that claim to be subject to additional examination. This initial triaging can be used to drastically cut down the downstream processing load of the ML engine.

2.5. Blockchain Integration

To achieve security and integrity, the system has Hyperledger Fabric, which is a limited access Blockchain platform. Once a claim is identified by an ML engine as potentially fraudulent or benign, the findings, metadata and blockchain record of the conclusion of the human reviewer (where needed) is hashed.

A fraud decision threshold is implemented using a simple deterministic rule, expressed as:

$$f(x) = \begin{cases} 1 & if\ s(x) \geq 0.85 \\ 0 & otherwise \end{cases}$$

Where s(x) represents the probability score output from the ML model. To support automated routing of results, the state transition function is defined as:

$$\delta(q,a) = \begin{cases} q_{approve} & if\ s(x) < 0.65 \\ q_{review} & if\ 0.65 \leq s(x) < 0.85 \\ q_{reject} & if\ s(x) \geq 0.85 \end{cases}$$

Where q represents the current state, while a signifies the action enabling smart contract-based branching.

Transaction costs related to Blockchain logging are modeled using:

$$Cost_{tx} = \sum_{i=1}^{n} g_i * p_i$$

With g_i indicating the gas units, and p_i representing the per-unit operation cost. This cost-awareness assures the optimal gas utilization, specifically in high-volume environments.

2.6. Machine Learning Model

The ML module functions as an intelligence layer, using past trends to categorise assertions as either authentic or fake. Three methods were taken into consideration: Deep Neural Networks (DNNs), Random Forest, and XGBoost. Each assertion is represented as a feature vector of 83 dimensions.

The binary classification problem is optimized using the Binary Cross-Entropy Loss function:

$$L = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)]$$

Where y_i represents the actual label and $\hat{y}_i$ Indicates the predicted probability. The model parameters θ are updated iteratively employing Gradient Descent,

$$\theta_{new} = \theta_{old} - \eta * \nabla_\theta L$$

With η as the learning rate.

To enhance the convergence, Learning Rate Decay is applied:

$$\eta_t = \eta_0 * e^{-\lambda t}$$

Minimising the learning rate over epochs. For stability and faster convergence, Momentum is added:

$$v_t = \gamma v_{t-1} + \eta \nabla_\theta L(\theta)$$

As well as adaptive optimizers such as Adam are utilized, computed as:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1)\nabla_\theta L(\theta), \quad v_t = \beta_2 v_{t-1} + (1-\beta_2)(\nabla_\theta L(\theta))^2$$

With parameter update:

$$\theta_t = \theta_{t-1} - \eta * \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}$$

To introduce non-linearity and increase model expressiveness, activation functions are employed:

The sigmoid outputs probabilities between 0 and 1, often used in the final fraud prediction layer.

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

Rectified Linear Unit (ReLU) introduces the non-linearity and accelerates convergence by zeroing out negative values.

$$f(x) = \max(0, x)$$

Leaky ReLU is utilized to solve the ReLU's "dying neuron" issue by allowing small gradients on negative inputs.

$$f(x) = \begin{cases} x & if\ x > 0 \\ \propto x & if\ x \leq 0 \end{cases}$$

Softmax is used in extended multi-class fraud types and modifies outputs to class-wise probabilities that sum to 1.

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

For evaluating information-based feature importance, the model calculates:

- **Shannon Entropy**
  The average amount of information in a dataset is the parameter that Shannon Entropy utilises to figure out how uncertain or impure it is. As a tool for gauging a node's mixedness before a split, it aids fraud detection by directing decision trees towards characteristics that mitigate uncertainty.

$$H(X) = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i)$$

- **Gini Impurity**
  When a data point is randomly labelled in a node, the Gini Impurity measures the probability of wrong classification. The most effective feature splits are chosen in the XGBoost model to identify false claims.

$$G = 1 - \sum_{i=1}^{C} p_i^2$$

- **Information Gain**
  Information Gain is the amount of entropy that drops after a dataset is split up based on a certain trait. It assists in figuring out which factors, like the amount of the claim or how consistent the diagnosis codes are, are the most important for detecting scams.

$$IG(D, A) = H(D) - \sum_{v \in Values(A)} \frac{|D_v|}{|D|} H(D_v)$$

- **Kullback–Leibler (KL) Divergence**
  The degree to which the expected probability distribution differs from the actual label distribution is measured by the KL Divergence. For model calibration & to determine the classifier's fraud pattern capture accuracy, it is important.

$$D_{KL}(P \| Q) = \sum_{i} P(i) \log \frac{P(i)}{Q(i)}$$

The enhancement in model explainability and compliance with regulations may be achieved by using these metrics to find the most discriminative elements that contribute to fraud patterns.

3. RESULT

The performance of the proposed Blockchain-Integrated RPA framework was evaluated using a real-world dataset of 4000 healthcare insurance applications, each comprising 83 features. The research primarily focused on the model's accuracy, processing efficiency, feature significance, and blockchain logging latency.

For a more complete evaluation, we also compared our hybrid model with several well-known traditional algorithms, including CNN, RNN, and Decision Trees. The hybrid model was also compared to a number of well-known conventional algorithms, such as CNN, RNN & Decision Trees, for a more comprehensive evaluation.

3.1. Model Performance Evaluation

Insurance claims frequently contain unbalanced classes and tabular datasets, which led to the selection of the suggested ML engine and the XGBoost classifier for its resilience in both areas. The model underwent optimisation across 30 epochs utilising gradient-based approaches, with a 70:30 split for training. With a steady convergence of the training curve, the final accuracy reached 92.4%, which is a respectable level for fraud detection tasks. The confusion matrix (Figure 2) that was generated as a result of the evaluation of the suggested Blockchain-Integrated RPA framework shows how well the model can classify claims that are fraudulent from those that are legitimate. Five hundred out of four thousand applications were found to be fake, which shows that the ML section performed an effective task of figuring out complicated fraud trends from the eighty-three organised traits. 1323 real claims (True Negatives) had been correctly sorted at the same time, showing that the framework can work well in situations that aren't fake and maintain the integrity of claim handling.

The system seems to maintain operational dependability without excessively interfering with valid claims, as seen by the small number of False Positives (87), which are real claims that are mistakenly marked as fraudulent. Reducing financial loss and retaining confidence in automated decision-making is crucial, and the modest failure rate of fraud detection is reflected in the fact that just 58 fraudulent claims were incorrectly categorised as legitimate (False Negatives). The combination of DL-based fraudulent classification engines, the rule-driven pre-screening to RPA bots, and the irreversible blockchain logging that upholds data integrity & smart contract rules is primarily responsible for this balance between minimising both forms of misclassification. The framework provides an effective framework in real-time fraud detection for health insurance processes because of these elements working together to produce high accuracy & decision accountability.

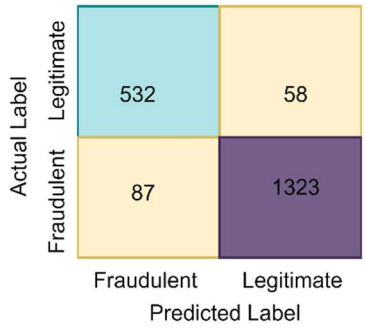

Figure 2. Confusion Matrix for Fraud Detection

Using real data, CNN and RNN models had lower classification accuracy (about 85–88%) and higher false positive rates. However, the suggested model constantly performed more effectively, with an improved true positive rate (532 frauds successfully flagged) and lower false positive rates (87). Additionally, the integration of RPA bots allowed for smooth rule-based filtering and claim intake, which are not natively supported by CNN or RNN architectures. Additionally, using blockchain technology adds an extra layer of security as well as auditability that isn't present in AI-only systems. Smart contracts require data immutability and openness. For real-time, scalable, and rule-compliant fraud detection in the healthcare insurance domain, this shows that the suggested mixed design is not only more effective but also better suited to it.

3.2. Feature Importance

The suggested approach makes use of XGBoost's integrated feature significance analysis as part of the model interpretability technique. This analysis ranks input features according to information gain, or the decrease in uncertainty (entropy) that occurs when a feature is used to divide the data. This approach is essential for directing the RPA layer & smart contract logic to concentrate on high-risk claim patterns, in addition to improving model transparency. Several factors that regularly influenced fraud detection options were identified by the feature ranking as illustrated in Figure 3.

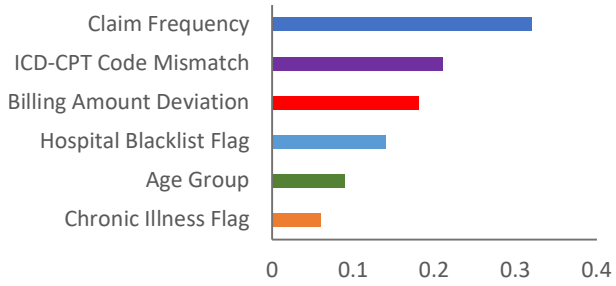

Figure 3: Top Features Influencing Fraud Detection (XGBoost Analysis)

Cases with a large number of claims in a short amount of time were strongly linked to scams, making Claim Frequency the most important factor. Most of the time, these kinds of frequency jumps mean that someone is intentionally filing multiple claims or taking advantage of the refund system in a calculated way. The International Classification of Diseases (ICD) code mismatches with Current Procedural Terminology (CPT) code mismatches were another high-impact indicator. These codes often pointed to billing for treatments that were not medically necessary, which is a common pattern of fraud in this field.

When the claimed amount was far higher than the typical treatment cost for comparable conditions, Billing Amount Deviation became a major warning indication. The RPA's rule engine automatically detected and assessed these anomalies using an ML model, which were subsequently cross-validated against pricing regulations. The hospital's blacklist flag included a binary characteristic that strongly indicated if the provider had a history of dubious billing practices. In addition to facilitating better categorisation, this flag was an essential component of smart contracts that enabled on-chain automatic audit triggers.

Finally, age group and chronic illness flags helped in assessing the credibility of long-term or high-cost claims. For example, claims involving geriatric patients with chronic conditions typically have predictable treatment paths and associated costs. Deviations from these expected patterns often signaled exaggeration or misrepresentation, which the model learned to flag reliably.

When combined, these characteristics improved the model's accuracy and closely matched actual fraud

typologies. High-importance characteristics may be implemented in RPA decision rules & blockchain smart contracts for improved automation and accountability, and their efficacy verifies the ML pipeline's architecture and supports the hybrid system's functionality.

3.3. Efficiency Comparison (Manual vs RPA Pipeline)

To rigorously assess the performance gains offered by the proposed hybrid system, a detailed time-efficiency comparison was conducted among the traditional manual review process as well as the RPA-enabled automated pipeline across 4 critical stages of healthcare insurance claim processing: data extraction & evaluation, claim triaging, rule-based tests, and fraud flagging for audit. Each stage was evaluated based on average processing times for a single claim.

Data capture and validation took about 7.5 minutes per claim in the manual review process because documents had to be found, provider and patient information had to be checked, and policy limits had to be cross-referenced. Another 4.2 minutes were devoted to claim triaging, which mostly relies on human judgement and case-specific expertise to prioritise claims according to urgency or complexity. By hand-searching ICD/CPT code references as well as insurance policy databases, rule-based examinations, which include making sure that diagnosis and treatment are consistent and finding cases of overbilling, took 5.3 minutes. Lastly, an extra 3.0 minutes were needed for audit flagging, which entails identifying claims that may need supplementary review. This process is often slowed down by subjective reviewer judgment and uneven flagging criteria. Manually processing each claim consumed an average of 20 minutes.

On the other hand, the pipeline that was enabled by RPA completed these identical steps consistently and at an astounding pace. Automated screen scraping, as well as real-time API access to hospital and policy systems, made it possible to get data and make sure it was correct in just 0.8 seconds. Automated systems used predetermined logic to claim information and historical fraud trends, allowing for 0.6-second claim triaging. Utilising deterministic processes, encoded logic inside the RPA engine and smart contracts in the blockchain, rule-based checks were conducted in 0.4 seconds. Audit flagging, which was before an arbitrary process, was automated and lowered to 0.3 seconds using ML fraud scores and blockchain-based risk assessments. This resulted in 600 times increase in operational efficiency as the entire time it took to process each claim dropped to 2 seconds.

Such massive drop-in processing time does not only increase output, but also allows scamming to be recognized and the decision made immediately, which is particularly critical in large public insurance programs and in businesses with thousands of cases per day. The RPA-based system is significantly more scalable, accessible, and flexible because it removes the bottlenecks of the operations behavior created by people and ensures the enforcement of the fraud regulations regularly by use of blockchain smart contracts.

3.4. Time of Blockchain Transactions logging

The pillar of the suggested hybrid detection framework that will be essential is the adoption of the permissioned blockchain network to provide the transparency, accountability, and auditability of the whole decision-making pipeline Hyperledger Fabric. The final determination, including pertinent information (such as the claim ID, prediction rating, timestamps & reviewer signature), is safely hashed and recorded as a transaction in the blockchain ledger after the processing of each healthcare insurance claim, either labelled as fraudulent or cleared as authentic.

From Figure 4, it is evident that an average of 1.7 seconds per claim was found for the transaction writing time, which encompasses consensus, smart contract implementation & ledger updating. The modular consensus and effective ordering service of Hyperledger Fabric, which performs better than public blockchains for latency-sensitive applications, allows for this quick writing speed. It is also ideal for real-time fraud detection and compliance audits, since the transaction retrieval time, which is the average time required to query and get a claim's decision record or audit trail, was 1.2 seconds.

Journal of Theoretical and Applied Information Technology
28th February 2026. Vol.104. No.4

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

Crucially, these latency numbers show that the RPA-ML pipeline has very little delay due to blockchain logging. The performance bottleneck is not created by the extra 1–2 seconds needed for blockchain logging, as RPA and ML inference finish the claim categorisation and decision-making process in less than 2 seconds. On the contrary, this minimal overhead is a worthwhile trade-off for the significant benefits in terms of data immutability, non-repudiation, and tamper-proof audit logging.

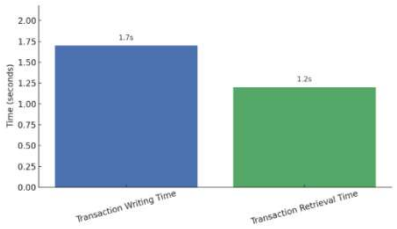

Figure 4. Blockchain Transaction Logging & Retrieval Time

The decision cannot be changed later on in the blockchain since each entry is cryptographically verified and timestamped. People who are trying to commit insurance scams would probably try to change logs or change the results of decisions if this tool weren't there. Hyperledger Fabric's decentralised design makes it much more resistant to hacking or insider manipulation that affects centralised systems. In order to provide end-to-end traceability, strengthen institutional confidence, and facilitate smooth regulatory compliance, the proposed approach involves anchoring the result of each claim to a distributed ledger.

Lastly, real-time auditability and operational integrity through entry into blockchain-based transaction tracking is used to improve the high throughput of the fraud detection system. This is the main feature of the sensitive sphere such as healthcare and insurance, which makes the structure technically robust and legally justifiable.

3.5. System Scalability and integrity.

The system scalability of the proposed Blockchain-Integrated RPA framework and integrity was adequately evaluated by concurrent batches of 100, 500, and 1000 data claims as shown in Figure 5. As batch sizes increased, the average processing durations only slightly increased from 1.9 to 2.5 seconds, and there were no failure rates in all of the cases. Strong throughput as well as fault tolerance under heavy load were shown by the notable absence of packet loss, theft of information, and system queuing delays. These findings prove that the framework can grow horizontally without compromising performance, making it ready for implementation in national healthcare insurance systems that manage many claims simultaneously.

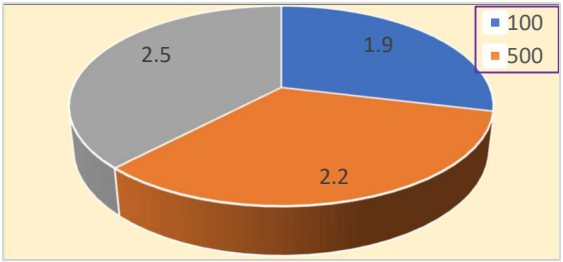

Figure 5: Scalability and Integrity of Proposed System

3.6. Comparison with Traditional Models

The XGBoost-based classifier, which is part of the Blockchain-RPA framework, was tested against a number of conventional ML models as well as models employing the identical dataset of 4000 health insurance applications using 83 features in order to assess the efficacy of the suggested hybrid methodology. Despite a mean training time of 22 seconds, the XGBoost model demonstrated the ideal balance between computational efficiency and predictive capability, achieving the greatest accuracy of 92.4% from Figure 6. The Decision Tree model was especially prone to overfitting, which reduced its reliability for generalised fraud detection. In comparison, simpler models such as Logistic Regression & Decision Tree were quicker but showed less accurate results (84.2% and 86.7%, respectively).

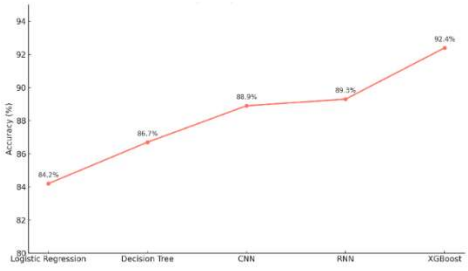

Figure 6 Model Accuracy Comparison for Fraud Detection

Accuracy rates of 89.3% and 88.9%, respectively, were shown by DL models including RNNs and CNNs, among others. But they needed a lot more time to train (40–56 seconds), and they weren't as well adapted to the highly tabular, organised structure of claim data. In regulatory situations that need clear audit trails, their limited value is due to the interpretability issues caused by their complicated designs. The XGBoost approach, on the other hand, has a readily comprehensible decision-making framework that integrates with RPA bots to enforce rules as well as the blockchain layer to store immutable records. Intelligent healthcare insurance claim fraud detection using the suggested hybrid architecture is preferable due to its combination of high accuracy, minimal latency, operational interpretability & scalability.

4. DISCUSSION

The suggested Blockchain-Integrated RPA Framework combines blockchain, automation, and ML to provide a scalable and reliable solution for healthcare insurance fraud detection. This hybrid framework makes use of the deterministic accuracy of RPA, the learning power of XGBoost-based classifiers, as well as the tamper-proof auditability for permissioned blockchain networks, in contrast to conventional systems that mainly depend on manual auditing or monolithic AI architectures. Real-time analysis, transparency, traceability & interpretability are important issues in fraud detection that are addressed by this collaborative design.

This approach is seen as technically solid and operationally efficient due to its high classification accuracy (92.4%) as well as minimal latency (<2 seconds includes blockchain writing time). By automating data collecting and early triaging, RPA bots drastically reduce processing time (from 20 minutes to 2 seconds per claim). Smart contracts automate audit triggers depending on ML risk assessments by enforcing deterministic logic. Additionally, XGBoost's feature significance improves explainability, which is a crucial part of compliance for the healthcare industry. Automation and audit rules are both enriched by feature metrics based on entropy and divergence, which in turn improve fraud signal detection.

DL models such as RNN and CNN are pretty accurate, but are unable to be used in real-time, regulation-heavy settings because they are "black boxes" and take a long time to train (40–56s). It's faster to train traditional models like DT as well as LR, but they don't do as well at finding scams or remembering them. It's possible for insurance communities to use the hybrid model on a large scale because it hits the perfect mix between being able to guess well, being easy to understand, and using minimal computing power. Table 1 describes the comparison of the suggested hybrid framework with that of the traditional methods.

Table 1: SOTA Analysis

Model/Framework	Accuracy	Average Training Time	Interpretability	Suitability for Tabular Data	Blockchain Integration
Proposed Hybrid (XGBoost + RPA + Blockchain)	92.4%	22s	High	Excellent	Yes
DT	86.7%	8s	High	Moderate	No
LR	84.2%	6s	High	Good	No
CNN	88.9%	56s	Low	Poor	No
RNN	89.3%	40s	Low	Fair	No
RF	87.6%	24s	Moderate	Good	No
XGBoost (Standalone)	91.8%	21s	High	Excellent	No

The suggested framework has several restrictions, but it performs well overall. The study may not apply to other healthcare

systems or policy environments because it was done on a single structured dataset with four thousand claims. There are currently no ways to differentiate between different kinds of fraud; instead, the implementation is based on binary fraud classification. Blockchain technology does increase auditability, but it also increases deployment costs and complicates infrastructure. Predefined RPA rules and fraud thresholds may need adjusting regularly to accommodate changing fraud tendencies. Research in the future should investigate multi-class fraud detection, validate the framework on bigger datasets from several sources, and evaluate the operational costs and scalability of real-world installations over the long term.

5. CONCLUSION

The goal of this research was to create a system for healthcare insurance claim fraud detection that is precise, transparent, auditable, and operable in real time. The results show that these goals were met to a large extent: the suggested RPA framework with blockchain integration achieved 92.4% classification accuracy, processed claims in 2.5 seconds or less, and made sure that decision logs couldn't be tampered with. The approach demonstrated superior prediction performance, interpretability, and operational practicality when compared to conventional ML and deep learning models. The practical applicability of the approach was confirmed through the direct application of the XGBoost feature importance and smart contracts to directly address regulatory and transparency needs. Even though the scalability and efficiency demonstration was productive, additional validation on diverse datasets should be carried out. Combined with other findings, the outcomes indicate that the future of healthcare insurance fraud detection can be reached with the use of blockchain technology alongside explainable intelligence and automation.

The framework is highly accurate, low-latent and auditable, research holes exist. There is no evidence of the model tested on multi-source, cross-institutional, and unstructured claim data; one structure, a dataset is tested. Binary classification not only lacks the categories of fraud, but also creates fraud tactics. Also, the adaptive learning approaches to updating dynamically the RPA rules and level of fraud were not studied. Future studies should investigate multi-class fraud detection, concept drift management, blockchain deployment cost-benefit analysis, and pilot applications to determine the cost-effectiveness and scalability of the method in the long-term.

The authors conceptualized and designed the study together. One author designed the approach, installed the system and conducted the experiments whereas the other monitored, validated and critically evaluated the text. All authors reviewed data, analyzed findings, edited articles, and gave final consent to the article.

REFERENCES:

[1] D. A. Patil and S. G, “A comprehensive survey on securing the social internet of things: protocols, threat mitigation, technological integrations, tools, and performance metrics,” *Sci. Rep.*, vol. 15, no. 1, p. 40190, 2025.

[2] M. A. Mohammed, M. Boujelben, and M. Abid, “A novel approach for fraud detection in blockchain-based healthcare networks using machine learning,” *Future Internet*, vol. 15, no. 8, p. 250, 2023.

[3] C. M. Selvamuthu, B. Lavaraju, and A. Sundaram, “A novel approach of streamlining claims processing and fraud prevention in health insurance through blockchain technology,” in *2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, IEEE, 2024, pp. 611–618. Accessed: Jan. 06, 2026. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10714863/

[4] S. Gupta, “AI, BLOCKCHAIN, AND AUTONOMOUS INNOVATION,” 2025, Accessed: Jan. 06, 2026. [Online]. Available: https://www.researchgate.net/profile/Shubha

m-Gupta-41/publication/391908172_AI_BLOCKCHAIN_AND_AUTONOMOUS_INNOVATION_Charting_the_Future_of_Intelligent_Enterprises/links/682ce6b7026fee1034f95eff/AI-BLOCKCHAIN-AND-AUTONOMOUS-INNOVATION-Charting-the-Future-of-Intelligent-Enterprises.pdf

[5] S. K. Jena, B. Kumar, B. Mohanty, A. Singhal, and R. C. Barik, "An advanced blockchain-based hyperledger fabric solution for tracing fraudulent claims in the healthcare industry," *Decis. Anal. J.*, vol. 10, p. 100411, 2024.

[6] F. S. Islam, "Artificial intelligence-driven smart waste-to-energy networks for climate-resilient circular resource management in vulnerable megacities," *Int. J. Environ. Clim. Change*, vol. 15, no. 7, pp. 381–415, 2025.

[7] S. Mewada *et al.*, "Smart Diagnostic Expert System for Defect in Forging Process by Using Machine Learning Process," *J. Nanomater.*, vol. 2022, no. 1, p. 2567194, Jan. 2022, doi: 10.1155/2022/2567194.

[8] K. Kapadiya *et al.*, "Blockchain-assisted healthcare insurance fraud detection framework using ensemble learning," *Comput. Electr. Eng.*, vol. 122, p. 109898, 2025.

[9] A. Chakraborty, G. Singh, V. Sirvastava, and S. A. Dhondiyal, "Blockchain-Enhanced Adversarial Machine Learning for Fraud Detection and Claims Automation in the Insurance Sector," in *2024 5th International Conference on Data Intelligence and Cognitive Informatics (ICDICI)*, IEEE, 2024, pp. 80–86. Accessed: Jan. 06, 2026. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10810927/

[10] N. Sharma and N. Jindal, "Emerging artificial intelligence applications: metaverse, IoT, cybersecurity, healthcare - an overview," *Multimed. Tools Appl.*, vol. 83, no. 19, pp. 57317–57345, Dec. 2023, doi: 10.1007/s11042-023-17890-6.

[11] A. A. Deshmukh *et al.*, "Event-based Smart Contracts for Automated Claims Processing and Payouts in Smart Insurance.," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 4, 2024, Accessed: Jan. 06, 2026. [Online]. Available: https://pdfs.semanticscholar.org/a7f3/8b3c3478b3a51f989869999de8ae198e530f.pdf

[12] T. J. Olorunlana, "Harnessing Technology for Effective Fraud Detection: Tools, Trends, and Case Studies", Accessed: Jan. 06, 2026. [Online]. Available: https://www.researchgate.net/profile/Taiwo-Olorunlana/publication/392917128_Harnessing_Technology_for_Effective_Fraud_Detection_Tools_Trends_and_Case_Studies/links/6858bd73e8fa0f5c2825bfc6/Harnessing-Technology-for-Effective-Fraud-Detection-Tools-Trends-and-Case-Studies.pdf

[13] N. N. I. Prova, "Healthcare fraud detection using machine learning," in *2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI)*, IEEE, 2024, pp. 1119–1123. Accessed: Jan. 06, 2026. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10696476/

[14] Z. Malveen, "Healthcare Insurance Fraud Detection Using AI", Accessed: Jan. 06, 2026. [Online]. Available: https://www.researchgate.net/profile/Zenith-Malveen/publication/391452611_Healthcare_Insurance_Fraud_Detection_Using_AI/links/6818b895ded43315574259e1/Healthcare-Insurance-Fraud-Detection-Using-AI.pdf

[15] O. A. Bello, A. Folorunso, O. E. Ejiofor, F. Z. Budale, K. Adebayo, and O. A. Babatunde, "Machine learning approaches for enhancing fraud prevention in financial transactions," *Int. J. Manag. Technol.*, vol. 10, no. 1, pp. 85–108, 2023.

[16] A. S. Mavai, D. K. Mishra, A. Sharma, and S. Bansal, "Role of big data analysis to improve the efficiency of healthcare system using machine learning techniques," *Discov. Internet Things*, vol. 5, no. 1, p. 153, Dec. 2025, doi: 10.1007/s43926-025-00225-2.

[17] P. Patro, R. Azhagumurugan, R. Sathya, K. Kumar, T. R. Kumar and M. V. S. Babu, "A hybrid approach estimates the real-time health state of a bearing by accelerated degradation tests, Machine learning," 2021 Second International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), Bengaluru, India, 2021, pp. 1-9, doi: 10.1109/ICSTCEE54422.2021.9708591.

[18] M. Malempati, *The Intelligent Ledger: Harnessing Artificial Intelligence, Big Data, and Cloud Power to Revolutionize Finance, Credit, and Security*. Deep Science Publishing, 2025. Accessed: Jan. 06, 2026. [Online]. Available: https://books.google.com/books?hl=en&lr=&id=jqxaEQAAQBAJ&oi=fnd&pg=PA1&dq=REAL-

TIME+INTELLIGENT+CLAIM+VERIFICATION+USING+BLOCKCHAIN-BASED+RPA+AND+XGBOOST+FOR+SCALABLE+HEALTHCARE+FRAUD+DETECTION&ots=SYLoCa3Op8&sig=67iNzXzJkDkX2OZAqCBCjI-QtjI

[19] R. K. Tulala, P. K, and B. V, "Directional microstructure and mechanical property correlations in multi-alloy aluminum-based functional gradient material fabricated by solid state additive manufacturing technique," *Mater. Res. Express*, vol. 12, no. 11, p. 116502, Nov. 2025, doi: 10.1088/2053-1591/ae171a.

[20] V. Mishra, S. Parakh, and V. Viradia, "Transfigurations of Healthcare Insurance (Payers) Claims with Artificial Intelligence: An Extensive Literature Review", Accessed: Jan. 06, 2026. [Online]. Available: https://www.researchgate.net/profile/Vineet-Mishra-12/publication/393446479_Transfigurations_of_Healthcare_Insurance_Payers_Claims_with_Artificial_Intelligence_An_Extensive_Literature_Review/links/686a4094e4632b045dca2b64/Transfigurations-of-Healthcare-Insurance-Payers-Claims-with-Artificial-Intelligence-An-Extensive-Literature-Review.pdf