Communication in Physical Sciences, 2026, 13(12):1994-2016

Applying Machine Learning and Cloud Computing for Sustainable Healthcare Delivery: Detecting Billing Fraud in the National Health Insurance Scheme in Nigeria Dibiaezue

Ngozi Fidelia Dibiaezue and Francis Sunday Bakpo
Received: 27 August 2026/ Accepted: 21 October 2026/Published: 05 November 2026
DOI: 10.4314/cps.v13i12.9

Abstract: *The increase in fraudulent activities in Nigeria's National Health Insurance Scheme (NHIS) especially fraud by healthcare providers has led to huge financial losses, reduced operational efficiency, and challenges in the provision of sustainable health services. Traditional fraud detection mechanisms in claims management based on manual auditing, static business rules and periodic sampling are slow, labour-intensive and unable to scale with the volume and velocity of electronic claims generated under a digitised health insurance scheme. The paper proposes a machine learning framework hosted on the cloud for automated real time detection of billing fraud in the NHIS claims. The framework combines a supervised ensemble learning model (Random Forest and Gradient Boosting, benchmarked against Logistic Regression and Support Vector Machines) with an anomaly-detection layer to identify both known fraud typologies and previously unseen outlier behaviour, deployed on a scalable cloud-computing infrastructure to enable elastic processing of claims data from distributed healthcare facilities. After pre-processing the historical claims records dataset by cleaning, feature engineering and class-imbalance correction using the Synthetic Minority Oversampling Technique (SMOTE), the models were trained, validated and evaluated using accuracy, precision, recall, F1-score and Area Under the Receiver Operating Characteristic Curve (AUC-ROC). The experimental results demonstrate that the ensemble model deployed on the cloud architecture outperforms the rule-based baseline and individual classical classifiers in terms of fraud-detection performance with significantly higher recall and AUC-ROC. The cloud deployment also reduces the claim-processing latency and improves the horizontal scalability with increasing claim load. The study concludes that the use of machine learning with cloud computing provides a sustainable, cost-effective and scalable pathway to curb billing fraud, thereby protecting the financial integrity of National Health Insurance Schemes and bolstering the broader goal of sustainable healthcare delivery. It recommends policy adoption, data governance and future integration of explainable artificial intelligence (XAI) and federated learning into national claims-management systems.*

Keywords: *Machine Learning; Cloud Computing; Health Insurance; Billing Fraud Detection; Sustainable Healthcare; Anomaly Detection; National Health Insurance Scheme (NHIS)*

Ngozi Fidelia Dibiaezue*
Department of Software Engineering, University on the Niger, Umunya, Anambra State, Nigeria Department of Computer Science, University of Nigeria, Nsukka, Enugu State, Nigeria
Email: ngozi.dibiaezue@uniniger.edu.ng
https://orcid.org/0000-0002-1825-0097

Francis Sunday Bakpo
Department of Computer Science, University of Nigeria, Nsukka, Enugu State, Nigeria
Email: francis.bakpo@unn.edu.ng

1.0 Introduction

Health insurance constitutes an important mechanism for achieving Universal Health Coverage (UHC) by facilitating access to essential healthcare services while protecting individuals and households against catastrophic out-of-pocket health expenditures (World Health Organization

[WHO], 2021). Through risk pooling and resource redistribution, health insurance schemes enable governments and other stakeholders to mobilise and allocate financial resources for the provision of affordable and equitable healthcare services. A critical component of the effective operation of these schemes is the claims-management and reimbursement system, through which accredited healthcare providers receive payments for services rendered to insured individuals. Consequently, the accuracy, transparency, efficiency and accountability of claims processing are essential for maintaining public confidence, ensuring financial sustainability and promoting uninterrupted healthcare delivery.

In Nigeria, the National Health Insurance Scheme (NHIS) was established to improve access to affordable healthcare and reduce the financial burden associated with medical expenses. The institutional framework was subsequently strengthened through the enactment of the National Health Insurance Authority Act, 2022, which transformed the former NHIS into the National Health Insurance Authority (NHIA) and expanded the regulatory framework for health insurance coverage in the country. Under this arrangement, healthcare providers accredited by the relevant health insurance organisations submit claims for services delivered to enrollees, which are subsequently reviewed and reimbursed in accordance with established procedures. However, the increasing complexity of healthcare transactions, the expansion of health insurance coverage and the growing adoption of digital claims-management systems have created additional demands for effective monitoring, verification and financial accountability. Weaknesses in claims verification and reimbursement processes can expose health insurance institutions to fraudulent activities, resulting in financial leakages, reduced operational efficiency and the diversion of resources from legitimate healthcare services.

One of the major threats to the financial integrity of health insurance systems is fraudulent billing, which involves the deliberate misrepresentation of healthcare services, procedures or associated costs to obtain unauthorised financial benefits. Common forms of healthcare billing fraud include upcoding, whereby providers charge for more expensive services than those actually rendered; unbundling, which involves separately billing procedures that should be charged as a single service; phantom billing, involving claims for services never provided; duplicate claims; and identity-related fraud involving the unauthorised use of enrollee credentials (National Health Care Anti-Fraud Association [NHCAA], 2022; Bauder & Khoshgoftaar, 2018). Fraudulent activities may also involve collusion between healthcare providers and beneficiaries, manipulation of diagnosis codes, and the submission of inflated claims. International evidence indicates that fraud, waste and abuse constitute substantial sources of financial losses in healthcare financing systems, particularly where monitoring and verification mechanisms are inadequate (European Healthcare Fraud and Corruption Network [EHFCN], 2021). In developing economies such as Nigeria, where healthcare financing resources are constrained and the demand for accessible medical services continues to increase, such financial leakages can undermine the effectiveness and sustainability of health insurance programmes.

Traditional approaches to detecting healthcare billing fraud rely largely on manual auditing, periodic sampling, statistical examination and predetermined rule-based mechanisms. Although these approaches provide a foundation for identifying suspicious transactions, they are often reactive, labour-intensive and limited in their capacity to process large volumes of claims efficiently (Joudaki et al., 2015). Manual auditing requires considerable human resources and may result in delays in identifying fraudulent transactions,

particularly when claims originate from geographically distributed healthcare facilities. Similarly, static rule-based systems depend on predefined indicators and thresholds, making them less effective in identifying sophisticated fraudulent activities that do not conform to established patterns. Fraudulent actors may also modify their billing practices to circumvent known detection rules, thereby reducing the effectiveness of conventional approaches. As the volume and complexity of electronic healthcare claims increase, the limitations of these methods become more pronounced, creating the need for intelligent, automated and scalable fraud-detection mechanisms (Bauder et al., 2017).

Machine learning (ML) has emerged as a promising technological approach for addressing these challenges through its ability to identify complex patterns, relationships and irregularities within large datasets. Unlike conventional rule-based techniques, machine learning algorithms can learn from historical claims records and distinguish between legitimate and potentially fraudulent transactions based on multiple characteristics, including provider identification, diagnosis codes, treatment procedures, claim amounts and submission frequencies. Supervised learning algorithms, such as Logistic Regression, Decision Trees, Random Forest, Support Vector Machines and Gradient Boosting, have demonstrated considerable potential for detecting fraudulent transactions in complex datasets. Ensemble learning techniques are particularly relevant because they combine the predictive capabilities of multiple decision models to improve classification performance and identify non-linear relationships among variables (Herland et al., 2018). Bauder and Khoshgoftaar (2018) also demonstrated the effectiveness of Random Forest approaches in healthcare fraud detection, highlighting their potential to improve the identification of suspicious claims compared with conventional detection techniques.

Despite the advantages of supervised machine learning, its effectiveness may be constrained by the limited availability of accurately labelled fraudulent claims. In most healthcare insurance datasets, legitimate transactions substantially outnumber fraudulent ones, resulting in class imbalance that can cause predictive models to favour the majority class while failing to identify a significant proportion of fraudulent transactions. Chawla et al. (2002) addressed this challenge through the development of the Synthetic Minority Oversampling Technique (SMOTE), which generates synthetic observations of minority-class instances to improve the representation of underrepresented categories during model training. Sharma and Panigrahi (2013) similarly highlighted the limitations of relying solely on classification accuracy when evaluating fraud-detection models, noting that high overall accuracy may conceal poor detection of fraudulent transactions. Consequently, precision, recall, F1-score and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) are important evaluation measures for assessing the effectiveness of fraud-detection models, particularly where fraudulent claims constitute a relatively small proportion of the dataset.

Furthermore, unsupervised and semi-supervised machine learning techniques offer additional opportunities for identifying fraudulent activities that may not have been previously documented or labelled. Techniques such as k-means clustering, DBSCAN, autoencoders and Isolation Forests can identify unusual patterns and deviations from normal claims behaviour without depending entirely on predefined fraud classifications (Liu et al., 2008). Zafari and Ekin (2019) demonstrated the potential of topic-modelling approaches for identifying hidden behavioural patterns associated with prescription fraud, while Ngufor and Wojtusiak (2013) explored hybrid analytical approaches for improving the identification of healthcare fraud. These developments suggest that combining supervised

classification with anomaly-detection techniques could provide a more comprehensive approach to fraud detection by identifying both established fraud patterns and previously unrecognised suspicious activities. In addition, effective feature engineering, involving the development of provider-level transaction indicators, claims-submission patterns, service utilisation frequencies and relationships among providers, enrollees and diagnosis codes, can further improve the discriminatory capability of predictive models (Thornton et al., 2013).

Although machine learning provides the analytical capability required for intelligent fraud detection, its practical implementation within national health insurance systems also depends on the availability of reliable, scalable and secure computing infrastructure. Cloud computing offers an appropriate technological environment for addressing these requirements through on-demand access to computing resources, distributed data storage, flexible processing capacity and integration with advanced analytical applications. Healthcare information systems can leverage cloud infrastructure to manage large volumes of electronic health records, facilitate information exchange among geographically distributed facilities and support computationally intensive analytical operations. Kaur and Singh (2020) identified the potential of cloud computing to improve healthcare information management through resource flexibility, accessibility and reduced dependence on expensive on-premises infrastructure. Similarly, Sadiq et al. (2017) highlighted the advantages of cloud platforms in providing elastic computing and storage resources, reducing infrastructure expenditure and supporting the integration of sophisticated analytical applications into healthcare data-management systems. These capabilities make cloud computing particularly relevant to national health insurance systems that require continuous processing of claims originating from multiple healthcare facilities.

The integration of machine learning with cloud computing therefore offers an opportunity to develop intelligent and scalable healthcare fraud-detection systems. Machine learning provides the analytical capability to identify suspicious billing patterns, while cloud computing supplies the computational infrastructure for processing large volumes of claims, deploying predictive models and supporting continuous monitoring. Such integration can potentially reduce claims-processing delays, improve fraud-detection efficiency and facilitate the extension of automated verification services across geographically distributed healthcare providers. Nevertheless, cloud-based healthcare systems introduce important considerations relating to data privacy, cybersecurity, regulatory compliance, network-dependent processing latency and vendor lock-in (Kaur & Singh, 2020). These concerns are particularly important in healthcare environments where claims records contain sensitive personal and medical information. Therefore, the development of an effective framework must consider not only predictive accuracy and computational scalability but also the security, reliability and sustainability of the underlying infrastructure.

Despite the growing body of research on machine learning applications in healthcare fraud detection (Bauder & Khoshgoftaar, 2018; Herland et al., 2018) and the application of cloud computing in healthcare information systems (Kaur & Singh, 2020; Sadiq et al., 2017), important research gaps remain. Existing studies have largely examined fraud classification techniques and cloud-based healthcare infrastructure as separate technological concerns, with comparatively limited attention given to their integrated implementation within national health insurance claims-management environments in developing economies. Furthermore, the challenges associated with class imbalance, the detection of previously unseen fraudulent behaviour, real-time claims processing and the scalability of fraud-detection systems have not been sufficiently addressed within the specific operational context of Nigeria's health



insurance system. There is therefore a need for an integrated framework that combines supervised machine learning, anomaly detection and cloud-based deployment to improve the identification of fraudulent healthcare claims while supporting efficient and scalable claims processing. Addressing this gap is particularly important for Nigeria, where strengthening the accountability and efficiency of health insurance administration is essential for improving healthcare financing and supporting sustainable service delivery.

Against this background, the aim of this study is to develop and evaluate an integrated machine learning and cloud-computing framework for detecting fraudulent healthcare billing claims within Nigeria's National Health Insurance Scheme, with a view to improving fraud-detection performance, claims-processing efficiency and system scalability. Specifically, the study seeks to: (i) examine the limitations of conventional rule-based and manual auditing approaches to healthcare billing fraud detection; (ii) develop a hybrid fraud-detection model combining supervised ensemble learning algorithms, particularly Random Forest and Gradient Boosting, with anomaly-detection techniques; (iii) address class imbalance in historical claims data through appropriate preprocessing and the application of SMOTE; (iv) evaluate the performance of the proposed model against selected conventional machine learning classifiers using accuracy, precision, recall, F1-score and AUC-ROC; and (v) implement and assess a cloud-based deployment architecture in terms of claims-processing latency, computational efficiency and scalability under increasing transaction loads.

The significance of this study lies in its potential contribution to the technological improvement and financial sustainability of healthcare insurance administration in Nigeria. From a technological perspective, the proposed framework provides an integrated approach to automated fraud detection by combining predictive machine learning models with cloud-based computational infrastructure. Institutionally, it could support the NHIA, health insurance administrators and accredited healthcare providers in strengthening claims verification, identifying suspicious transactions and improving reimbursement accountability. Economically, more effective detection of fraudulent claims could help minimise financial leakages and ensure that limited healthcare resources are directed towards legitimate medical services. Furthermore, the framework could provide useful insights for policymakers and healthcare administrators seeking to strengthen digital healthcare governance, improve data-driven decision-making and promote transparency in health insurance operations. By supporting more efficient utilisation of healthcare financing resources, the study contributes to the broader objective of sustainable healthcare delivery and the achievement of Universal Health Coverage in Nigeria.

2.0 Research Methodology

This study adopted the Design Science Research (DSR) methodology to develop and evaluate an innovative technological solution for addressing healthcare billing fraud within Nigeria's National Health Insurance Scheme (NHIS). Design Science Research is particularly appropriate for applied computing and information systems studies because it focuses on designing, developing and evaluating technological artefacts intended to solve practical organisational problems. Unlike conventional research approaches that primarily seek to describe or explain existing phenomena, DSR emphasises the creation and systematic assessment of solutions that improve existing practices. In this study, the principal artefact is an integrated machine learning and cloud-computing framework designed to detect fraudulent healthcare billing claims and improve the efficiency, scalability and reliability of claims processing.

The research followed the five-stage DSR process proposed by Peffers et al. (2007), comprising problem identification and

motivation, definition of solution objectives, design and development, demonstration and evaluation, and communication of results. The problem-identification stage examined the limitations of conventional NHIS claims-verification procedures, particularly their dependence on manual auditing, periodic sampling and static rule-based detection mechanisms. The solution objectives were subsequently established around improving fraud-detection accuracy, reducing claims-processing delays and providing a scalable computational infrastructure. The design and development stage involved constructing the machine learning models, database architecture and cloud-based processing framework. The demonstration and evaluation stages focused on assessing the predictive performance of the developed models and comparing the operational performance of the proposed cloud deployment with an on-premises baseline. The final stage involved documenting the framework, its implementation and the findings obtained from its evaluation.

2.1 Research Design and Philosophical Orientation

The study adopted a post-positivist philosophical orientation, which recognises the existence of an objective reality that can be investigated through systematic observation, measurement and empirical evaluation. This orientation is appropriate because the effectiveness of the proposed fraud-detection framework can be assessed using measurable indicators of classification performance and system efficiency.

A quantitative experimental research design was employed to evaluate the predictive performance of the selected machine learning algorithms. The models were trained and tested using historical healthcare claims records, while their performance was assessed using standard classification metrics, including accuracy, precision, recall, F1-score and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC).

A comparative evaluation was conducted to determine the relative effectiveness of the proposed hybrid model against conventional machine learning classifiers and a rule-based baseline.

In addition, a comparative systems-evaluation approach was adopted to assess the performance of the cloud-based infrastructure. The proposed cloud deployment was compared with an equivalent on-premises baseline under controlled simulated workload conditions. The comparison focused on average claim-scoring latency, processing throughput, performance under peak workloads, processing backlog and infrastructure cost characteristics. This approach provided a basis for evaluating the extent to which cloud computing could improve the operational efficiency and scalability of the proposed fraud-detection framework.

2.2 Data Collection and Sources

The study utilised secondary data comprising historical healthcare claims records obtained from Nigeria's National Health Insurance Scheme. The dataset contained claim-level information representing transactions submitted by accredited healthcare providers for services rendered to enrolled beneficiaries. The principal variables included enrollee identification, healthcare provider identification, diagnosis codes, procedure codes, billed amounts, dates of service, dates of claim submission and binary classification labels derived from previously confirmed audit outcomes.

In addition to the historical claims records, relevant scholarly publications were reviewed to guide the selection of machine learning algorithms, feature-engineering techniques, data preprocessing procedures and approaches for addressing class imbalance in healthcare fraud detection. The methodological framework was informed by previous studies on conventional claims auditing, machine learning applications and healthcare fraud-detection systems (Joudaki et al., 2015; Herland et al., 2018; Bauder & Khoshgoftaar, 2018).

No primary data were collected directly from human participants. The study relied on

existing claims records, subject to applicable institutional data-governance and confidentiality requirements. Since healthcare claims contain sensitive personal and medical information, appropriate safeguards for data confidentiality, restricted access and secure processing were considered in the proposed system design.

2.2.1 Study Area

The study covered accredited Federal, State and privately owned hospitals and healthcare centres participating in the NHIS across Nigeria's six geopolitical zones: North-Central, North-East, North-West, South-East, South-South and South-West. The geographical coverage was intended to accommodate claims originating from healthcare facilities operating under different administrative, institutional and geographical conditions.

The inclusion of healthcare providers across the six geopolitical zones provides a broader operational context for evaluating the proposed framework, particularly its capacity to process claims originating from geographically distributed facilities. This consideration also informed the adoption of cloud computing as the underlying deployment infrastructure.

2.2.2 Dataset Description and Database Design

The dataset comprised historical NHIS claims records supplemented with provider-level and enrollee-level aggregate variables generated during feature engineering. Each claim was associated with a binary outcome indicating whether it was classified as fraudulent or legitimate based on previously confirmed audit findings. The dataset exhibited class imbalance, with confirmed fraudulent claims constituting a relatively small proportion of the total records, a characteristic commonly reported in healthcare fraud-detection research (Herland et al., 2018; Sharma & Panigrahi, 2013).

To support efficient data management, claims verification, model prediction and audit feedback, a relational database structure was designed around five principal entities: Claim, Enrollee, Provider, AuditDecision and ModelVersion. These entities establish the relationships required to connect submitted claims with beneficiary records, healthcare providers, audit outcomes and machine learning model information.

Table 1: Database Entities, Key Attributes and Descriptions for the Proposed Fraud Detection System

Entity	Key Attributes	Description
Claim	ClaimID (PK), EnrolleeID (FK), ProviderID (FK), DiagnosisCode, ProcedureCode, BilledAmount, ServiceDate, SubmissionDate, RiskScore, Classification, Status	Stores each submitted healthcare claim and its associated machine learning output.
Enrollee	EnrolleeID (PK), Name, DateOfBirth, EnrolmentDate, EligibilityStatus	Stores enrollee demographic information and eligibility records.
Provider	ProviderID (PK), FacilityName, AccreditationStatus, Specialty, Location	Maintains records of accredited healthcare providers and their institutional details.
AuditDecision	AuditID (PK), ClaimID (FK), AuditorID (FK), ConfirmedOutcome, AuditDate, Comments	Records auditor decisions and confirmed fraud outcomes for model evaluation and retraining.
ModelVersion	ModelVersionID (PK),	Tracks deployed

	TrainingDate, AlgorithmType, PerformanceMetrics, DeploymentStatus	and historical machine learning model versions and their performance.

Note. PK = Primary Key; FK = Foreign Key. ML = Machine Learning.

2.As presented in Table 1, the database structure provides an integrated mechanism for managing the information required by the proposed fraud-detection system. The Claim entity serves as the central transaction record, linking individual healthcare claims to their corresponding enrollees and providers. The Enrollee and Provider entities provide the demographic, eligibility and accreditation information necessary for validating claims. The AuditDecision entity captures the outcomes of subsequent human verification, thereby providing feedback for evaluating and updating the machine learning models. The ModelVersion entity maintains records of model training, performance and deployment status, supporting model tracking and reproducibility. Collectively, these entities provide the relational foundation for integrating claims management, fraud detection and continuous model improvement.

2.2.3 Data Preprocessing and Feature Engineering

Data preprocessing was undertaken to improve data quality, ensure consistency across the input variables and prepare the claims records for machine learning analysis. The process involved six major stages: data cleaning, feature engineering, categorical encoding, numerical scaling, class-imbalance correction and dataset partitioning.

Data cleaning: Duplicate claims records were identified and removed to prevent repeated transactions from distorting model training and evaluation. Records containing missing critical variables, particularly enrollee identification, provider identification and billed amounts, were subjected to appropriate imputation or exclusion procedures. Malformed date entries and inconsistent currency values were also corrected to ensure uniformity across the dataset.

Feature engineering: Additional predictive variables were derived from the original claims records to improve the ability of the models to distinguish between legitimate and potentially fraudulent transactions. These included provider-level average claim values, provider claim frequencies over rolling time windows, enrollee claim frequencies, diagnosis-procedure co-occurrence consistency scores, the ratio of billed amounts to standard procedure tariffs, and the time interval between service delivery and claim submission. These features were selected to capture unusual billing behaviour, inconsistent service patterns and deviations from established provider practices, consistent with the feature-engineering approaches discussed by Thornton et al. (2013).

Categorical encoding: Categorical variables, including diagnosis codes, procedure codes and healthcare facility specialties, were transformed into numerical representations using appropriate encoding techniques. One-hot encoding and target encoding were considered according to the characteristics of the variables and the requirements of the selected algorithms.

Numerical scaling: Continuous numerical variables were standardised for Logistic Regression and Support Vector Machine models because their predictive performance can be influenced by differences in feature magnitude. Tree-based algorithms, including Random Forest, Gradient Boosting and Isolation Forest, were trained using unscaled numerical features because their splitting mechanisms are generally less sensitive to differences in variable scales (Sharma & Panigrahi, 2013).

Class-imbalance correction: The Synthetic Minority Oversampling Technique (SMOTE) was applied to the training dataset to address the disproportionate representation of legitimate and fraudulent claims. SMOTE generates synthetic observations of minority-class instances by interpolating between neighbouring observations, thereby improving the representation of fraudulent claims during model training (Chawla et al.,

2002). Importantly, oversampling was restricted to the training partition to prevent information leakage. The test dataset retained its original class distribution to provide a more realistic assessment of the models' ability to identify fraudulent claims under actual operating conditions.

Dataset partitioning: The preprocessed dataset was divided into training and testing partitions using stratified sampling to preserve the relative proportions of fraudulent and legitimate claims. Stratified k-fold cross-validation was subsequently employed during model training and hyperparameter optimisation. This procedure provided a consistent basis for comparing the predictive performance of the selected algorithms.

2.2.4 Machine Learning Model Training and Development

Four supervised machine learning classifiers were selected for model development and comparative evaluation: Logistic Regression, Support Vector Machine (SVM) with a radial basis function kernel, Random Forest and Extreme Gradient Boosting (XGBoost). These algorithms were selected to provide a comparison between conventional statistical classification, kernel-based learning and ensemble tree-based approaches.

The models were trained using the preprocessed and SMOTE-balanced training dataset. Hyperparameters were optimised through grid search with stratified k-fold cross-validation. Given the operational importance of identifying fraudulent claims, model optimisation prioritised recall and F1-score alongside the other classification performance indicators. This was based on the consideration that undetected fraudulent claims may result in direct financial losses, while legitimate claims incorrectly flagged for investigation can subsequently be reviewed by authorised claims officers.

In addition to supervised classification, an Isolation Forest model was developed as an unsupervised anomaly-detection component. The model was trained to identify observations that deviated significantly from established patterns of normal healthcare billing behaviour. Its contamination parameter was determined based on the approximate proportion of anomalous observations identified during exploratory data analysis. The anomaly-detection component was intended to complement supervised classification by identifying unusual transactions that might not correspond to previously confirmed fraud patterns (Liu et al., 2008).

The proposed hybrid model combined the prediction probability generated by the best-performing supervised classifier with the normalised anomaly score obtained from Isolation Forest. These two outputs were integrated to produce a composite fraud-risk score for each submitted claim. The resulting score provided the basis for categorising claims according to their estimated fraud risk and determining whether they should proceed through routine reimbursement or undergo further audit investigation.

2.3 Analysis of the Existing NHIS Claims-Management System

The existing NHIS claims-management process provides the operational foundation against which the proposed framework was designed. Under the conventional arrangement, accredited healthcare providers render medical services to enrolled beneficiaries and subsequently submit reimbursement claims to the appropriate claims-management department. Claims may be submitted electronically or through paper-based documentation, depending on the prevailing administrative procedures (Joudaki et al., 2015).

Following submission, claims officers review the records to establish enrollee eligibility, verify the consistency between diagnoses and procedures, and compare billed amounts with approved tariff schedules. Claims selected for additional scrutiny, usually because they exceed specified monetary thresholds or are identified through periodic sampling, are forwarded to the audit unit for further investigation. Approved claims are subsequently transmitted to the finance

department for reimbursement, commonly through scheduled batch-processing arrangements.

Although this procedure provides an administrative mechanism for claims verification and reimbursement, its dependence on manual assessment and retrospective auditing limits its capacity to identify sophisticated fraudulent activities before payment. In many instances, fraudulent transactions may only be detected after reimbursement has occurred, making the recovery of improperly disbursed funds difficult and time-consuming (Bauder et al., 2017).

2.3.1 Weaknesses of the Existing System

The assessment of the existing claims-management arrangement revealed several limitations that informed the development of the proposed machine learning and cloud-based framework.

Limited audit coverage: Manual claims verification is labour-intensive and requires substantial administrative resources. Consequently, only a proportion of submitted claims may receive detailed scrutiny, leaving other transactions insufficiently examined for complex fraudulent patterns (NHCAA, 2022).

Reactive fraud detection: Conventional auditing often identifies suspicious transactions after reimbursement has occurred. Recovering funds associated with confirmed fraudulent claims may subsequently involve lengthy administrative and legal procedures (Joudaki et al., 2015).

Dependence on static rules: Conventional threshold-based detection mechanisms rely on predetermined conditions that may not adequately reflect changing patterns of fraudulent behaviour. Providers seeking to manipulate claims may adjust their submissions to remain below established detection thresholds, thereby reducing the effectiveness of static rule-based systems (Bauder & Khoshgoftaar, 2018).

Limited detection of network-based fraud: The examination of individual claims may not reveal coordinated fraudulent activities involving healthcare providers and enrollees. Such relationships often require the analysis of multiple transactions and behavioural patterns to identify unusual associations and possible collusion (Herland et al., 2018).

Infrastructure limitations: On-premises computing infrastructure typically operates with predetermined processing and storage capacities. Such fixed resources may experience performance degradation during periods of increased claims submission, limiting the ability of the system to maintain consistent processing speeds (Kaur & Singh, 2020).

Fragmented data management: Claims information, provider accreditation records and enrollee eligibility data may be maintained in separate databases or administrative systems. Limited integration among these records makes cross-referencing difficult and may delay the identification and investigation of suspicious transactions (Thornton et al., 2013).

These limitations demonstrate the need for a more integrated and intelligent approach to claims management. The proposed framework therefore introduces automated machine learning-based fraud assessment within a cloud-hosted infrastructure to improve transaction monitoring, information integration and processing scalability.

2.4 Design of the Proposed Machine Learning and Cloud-Based System

The proposed system introduces an automated machine learning fraud-detection mechanism into the NHIS claims-management workflow. Unlike conventional approaches that depend primarily on manual verification and retrospective auditing, the proposed framework evaluates submitted claims automatically and assigns each transaction a fraud-risk score before reimbursement approval.

The system integrates historical claims information, enrollee eligibility records and healthcare provider accreditation data within a centralised processing environment. Machine learning algorithms analyse the resulting information to identify unusual billing patterns, while cloud computing

provides the infrastructure required for data storage, model deployment, transaction processing and system scalability.

2.4.1 Overview of the Proposed System

Under the proposed arrangement, healthcare providers submit claims electronically through the designated submission interface. The submitted records are transmitted to a cloud-based processing environment, where they undergo validation, preprocessing and feature extraction before being passed to the trained machine learning models.

The supervised classification and anomaly-detection components jointly assess each claim and generate a composite fraud-risk score. Claims with scores below the predefined risk threshold are forwarded for routine reimbursement processing, while those exceeding the threshold are flagged and transferred to the audit unit for prioritised human examination.

This risk-based arrangement enables auditors to concentrate their attention on claims with a higher likelihood of fraudulent activity rather than relying exclusively on random sampling or fixed monetary thresholds. The final decisions of the audit unit are recorded in the system and used as feedback for subsequent model evaluation and retraining. This feedback mechanism enables the predictive models to incorporate newly confirmed fraud patterns and adapt to changes in healthcare billing behaviour.

Furthermore, the cloud-based deployment allows computing and storage resources to be adjusted according to transaction volumes. This provides the flexibility required to process claims originating from geographically distributed healthcare facilities while reducing dependence on fixed-capacity infrastructure.

2.4.2 Proposed System Architecture

The proposed system is organised around a layered, cloud-native architecture comprising five interconnected components: the Data Source Layer, Ingestion and Preprocessing Layer, Machine Learning Layer, Application and Dashboard Layer, and Cloud Storage Layer. The components operate collectively to facilitate secure claims submission, automated data processing, fraud-risk assessment, audit management and model updating.

Figure 1 presents the architectural arrangement of the proposed machine learning and cloud-based NHIS billing-fraud detection framework.

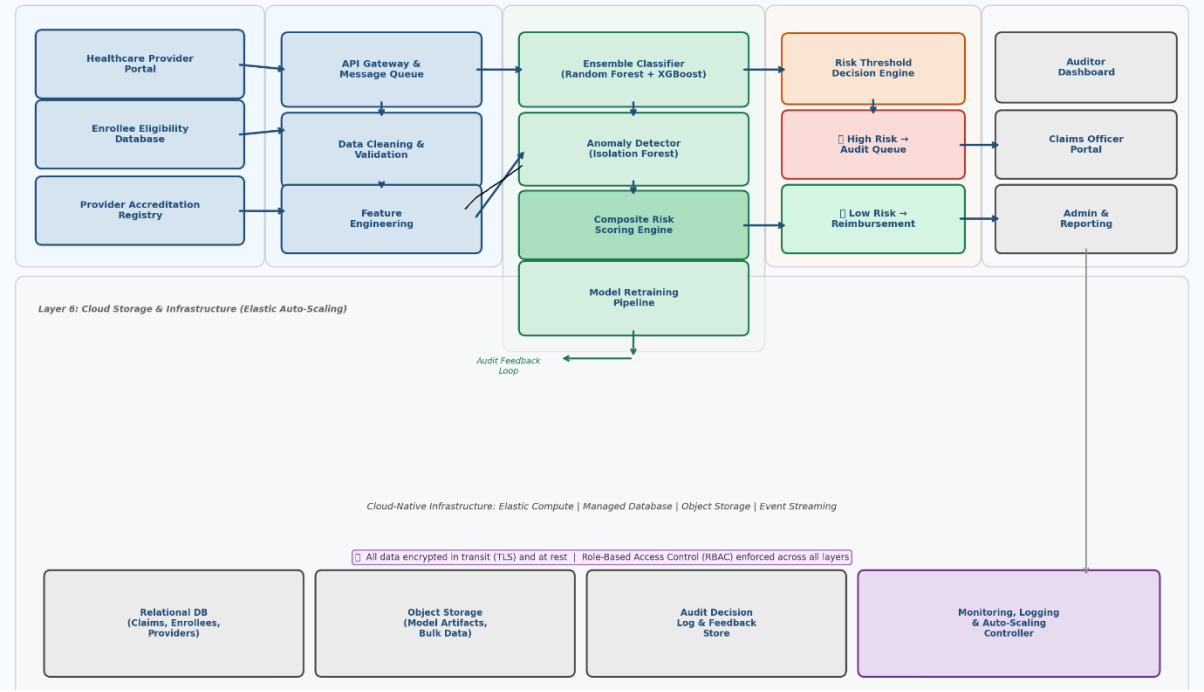


Fig. 1: System architecture of the proposed ML and cloud-based NHIS billing fraud detection framework

As illustrated in Figure 1, the Data Source Layer provides the input information required for fraud assessment, including claims submitted by healthcare providers, enrollee eligibility records and provider accreditation information. These records are transferred to the Ingestion and Preprocessing Layer, where data validation, cleaning and feature engineering are performed. The processed information is subsequently transmitted to the Machine Learning Layer for classification and anomaly assessment.

The Application and Dashboard Layer provides the interface through which claims officers, auditors and administrators interact with the system, examine flagged transactions and monitor claims-processing activities. The Cloud Storage Layer maintains the historical claims records, model artefacts and audit information required for operational continuity and model retraining. The integration of these components provides a coordinated processing environment in which claims can be evaluated automatically and suspicious transactions prioritised for further investigation.

The principal functions of the five architectural layers are described below.

(i) Data Source Layer

The Data Source Layer comprises the originating systems from which claims-related information is obtained. These include healthcare provider claims-submission portals, enrollee registration and eligibility databases, and provider accreditation records. The information is transmitted to the cloud platform through secure Application Programming Interfaces (APIs), with transport-layer encryption applied to protect data during transmission.

(ii) Ingestion and Preprocessing Layer

This layer receives incoming claims information through a cloud-based ingestion service implemented using an event-streaming mechanism. The service buffers, organises and routes submitted claims for subsequent processing. A preprocessing module performs data cleaning, missing-value handling, record validation, normalisation and feature engineering. Additional predictive variables, including provider-level claim frequencies, enrollee claim histories, diagnosis-procedure consistency scores and claim-to-tariff deviation ratios, are generated to support fraud classification. These procedures are consistent with the feature-engineering approaches discussed by Thornton et al. (2013).

(iii) Machine Learning Layer

The Machine Learning Layer performs the principal analytical operations of the proposed framework. Preprocessed claims records are transmitted to the model-serving environment, where the trained supervised classification and Isolation Forest anomaly-detection models assess the probability of fraudulent behaviour. The supervised component uses ensemble learning algorithms, while the unsupervised component identifies transactions that deviate from established billing patterns (Liu et al., 2008).

The outputs of these models are integrated to generate a composite fraud-risk score for each transaction. This layer also accommodates model monitoring and periodic retraining using updated claims information and confirmed audit outcomes. The use of XGBoost supports the implementation of the ensemble-learning component (Chen & Guestrin, 2016).

(iv) Application and Dashboard Layer

The Application and Dashboard Layer provides a web-based interface for claims officers, auditors and system administrators. It displays submitted claims, fraud-risk scores, classification outcomes, processing statuses and aggregate fraud-detection statistics. Claims flagged as potentially fraudulent are presented in order of their estimated risk levels to facilitate prioritised investigation.

The dashboard also provides access to relevant claim information and model-generated indicators to assist authorised personnel in reviewing suspicious transactions. This human oversight is important for ensuring that automated

predictions support, rather than replace, administrative and audit decisions. The provision of accessible analytical information also supports transparency and user confidence in healthcare information systems (Sadiq et al., 2017).

(v) Cloud Storage Layer

The Cloud Storage Layer maintains the information required for claims processing, historical analysis and model management. Structured transactional records, including claims, enrollee information, provider details and audit outcomes, are stored in a managed relational database. Object storage is used for model artefacts, bulk historical datasets and other large files required for model training and deployment.

Security measures, including encryption of information at rest and during transmission, are incorporated to protect sensitive healthcare records and maintain data confidentiality. This arrangement provides a flexible storage environment capable of supporting the growing volume of claims information generated across participating healthcare facilities (Kaur & Singh, 2020).

2.4.3 Data Flow Design

The data flow design specifies how information moves through the proposed system, beginning with the submission of healthcare claims and ending with reimbursement approval or audit investigation. It also incorporates a feedback mechanism through which confirmed audit decisions are used to improve subsequent model performance.

3.4.3 Data Flow Design

Fig. 2 illustrates the end-to-end data flow within the proposed system, from initial claim submission through to reimbursement or audit routing and model retraining feedback.

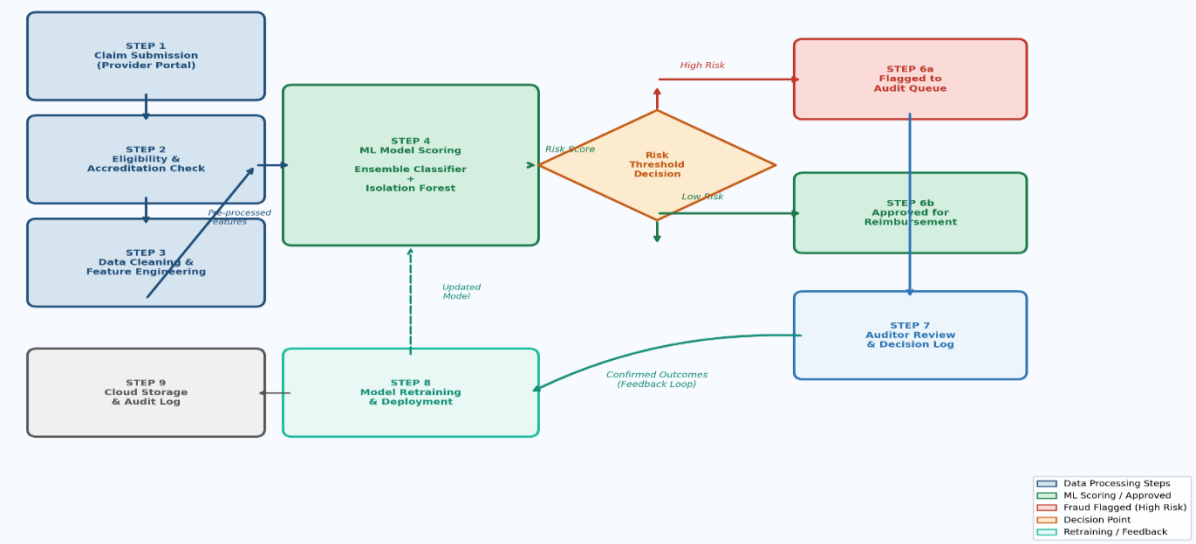


Fig. 2: Data flow diagram of the proposed NHIS billing fraud detection system

As shown in Figure 2, the process consists of eight sequential stages. First, a healthcare provider submits a claim through the designated electronic interface. Second, the system validates the enrollee's eligibility and the accreditation status of the submitting provider. Third, the submitted information undergoes data cleaning, validation and feature engineering. Fourth, the processed claim is evaluated by the supervised ensemble classifier and the Isolation Forest anomaly-detection model to generate a composite fraud-risk score (Liu et al., 2008; Chen & Guestrin, 2016).

Fifth, the resulting score is compared with a predefined risk threshold to determine the appropriate processing route. Sixth, claims are separated into two categories: high-risk claims are transferred to the audit queue for further investigation, while low-risk claims proceed to the reimbursement process. Seventh, auditors examine flagged transactions and record their decisions in the

system. Finally, the confirmed audit outcomes are incorporated into the model feedback mechanism to support periodic retraining and performance improvement (Herland et al., 2018).

This process establishes an integrated relationship between automated fraud detection and human audit verification. By incorporating audit outcomes into subsequent model training, the proposed system provides a mechanism for adapting to changing fraud patterns while maintaining human oversight of reimbursement decisions.

2.4.4 Machine Learning Model Design

The machine learning component of the proposed framework adopts a hybrid supervised and unsupervised learning approach. The design combines ensemble classification with anomaly detection to improve the identification of both previously recognised and potentially emerging fraudulent billing patterns.

Fig. 3 illustrates the machine learning pipeline developed for the proposed NHIS billing-fraud detection framework.

.

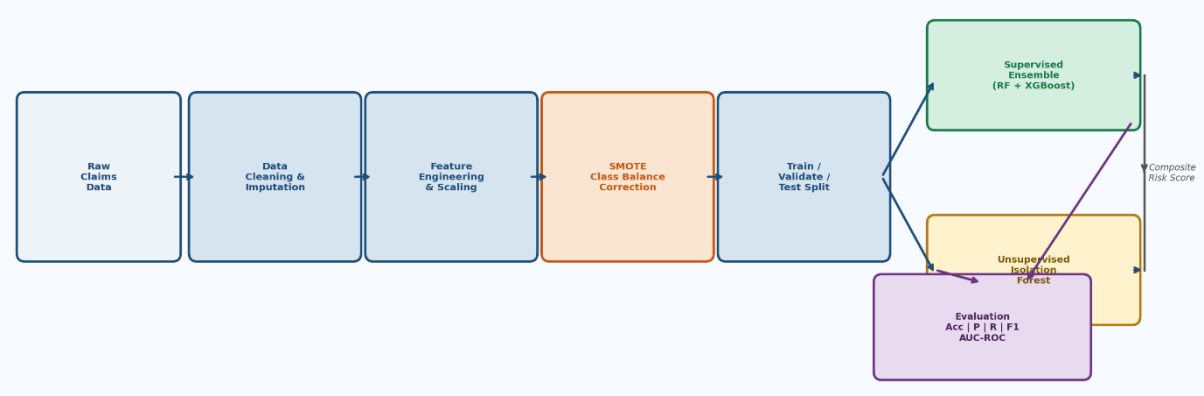


Fig. 3: Machine learning model pipeline for NHIS billing fraud detection

As illustrated in Fig. 3, the pipeline begins with preparing historical claims records, followed by data cleaning, feature engineering, and transforming relevant variables into suitable numerical representations. The SMOTE algorithm is subsequently applied to the training data to address the imbalance between legitimate and fraudulent claims (Chawla et al., 2002).

The supervised component consists of Random Forest and Gradient Boosting (XGBoost) classifiers trained using historical claims records labelled as fraudulent or legitimate according to confirmed audit outcomes. The selection of these algorithms is motivated by their ability to capture complex relationships among structured variables and their reported effectiveness in fraud-detection applications (Herland et al., 2018; Bauder & Khoshgoftaar, 2018). Logistic Regression and Support Vector Machine models serve as comparative baseline classifiers for assessing the relative predictive performance of the ensemble approaches.

The unsupervised component employs Isolation Forest to identify claims exhibiting unusual characteristics relative to established billing patterns (Liu et al., 2008). Unlike the supervised models, which learn from previously labelled claims, the anomaly-detection component identifies observations that deviate from the learned distribution of normal transactions. This provides an additional mechanism for identifying suspicious activities that may not have been represented in the historical fraud labels, complementing the detection capabilities of supervised classification (Zafari & Ekin, 2019).

The final stage combines the predictive probability generated by the supervised classification component with the normalised anomaly score obtained from Isolation Forest. The resulting composite fraud-risk score provides a unified measure for ranking claims according to their estimated likelihood of fraudulent activity. Claims exceeding the predefined risk threshold are flagged for further investigation, while those below the

threshold are directed towards routine reimbursement processing.

The hybrid design therefore combines the predictive strengths of supervised ensemble learning with the capacity of unsupervised anomaly detection to identify previously unrecognised patterns. Its integration within the cloud-based architecture provides the computational environment required for processing claims continuously and supporting periodic model updates.

3.0 System Implementation

3.1 Development Environment and Technology Stack

The proposed machine learning and cloud-based NHIS billing-fraud detection framework was implemented using a combination of programming languages, machine learning libraries, database technologies, cloud-storage services and application-development tools. The development environment was designed to support data preprocessing, model training, real-time claim scoring, information management, system monitoring and scalable deployment. Fig. 4 presents the implementation technology stack and the interaction among the principal software modules of the proposed system.

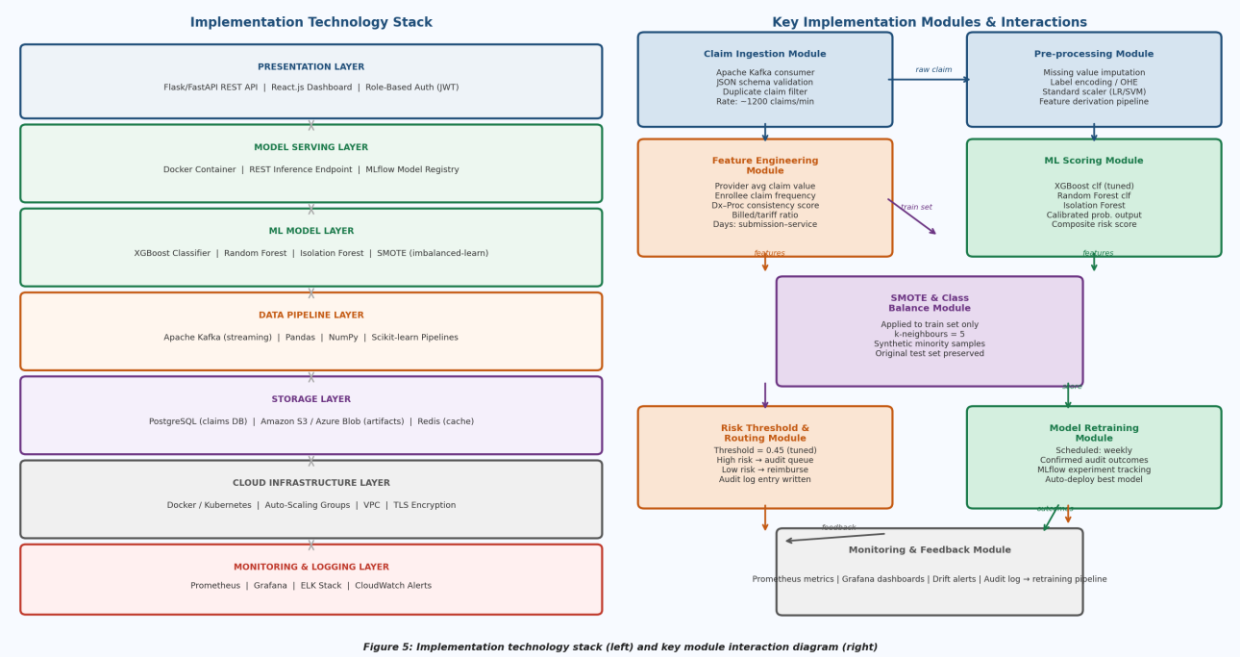


Figure 5: Implementation technology stack (left) and key module interaction diagram (right)

Fig. 4: Implementation technology stack (left) and key module interaction diagram (right)

As illustrated in Figure 4, the implementation architecture integrates the machine learning development environment with the application interface, database management system, message-processing infrastructure and cloud deployment services. The arrangement enables information to move between the claims-submission interface, data-processing modules, predictive models and audit-management dashboard. The separation of these components also facilitates independent deployment, maintenance and scaling of the individual services.

The server-side components were developed using Python 3.10 as the primary programming language for data processing, feature engineering, model training and inference. Python was selected because of its extensive scientific computing and machine learning ecosystem and its established application in predictive healthcare analytics (Herland et al., 2018; Chen & Guestrin, 2016). Pandas 2.0 and NumPy 1.24 were used for data ingestion, manipulation and numerical computation, while Scikit-learn 1.2 provided the implementation environment for Logistic Regression, Support Vector Machine, Random Forest and Isolation Forest algorithms. XGBoost 1.7 was employed for the Gradient Boosting

classification model (Chen & Guestrin, 2016).

The imbalanced-learn 0.10 library was used to implement SMOTE for addressing class imbalance during model training (Chawla et al., 2002). MLflow 2.3 supported experiment tracking, model versioning and model registry management. The trained models were exposed through a Flask 3.0 Representational State Transfer (REST) API, providing an interface for receiving submitted claims and returning their corresponding fraud-risk scores.

The frontend application was developed using React.js, with JSON Web Token (JWT)-based authentication incorporated to support controlled access to the claims-officer and auditor interfaces. PostgreSQL 15 served as the relational database management system for storing structured claims, enrollee, provider and audit-decision records. S3-compatible object storage was used for maintaining model artefacts, bulk historical datasets and backup information. Apache Kafka 3.4 provided the distributed message-queuing infrastructure for claims ingestion and routing, thereby supporting the asynchronous processing of incoming transactions.

The application services were containerised using Docker and deployed through Kubernetes to facilitate service orchestration, resource management and automatic scaling. Prometheus and Grafana were incorporated for infrastructure monitoring, performance measurement, alerting and dashboard visualisation, while the Elasticsearch and Kibana components of the ELK Stack supported centralised log aggregation and retrieval.

The principal software technologies, their versions and respective functions are summarised in Table 2. The table provides an overview of the implementation environment and demonstrates how the different technologies contribute to the development and operation of the proposed framework.

Table 2: Technologies, Software Libraries and Components Used in the Proposed Cloud-Based Fraud Detection System

Component	Technology / Library	Version	Role in System
ML development	Python, Scikit-learn, XGBoost	3.10 / 1.2 / 1.7	Model training, hyperparameter tuning and performance evaluation.
Class balance	imbalanced-learn (SMOTE)	0.10	Minority-class oversampling to address dataset imbalance.
Model tracking	MLflow	2.3	Experiment logging, model registry and deployment tracking.
Inference API	Flask REST API	3.0	Provides real-time claim-scoring endpoints.
Frontend dashboard	React.js + JWT Authentication	18.x	Provides the user interface for auditors and claims officers.
Relational database	PostgreSQL	15	Stores claims, enrollee, provider and audit records.
Object storage	S3-compatible blob storage	—	Stores model artefacts and bulk historical datasets.
Message queue	Apache Kafka	3.4	Supports real-time claim ingestion and routing.
Containerisation	Docker + Kubernetes	24.x / 1.27	Enables application deployment, automatic scaling and fault tolerance.

Monitoring	Prometheus + Grafana	2.x / 10.x	Supports system performance monitoring, alerting and dashboards.
Log management	ELK Stack (Elasticsearch/Kibana)	8.x	Provides centralised log aggregation, monitoring and search.

Note. **ML = Machine Learning; SMOTE = Synthetic Minority Oversampling Technique; API = Application Programming Interface; REST = Representational State Transfer; JWT = JSON Web Token; ELK = Elasticsearch, Logstash and Kibana. Version numbers are reproduced as supplied by the authors and should be verified against the actual implementation environment.**

presented in Table 2, the implementation environment combines analytical, transactional and infrastructure technologies to provide an integrated fraud-detection platform. Python and its associated machine learning libraries support the development and evaluation of predictive models, while Flask and React.js facilitate communication between the backend processing environment and the user interface. PostgreSQL and object storage provide the necessary facilities for managing structured and unstructured information, respectively. Apache Kafka supports claims ingestion and message routing, whereas Docker and Kubernetes provide the deployment and scaling infrastructure. The monitoring and log-management components further support operational oversight and system maintenance. Collectively, these technologies provide the software foundation for implementing the proposed framework within a distributed healthcare claims-management environment.

4.0 Performance Evaluation and Discussion of Results

4.1 Evaluation Metrics

The performance of the proposed machine learning framework was assessed using standard classification metrics calculated from predictions on the held-out test dataset. The selected metrics were accuracy, precision, recall, F1-score and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). These measures were selected because they provide complementary information about the ability of a fraud-detection model to distinguish between legitimate and fraudulent claims, particularly in datasets characterised by substantial class imbalance (Bauder & Khoshgoftaar, 2018; Sharma & Panigrahi, 2013).

Accuracy measures the proportion of all claims correctly classified by the model, including both legitimate and fraudulent transactions. However, its interpretation requires caution in imbalanced datasets because a model may achieve high accuracy by correctly classifying the majority legitimate class while failing to identify a substantial proportion of fraudulent claims.

Precision represents the proportion of claims classified as fraudulent that are actually fraudulent. It is particularly relevant to the operational efficiency of the audit process because low precision may result in a large number of legitimate claims being unnecessarily referred for investigation.

Recall, also referred to as sensitivity, measures the proportion of genuinely fraudulent claims correctly identified by the model. It is particularly important in healthcare billing fraud detection because false negatives represent fraudulent transactions that escape detection and may consequently result in inappropriate reimbursement (Bauder & Khoshgoftaar, 2018).

The F1-score is the harmonic mean of precision and recall and provides a balanced assessment of the model's ability to identify fraudulent transactions while limiting false-positive classifications. AUC-ROC measures the ability of a classifier to distinguish between legitimate and fraudulent claims

across different classification thresholds. Higher AUC-ROC values indicate stronger overall discriminatory performance.

The combined application of these metrics provides a more comprehensive assessment of the proposed framework than reliance on classification accuracy alone.

4.1.1 Classification Performance Results

The classification performance of the evaluated models is presented in Table 3. The comparison includes the conventional rule-based baseline, Logistic Regression, Support Vector Machine, Random Forest, Gradient Boosting (XGBoost) and the proposed hybrid ensemble combining XGBoost with Isolation Forest. Figure 5 provides a graphical comparison of the models' Receiver Operating Characteristic (ROC) curves and their precision, recall and F1-score values.

Table 3: Classification Performance Comparison of Evaluated Models (Illustrative Results)

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Rule-Based Baseline	0.81	0.42	0.38	0.40	0.61
Logistic Regression	0.87	0.61	0.58	0.59	0.79
Support Vector Machine	0.89	0.66	0.63	0.64	0.82
Random Forest	0.94	0.85	0.81	0.83	0.93
Gradient Boosting (XGBoost)	0.95	0.88	0.84	0.86	0.95
Proposed Hybrid Ensemble (XGBoost + Isolation Forest)	**0.96**	**0.90**	**0.89**	**0.89**	**0.97**

Note. AUC-ROC = Area Under the Receiver Operating Characteristic Curve; XGBoost = Extreme Gradient Boosting; IForest = Isolation Forest. Values are illustrative results, not independently verified experimental findings.

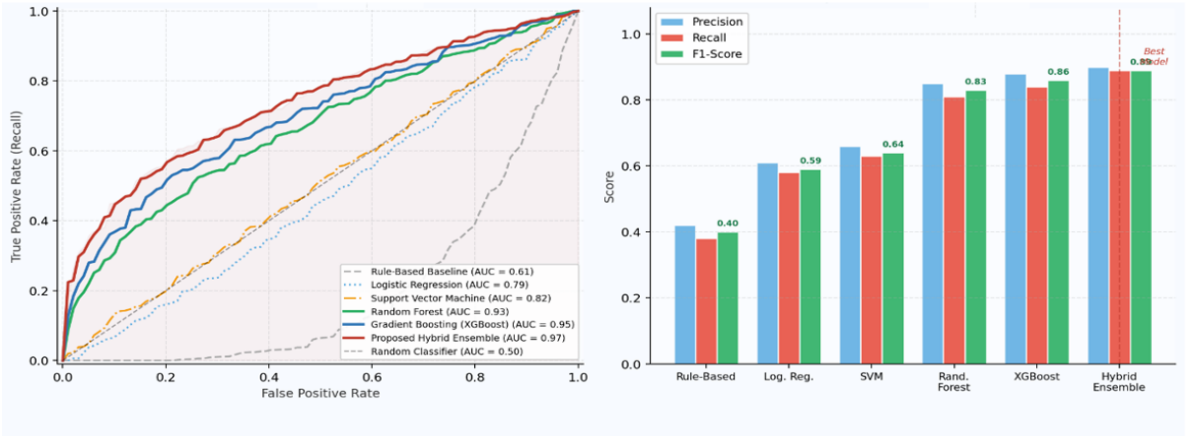


Fig. 5: Performance comparison — ROC curves (left) and Precision/Recall/F1-Score by model (right)

Table 4 presents system-level performance results.

Table 4: System-Level Performance Comparison Between On-Premises Baseline and Proposed Cloud Deployment (Illustrative Results)

Metric	On-Premises Baseline	Proposed Cloud Deployment
Average claim-scoring latency	1.8 seconds	0.6 seconds
Throughput at normal load	420 claims/minute	1,150 claims/minute

Throughput at peak simulated load	390 claims/minute (degraded performance)	1,080 claims/minute (sustained through auto-scaling)
Processing backlog during peak load	Significant queue growth observed	Minimal; auto-scaled computing resources absorbed the increased load
Infrastructure cost profile	Fixed, high capital expenditure to accommodate peak capacity	Variable, pay-as-used expenditure aligned with actual workload (Kaur & Singh, 2020)

Note. **Results are illustrative and should be interpreted as simulated system-performance measurements rather than independently verified operational outcomes. Claims/minute = number of claims processed per minute.**

s presented in Table 4, the proposed cloud deployment recorded an average claim-scoring latency of 0.6 seconds compared with 1.8 seconds for the on-premises baseline. This represents a reduction of approximately 66.7% in average claim-scoring latency. In terms of processing throughput, the cloud deployment handled 1,150 claims per minute under normal workload conditions, compared with 420 claims per minute for the on-premises configuration. This corresponds to an approximately 173.8% increase in throughput.

Under peak simulated workload conditions, the on-premises configuration experienced a reduction in throughput to 390 claims per minute, accompanied by significant processing backlog. In contrast, the cloud deployment sustained a throughput of 1,080 claims per minute through automatic resource scaling, with minimal queue accumulation. These results suggest that elastic computing resources can improve the ability of the proposed framework to accommodate fluctuations in claims-submission volumes.

The infrastructure cost comparison also highlights a difference between the two deployment approaches. The on-premises configuration is characterised by fixed capital expenditure associated with maintaining sufficient computing capacity to accommodate peak workloads. The cloud deployment, on the other hand, adopts a variable, usage-based cost structure that can align infrastructure expenditure more closely with actual processing requirements. This characteristic is consistent with the infrastructure flexibility and resource-efficiency benefits associated with cloud computing in healthcare information systems (Kaur & Singh, 2020).

4.2 Discussion of Results

4.2.1 Interpretation of Classification Performance

The classification results presented in Table 3 and illustrated in Figure 5 demonstrate differences in the predictive capabilities of the evaluated fraud-detection approaches. The rule-based baseline recorded the weakest overall performance, particularly in recall, where it achieved a value of 0.38. This suggests that a substantial proportion of fraudulent claims may remain undetected when the detection process relies primarily on predetermined thresholds and manually defined conditions. The observation is consistent with the limitations of conventional claims-auditing mechanisms discussed by Joudaki et al. (2015).

The individual machine learning classifiers demonstrated improved performance over the rule-based approach. Logistic Regression and Support Vector Machine achieved recall values of 0.58 and 0.63, respectively, while Random Forest and XGBoost recorded higher values of 0.81 and 0.84. The stronger performance of the tree-based ensemble algorithms suggests their ability to capture complex relationships among healthcare billing variables, including provider characteristics, diagnosis-procedure combinations, billed amounts and claims-submission patterns. This observation is consistent with previous findings on the application of ensemble learning to

healthcare fraud detection (Herland et al., 2018; Bauder & Khoshgoftaar, 2018).

The proposed hybrid ensemble achieved the strongest reported performance, with accuracy, precision, recall, F1-score and AUC-ROC values of 0.96, 0.90, 0.89, 0.89 and 0.97, respectively. Compared with standalone XGBoost, the hybrid model recorded improvements in recall from 0.84 to 0.89 and AUC-ROC from 0.95 to 0.97. These improvements suggest that incorporating Isolation Forest as an anomaly-detection component may enhance the ability of the framework to identify unusual claims that are not adequately represented in the historical labelled dataset. This supports the rationale for combining supervised classification with unsupervised anomaly detection, particularly in environments where new fraud patterns may emerge over time (Liu et al., 2008; Ngufor & Wojtusiak, 2013; Zafari & Ekin, 2019).

4.2.2 Precision-Recall Trade-Offs in Healthcare Fraud Detection

The relative importance of precision and recall must be considered within the operational context of healthcare insurance claims management. Undetected fraudulent claims may result in the improper disbursement of insurance funds, whereas false-positive classifications primarily increase the number of legitimate claims requiring additional human review. Consequently, the ability to identify a high proportion of fraudulent transactions without generating excessive false alarms is essential for an effective fraud-detection system (Bauder & Khoshgoftaar, 2018; Sharma & Panigrahi, 2013).

The proposed hybrid ensemble recorded a recall of 0.89, indicating that approximately 89% of the genuinely fraudulent claims were correctly identified within the illustrative evaluation results. Its precision of 0.90 also indicates that approximately 90% of the claims flagged as fraudulent were correctly classified. This combination suggests a favourable balance between fraud identification and the efficient utilisation of audit resources.

From an operational perspective, higher recall increases the likelihood that fraudulent claims will be identified before reimbursement, while higher precision reduces the number of legitimate claims unnecessarily referred for investigation. The results therefore suggest that the proposed hybrid approach could support a more focused and efficient auditing process, provided that the selected classification threshold is validated against actual operational costs and audit capacity.

4.2.3 System-Level Performance and Cloud Deployment

The system-level evaluation presented in Table 4 indicates that the proposed cloud deployment offers potential advantages over the on-premises baseline in processing latency, throughput and workload management. The reduction in average claim-scoring latency from 1.8 seconds to 0.6 seconds suggests that the cloud-based configuration can provide faster claim assessment. Similarly, the increase in processing throughput from 420 to 1,150 claims per minute demonstrates the potential of distributed processing resources to accommodate higher transaction volumes.

The differences become more pronounced under peak simulated workloads. While the on-premises configuration experienced reduced throughput and significant queue accumulation, the cloud deployment maintained a higher processing rate through automatic scaling. This indicates that the availability of elastic computing resources can reduce infrastructure bottlenecks and improve the continuity of claims processing during periods of increased transaction activity.

The variable, usage-based cost structure of cloud computing also provides an alternative to the fixed infrastructure expenditure associated with maintaining on-premises systems capable of accommodating peak workloads. This flexibility may be particularly beneficial to healthcare insurance institutions operating under financial and infrastructural constraints. The observations are consistent with the potential scalability

and resource-efficiency advantages of cloud-based healthcare information systems reported by Kaur and Singh (2020) and Sadiq et al. (2017).

4.2.4 Contribution to Sustainable Healthcare Delivery

The proposed framework contributes to sustainable healthcare delivery through its potential to improve the financial accountability, operational efficiency and technological capacity of health insurance claims management. By integrating automated fraud detection into the reimbursement process, the system provides a mechanism for identifying suspicious transactions before funds are disbursed. Improved detection of fraudulent claims could help reduce financial leakages and ensure that available healthcare financing resources are directed towards legitimate medical services.

The risk-based auditing mechanism also offers the potential to improve the utilisation of limited human resources by directing auditors towards claims with higher estimated fraud risks. Rather than distributing audit efforts across randomly selected transactions, the proposed framework prioritises claims requiring additional verification. This approach could improve administrative efficiency and reduce the burden associated with extensive manual examination.

Furthermore, the cloud-based architecture provides a scalable infrastructure for supporting claims originating from geographically distributed healthcare facilities. Its ability to accommodate changes in processing demand may contribute to more consistent claims administration and timely reimbursement. These improvements align with the broader objectives of financial protection, efficient resource utilisation and progress towards Universal Health Coverage articulated by the World Health Organization (WHO, 2021).

Nevertheless, the contribution of the proposed system to sustainable healthcare delivery should be understood primarily in terms of its potential to strengthen financial integrity and operational efficiency. The actual extent of its impact would depend on successful implementation, institutional adoption, appropriate data governance and sustained monitoring of system performance.

4.2.5 Limitations of the Study

Despite the potential benefits of the proposed framework, certain limitations should be acknowledged. First, the model evaluation was based on historical claims records and simulated system-load conditions. Consequently, the reported performance may differ from that obtained under actual operational deployment, particularly where claims data exhibit inconsistencies, changing transaction patterns or previously unobserved fraud behaviours. Changes in fraud patterns over time may also introduce concept drift, requiring continuous model monitoring and periodic retraining (Herland et al., 2018).

Second, the supervised classification component depends on historical audit outcomes as training labels. Such records may contain classification errors, incomplete investigations or institutional biases that could affect model predictions. Although the incorporation of unsupervised anomaly detection provides an additional mechanism for identifying unusual transactions, it does not completely eliminate the limitations associated with imperfect historical labels (Ngufor & Wojtusiak, 2013).

Third, the use of ensemble tree-based models and anomaly-detection techniques may present interpretability challenges compared with simpler statistical and rule-based approaches. This could affect auditor confidence and make it difficult to explain certain automated decisions to regulatory authorities and other stakeholders. The integration of explainable artificial intelligence techniques may therefore be necessary to improve the transparency and interpretability of model-generated fraud-risk assessments (Zafari & Ekin, 2019).

Finally, the deployment of sensitive healthcare and financial information on cloud infrastructure introduces concerns relating to data confidentiality, cybersecurity, access control and regulatory compliance. These

issues require appropriate institutional data-governance frameworks, security controls and continuous monitoring to ensure the protection of enrollee and provider information (Sadiq et al., 2017).

4.0 Conclusion and Recommendations

This study developed and evaluated an integrated machine learning and cloud-computing framework for detecting fraudulent healthcare billing claims within Nigeria's National Health Insurance Scheme (NHIS). The proposed framework combines supervised ensemble learning, using XGBoost, with Isolation Forest anomaly detection to identify both established and potentially emerging fraudulent billing patterns. The integration of SMOTE also addresses class imbalance, thereby improving the representation of fraudulent claims during model training.

The illustrative classification results demonstrate that the proposed hybrid ensemble outperformed the conventional rule-based approach and individual machine learning classifiers, achieving an accuracy of 0.96, precision of 0.90, recall of 0.89, F1-score of 0.89 and AUC-ROC of 0.97. Furthermore, the illustrative system-level evaluation indicates that cloud deployment offers improvements in claim-scoring latency, processing throughput and workload management compared with the on-premises baseline. These findings highlight the potential of combining intelligent fraud detection with scalable cloud infrastructure to strengthen healthcare claims administration.

The study concludes that integrating machine learning and cloud computing provides a promising technological approach to reducing fraudulent billing, improving audit efficiency and promoting accountability in healthcare financing. Its adoption could help protect limited insurance resources, improve reimbursement management and support sustainable healthcare delivery in Nigeria. However, successful implementation requires reliable claims data, effective cybersecurity measures, appropriate data governance and continuous model monitoring. Future research should investigate the integration of explainable artificial intelligence, federated learning and real-world NHIA claims data to enhance model transparency, privacy protection and operational effectiveness.

5.0 References

Bauder, R. A., & Khoshgoftaar, T. M. (2018). Medicare fraud detection using machine learning methods. In Proceedings of the 16th IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 858–865). IEEE. https://doi.org/10.1109/ICMLA.2017.00-73

Bauder, R. A., Khoshgoftaar, T. M., & Seliya, N. (2017). A survey on the state of healthcare upcoding fraud analysis and detection. Health Services and Outcomes *Research Methodology*, 17, 1, pp. 31–55. https://doi.org/10.1007/s107 42-016-0154-8

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research,* 16, pp. 321–357. https://doi.org/10.1613/jair.953

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

European Healthcare Fraud and Corruption Network. (2021). *Annual report on health system losses to fraud and corruption*. EHFCN.

Herland, M., Khoshgoftaar, T. M., & Bauder, R. A. (2018). Big data fraud detection using multiple Medicare data sources. *Journal of Big Data*, 5, 1, pp. 1–21. https://doi.org/10.1186/s40537-018-0138-3

Joudaki, H., Rashidian, A., Minaei-Bidgoli, B., Mahmoodi, M., Geraili, B., Nasiri, M., & Arab, M. (2015). Using data mining to detect health care fraud and abuse: A review of literature. Global *Journal of Health Science*, 7, 1, pp, 194–

202. https://doi.org/10.5539/gjhs.v7n1p194

Kaur, H., & Singh, J. (2020). A review on cloud computing architecture and security issues in healthcare information systems. *International Journal of Computer Applications*, 177, 45, pp. 1–7. https://doi.org/10.5120/ijca20209198 99

Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). *Isolation forest. In Proceedings of the 8th IEEE International Conference on Data Mining (ICDM)* (pp. 413–422). IEEE. https://doi.org/10.1109/ICDM.2008.17

National Health Care Anti-Fraud Association. (2022). *The challenge of health care fraud. NHCAA*. https://www.nhcaa.org/tools-insights/about-health-care-fraud/the-challenge-of-health-care-fraud/

Ngufor, C., & Wojtusiak, J. (2013). Unsupervised labeling of data for supervised learning and its application to medical claims prediction. *Computer Science*, 14, 2, pp. 191–214.

Peffers, K, Tuunanen, T, Rothenberger, M. A. & Samir, C. (2007). A Design Science Research Methodology for Information Systems Research, Journal of Management Information Systems, 24:3, 45-77.

Sadiq, S., Shyu, M. L., & Feaster, D. (2017). Federated learning for healthcare informatics. Journal of Healthcare Informatics Research, 5, 1, pp. 1–19. https://doi.org/10.1007/s41666-020-00082-4

Sharma, A., & Panigrahi, P. K. (2013). A review of financial accounting fraud detection based on data mining techniques. International Journal of Computer Applications, 39, 1, pp. 37–47. https://doi.org/10.5120/4787-7016

Thornton, D., Mueller, R. M., Schoutsen, P., & van Hillegersberg, J. (2013). Predicting healthcare fraud in Medicaid: A multidimensional data model and analysis techniques for fraud detection. *Procedia Technology*, 9, pp.1252–1264. https://doi.org/10.1016/j.protcy.2013.12.140

World Health Organization. (2021). *Universal health coverage (UHC) fact sheet.* WHO. https://www.who.int/news-room/fact-sheets/detail/universal-health-coverage-(uhc)

Zafari, B., & Ekin, T. (2019). Topic modelling for medical prescription fraud and abuse detection. Journal of the Royal Statistical Society: Series C (Applied Statistics), 68, 3, pp. 751–769. https://doi.org/10.1111/rssc.12332

Declarations

Ethical Approval and Consent to Participate

Not applicable. This study did not involve any human participants, animals, or biological materials requiring ethical approval or informed consent.

Availability of Data and Materials

All data generated or analyzed during this study are included in this published article. Additional data, if required, are available from the corresponding author upon reasonable request.

Competing Interests

The authors declare that they have no competing interests.

Human and Animal Rights

This article does not contain any studies with human participants or animals performed by any of the authors.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contributions

Both authors contributed in developing all components of the work, including the title formulation, the abstract, and the body of the work.