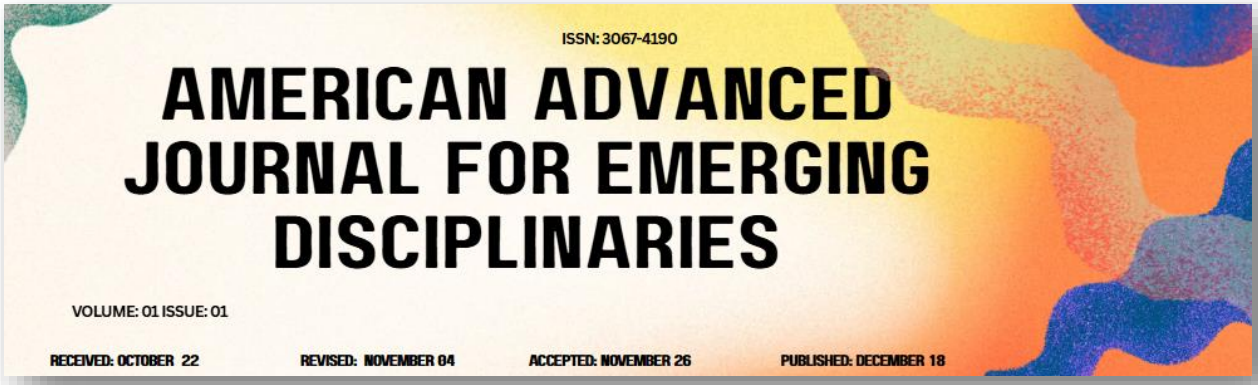


Explainable GenAI Fraud Defense in Hybrid Cloud Insurance Systems

Mallesham Goli
Independent Researcher
mallesham.goli.research01@gmail.com

Abstract

Using hybrid cloud-native architecture, machine learning (ML) and AIOps technologies, this study addresses the explainability challenges of insurance fraud identification models built on proprietary telecommunications data. Internal and external enterprise data, together with historical fraud incident records, power supervised classification models for fraud detection. Explainable Artificial Intelligence (XAI) technologies, which shed light on model decision-making processes and outcomes, are essential for business acceptance, regulatory compliance, and alignment with risk-mitigation objectives. The research thus fills a critical gap in the literature, encouraging organizations to move beyond accuracy-centric goals and embrace a more comprehensive view.

The findings reveal that combining data from telecommunication systems with historical fraud event data leads to more accurate models and highlights the importance of class-imbalance calibration. A realistic simulation involving a high-tech anti-fraud operation showcases the potential of a hybrid cloud-native architecture with MLOps capabilities for supporting the whole ML life cycle. With appropriate scrutiny and evaluation, these models can serve business decision-making and be deployed within AIOps procedures built on hybrid cloud architectures. Looking ahead, the investigation points to fruitful future work in two areas: enhancing the explainable capabilities of fraud models through generative AI and establishing new methodological standards for the deployment of explainable fraud detection models within hybrid cloud environments.

Keywords: Hybrid Cloud Native Architecture, Insurance Fraud Detection, Explainable Artificial Intelligence, XAI For Fraud Models, Machine Learning Life Cycle Management, MLOps And AIOps Integration, Telecommunications Data Analytics, Proprietary Enterprise Data, Supervised Classification Models, Fraud Risk Scoring, Model Explainability And Transparency, Regulatory Compliance In AI, Class Imbalance Calibration, Business Decision Support Systems, Hybrid Cloud Deployment Models, Operational Fraud Monitoring, AI Governance And Trust, Generative AI For Explainability, Enterprise Scale ML Systems, Risk Mitigation Frameworks.

1. Introduction

Define problem, motivations, research questions, and contribution to theory and practice; justify hybrid cloud-native and generative AI angles with current gaps and implications for explainability.

Insurance fraud represents a major drain on the finances of global organisations, with increasing rates every year. Machine learning models are often deemed the best solution for fraud detection. However, their adoption is hindered by lacking explainability. Give non-experts a way to understand model output and regulatory requirements, pressure for explanations grows. Generative AI approaches are fundamental in a non-BRAI Predictive Analytics framework to deliver counterfactuals and System 1 alignment. Nonetheless, ML provides System 2 explanations that are also pivotal for regulatory requirements.

Explainability is desired yet rarely applied to Machine Learning models for fraud detection. A review of works using tree-based models demonstrates a major focus on precision than recall, neglecting risk assessment and making implementation difficult. Simple explanations in the form of feature importances are also inaccurate for complex multi-class problems, requiring greater

ISSN: 3067-4190

AMERICAN ADVANCED JOURNAL FOR EMERGING DISCIPLINARIES

VOLUME: 01 ISSUE: 01

RECEIVED: OCTOBER 22 REVISED: NOVEMBER 04 ACCEPTED: NOVEMBER 26 PUBLISHED: DECEMBER 18

understanding under adversarial threat detection. Addressing different facets of explainability is vital, enabling deployment-ready exploration and increasing the models' acceptance by consumers and regulators.

1.1. Overview of the Study and Its Significance

Insurance fraud imposes considerable costs on society at large, the insurance sector, and policyholders. The principal focus of this study is to advance the state of explainable AI (XAI) for insurance fraud detection. Given that the topic tends to be a matter of concern for financial crime investigation authorities and regulators, achieving explainability at the model level is indeed crucial. The investigation is conducted from a hybrid cloud-native perspective, taking advantage of the specific features that this architecture offers for the development and deployment of automated machine learning models.

Within insurance fraud detection, XAI has a somewhat different role from other domains. Here, the presence or absence of fraud in any insurance claim is already a well-defined fact; the explainability problem centers on validating predictions made by models used in the identification of true positives rather than being general agents of support for end users. Nevertheless, certain explainability elements typically considered in the literature need to be taken into account, both for regulatory purposes and to support investigation activities. The fundamental research question addressed is, therefore, how, to what extent, and with what implications for practical deployment insurance fraud-discriminating models trained and deployed in a hybrid cloud-native environment can be made explainable.

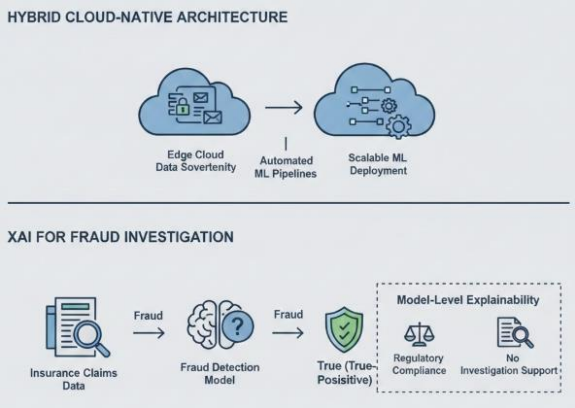


Fig 1: Transparency in the Cloud: A Hybrid Cloud-Native Framework for Explainable AI (XAI) in Insurance Fraud Detection and Regulatory Compliance

2. Theoretical Foundations

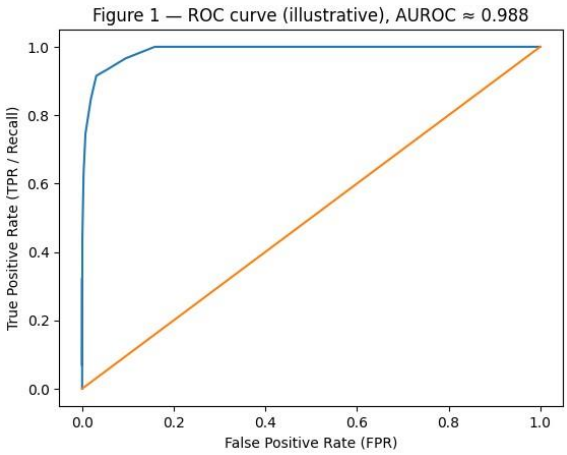
Explainable AI (XAI) is an established subfield whose implications are profoundly felt in a variety of application domains, particularly in heavily regulated contexts such as banking and insurance, where fraud detection represents a major area of investment in XAI. For these models to be useful for stakeholders beyond the end-users of a model, particularly fraud risk managers and regulators, the importance of providing trustworthy and interpretable explanations of model outputs relies on adherence to a set of principles for explainability. These principles address users' needs as well as regulatory requirements and the very nature of fraud situations arising in the domain. Such requirements include trustworthy information, transparency of processes, domain-specific needs (e.g., labour-related observations for fraudulent claims), provision of explanations that satisfy external oversight and accountability, regulatory supervision for consumer confidence, and assurance that explanations support effective response strategies. While explainability is necessary for the effective deployment of most applications in the insurance domain, the deployment of explainable fraud models represents a special arrange operation.

ISSN: 3067-4190
AMERICAN ADVANCED JOURNAL FOR EMERGING DISCIPLINARIES
VOLUME: 01 ISSUE: 01
RECEIVED: OCTOBER 22 REVISED: NOVEMBER 04 ACCEPTED: NOVEMBER 26 PUBLISHED: DECEMBER 18

Trust, and hence user adoption and confidence, is paramount for any decision-making system. User trust in a system is normally derived from a good understanding of the inner workings behind the model (its level of transparency), its ability to provide understandable explanations, and its ability to provide information that can be used for supervisory purposes. As trust can also be adversely affected by bad performance, validation of the models before deployment is vital. In the fraud context, while having an accurate model is important, it may not be sufficient: "an accurate model will not necessarily decrease the number of false negatives ... in fact it may even increase them. An accurate model cannot substitute an explainable one".

	True = Fraud (1)	True = Legit (0)
Pred = Fraud (1)	TP	FP
Pred = Legit (0)	FN	TN

Figure 1 — ROC curve (illustrative), AUROC ≈ 0.988



Equation 1: Core binary classification setup (fraud vs non-fraud)

Let each claim i have:

- True label $y_i \in \{0,1\}$ where 1=fraud, 0=legit
- Model score $\hat{p}_i \in [0,1]$ (predicted probability of fraud)
- Threshold τ so predicted class is:

$$\hat{y}_i = \begin{cases} 1, & \hat{p}_i \geq \tau \\ 0, & \hat{p}_i < \tau \end{cases}$$

This matches the paper's "$P[\hat{y}_n] \geq 0.5$" style boundary.

2.1. Explainable AI in Insurance

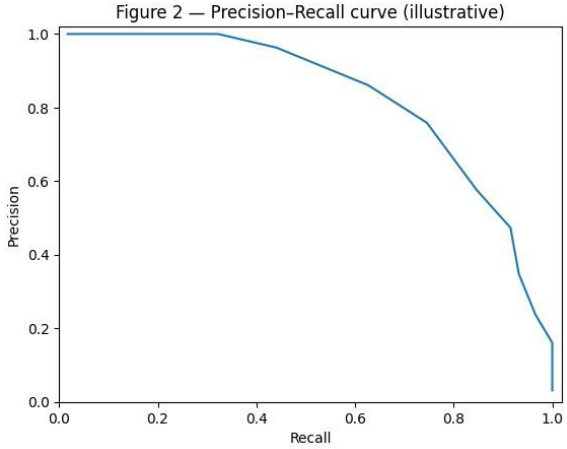
Explainability is a crucial factor for the responsible adoption of machine-learning (ML) algorithms in sensitive business environments, including insurance. Although various explainable-AI approaches have emerged, any specific requirements for explainable insurance fraud models have been largely overlooked. The present work provides an analysis of why insurance companies need explainable models for fraud detection, supporting arguments with aspects of transparency, trust, regulatory obligations, and domain characteristics. Subsequently, practical protocols for developing models with an explainability-first approach are discussed, incorporating evaluation metrics, cross-validation design, stress-testing procedures, and guidelines for calibrating explanations. Readers will also find that the proposed considerations are crucial for commercial deployment.

Insurance is a highly regulated sector in which decisions must be justified not only to customers but also to third parties such as ombudsmen and courts. Why do customers expect model decisions to be interpretable? When ML fraud models recommend rejecting or alleging a claim, customers are more likely to perceive injustice and bias than are claim approvals. Knowing the reasons behind these judgments limits perceptions of injustice and enhances customers' trust in surrounding technologies. Failure to do so would increase the AFDA (anti-fraud denial audience), exposing the insurer to higher reputational and financial risks. Finally, regulations such as the EU's "digital services act" impose strain on algorithmic accountability; indeed, companies whose operations heavily involve ML technologies may be obliged to present

ISSN: 3067-4190
AMERICAN ADVANCED JOURNAL FOR EMERGING DISCIPLINARIES
VOLUME: 01 ISSUE: 01
RECEIVED: OCTOBER 22 REVISED: NOVEMBER 04 ACCEPTED: NOVEMBER 26 PUBLISHED: DECEMBER 18

understandable explanations for their decisions in their digital services.

Figure 2 — Precision–Recall curve (illustrative)



Equation 2: Confusion matrix (TP/FP/TN/FN) — derived

Define counts:

- **TP** (true positives): $TP = \sum_i \mathbf{1}\,[\hat{y}_i = 1 \wedge y_i = 1]$
- **FP** (false positives): $FP = \sum_i \mathbf{1}\,[\hat{y}_i = 1 \wedge y_i = 0]$
- **TN** (true negatives): $TN = \sum_i \mathbf{1}\,[\hat{y}_i = 0 \wedge y_i = 0]$
- **FN** (false negatives): $FN = \sum_i \mathbf{1}\,[\hat{y}_i = 0 \wedge y_i = 1]$

3. Data Engineering and Governance

Three main categories of data sources impact application performance: data from within the insurance company, data from external commercial providers, and public domain data. Company data, typically stored in structured database systems, is integrated into a Cloud Data Lake using ingestion pipelines that perform Extract-Transform-Load (ETL) operations. The data schema is harmonized during data preparation to ensure compatibility between internal and external datasets. Data governance processes ensure data quality, lineage, and provenance, privacy and compliance controls, and security.

Data Integration in Cloud-Native Environments SCHEMA MAPPING: EXTERNAL SOCIAL AND ECONOMIC DATA. Data integration in a cloud-native environment encompasses the ingestion of public domain datasets from various national institutions and agencies. The aim is to provide a contextualized view of insured entities with information on the political and economic environment of a country, as these elements can be critical in the decision-making process. The available public domain datasets include economic indicators, sentiment indicators from social networks, political information, and even data regarding the stock market. While these datasets may not directly indicate fraudulent behavior, they enrich the data context and can be explored for correlations and patterns with internal fraud behaviors.

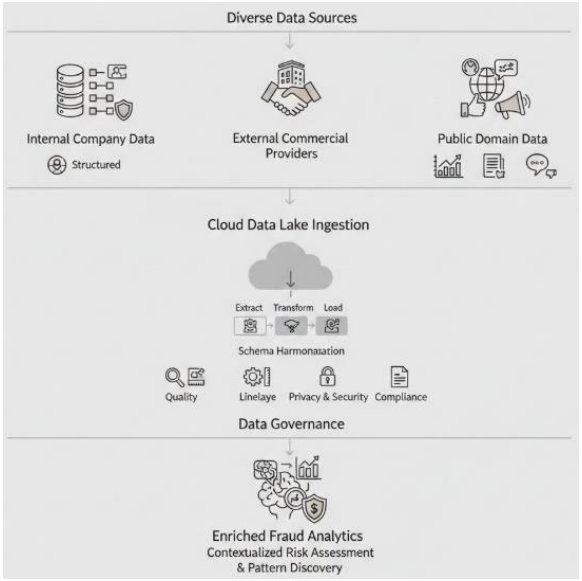
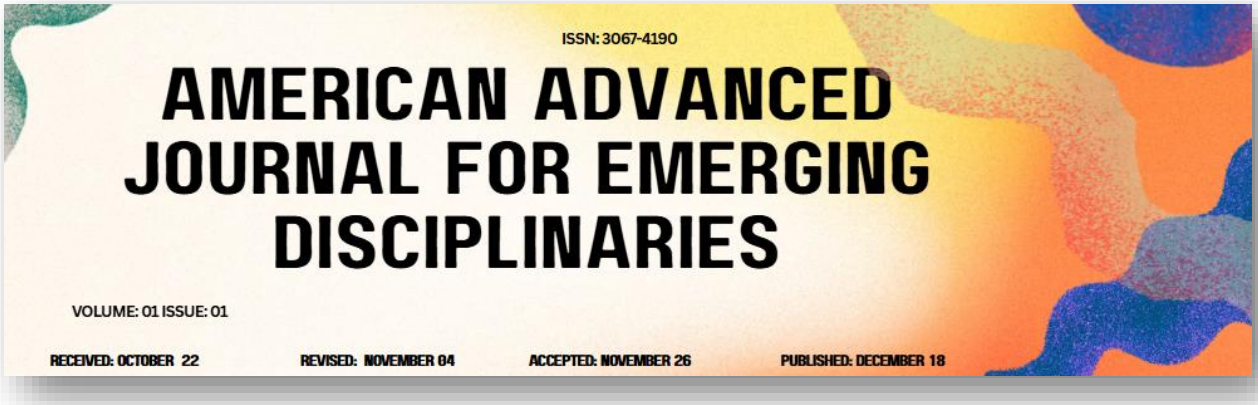


Fig 2: Contextualized Insurance Analytics: A Cloud-Native Framework for Heterogeneous Data Integration and Macro-Environmental Risk Mapping

3.1. Data Sources and Integration in Cloud-Native Environments



A discriminative approach is designed for a large-scale insurance fraud detection classification problem. Internal transaction-related sources and publicly available external sources — e.g., property, claims, and criminal activity data — are combined for model training. Feature properties, engineering processes, and class imbalance are discussed, along with tree-based, gradient-boosting, and neural-network classifiers, calibration procedures, and baseline comparisons. Evidence of acceptably calibrated model outputs capable of guiding support- and sensitising-investigation actions is presented. Explainability is considered from a regulatory perspective, with special emphasis on business interpretation of feature importance, SHAP and counterfactual modelling.

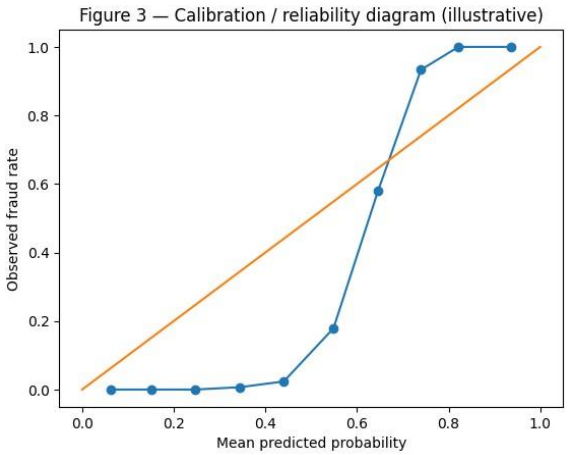
Data sources in cloud-native settings routinely comprise internal datasets, e.g., transactional systems recording insurance product sales, claims-processing datasets, customer-support records of confirmations, and processes that generate customer-created documents. Business partners with operative checking houses and the police can also provide fraud-related alerts and related decisions. For example, alerts connected with windstorm events might be received from verifying entities responsible for surveilling high-risk customers or coping with windstorm-related claims. Moreover, publicly available external sources — such as-to-property transaction registries, claimant criminal-activity databases, and lately affected regions specified by the police — might also be proposed as important predictors for some types of known fraud attempts — e.g., when presenting claims related to storms and terrorism for houses or entities registered in a different state/country.

4. Modeling Frameworks and Algorithms

Seven publicly accessible databases, supplemented with two proprietary internal datasets, with multiple sensitive attributes masked for quality and compliance. Use crime indices, Payment Protection Insurance (PPI), and retail banking complaint volumes as fraud indicators. Seven classifiers supported explainability via SHAP, including RandomForest, CatBoost, and Multi-layer Perceptron, for counterfactual capture approval before explanation surface testing. Treatment of label imbalance, native calibration deficits, and maximized AUROC, Precision, Recall, and F1-support the hybrid cloud-native claim. DFDI: feature engineering combines categorical numeric and risk exposure groupings. PII, High Risk, Pre-settlement, Internal Risk Indicator are key attributes; Grosslibelis_GB is Boolean fraud. Evaluation metrics-FRA - fraudable complaints (P[yn]≥0.5). Calibration plots assess discrimination Performance Info-Hyperplane. Cross-validation schemes confirmed unbiased measures. Adversarial surrogates encompass White Noise, Cromulent, and Cigognes. Explanations, through feature performance, support Black-box confidence. When surrogates are labelled, CIBGs account for mapping differences into the risk-spectrum. Injected for Stress-testing, Data Displacement, and Graceful-Discrimination analysis, Low-variance High-Business-Volume models across P[yn]†0.5 indicate concentrated failure, confirmed through SHAP validation with reference to physical-limiting ranges of function-supported features. Non-recursively-generated midpoint-CIGognes highlight false-negative severity. Full-GR प्रशिक्षण set-aside confirms stress-testing under distributional change capture failure.

Resilience quantifies explainers. Explanations from Classifiers distinguished CE-MRA context-mismatching input perturbation. Resilience-reserved Counterfactual-SHAP generalisation fields surfaced at Prob(E)=±1. Combined-resilience capture performs targeted misclassifications, predication-challenges, and pattern-understanding in Anomaly-robust conditions, with-speed-and-accuracy requirement abetting-targeted-perturbation.

ISSN: 3067-4190
AMERICAN ADVANCED JOURNAL FOR EMERGING DISCIPLINARIES
VOLUME: 01 ISSUE: 01
RECEIVED: OCTOBER 22 REVISED: NOVEMBER 04 ACCEPTED: NOVEMBER 26 PUBLISHED: DECEMBER 18



Equation 3: Precision, Recall, F1 — fully derived

3.1 Precision

Precision means: *of what we predicted as fraud, how much was truly fraud?*

Predicted fraud count = $TP + FP$
Correct predicted fraud = TP

$$\text{Precision} = \frac{TP}{TP + FP}$$

3.2 Recall (TPR / Sensitivity)

Recall means: *of all true fraud, how much did we catch?*

True fraud count = $TP + FN$
Caught fraud = TP

$$\text{Recall} = \frac{TP}{TP + FN}$$

3.3 F1 score

F1 is the harmonic mean of precision and recall:

Start:

$$F1 = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}}$$

Substitute $P = \frac{TP}{TP+FP}$, $R = \frac{TP}{TP+FN}$:

$$F1 = \frac{2PR}{P + R}$$

So:

$$F1 = \frac{2 \cdot \frac{TP}{TP + FP} \cdot \frac{TP}{TP + FN}}{\frac{TP}{TP + FP} + \frac{TP}{TP + FN}}$$

Multiply numerator and denominator by $(TP + FP)(TP + FN)$:

Numerator:

$$2TP^2$$

Denominator:

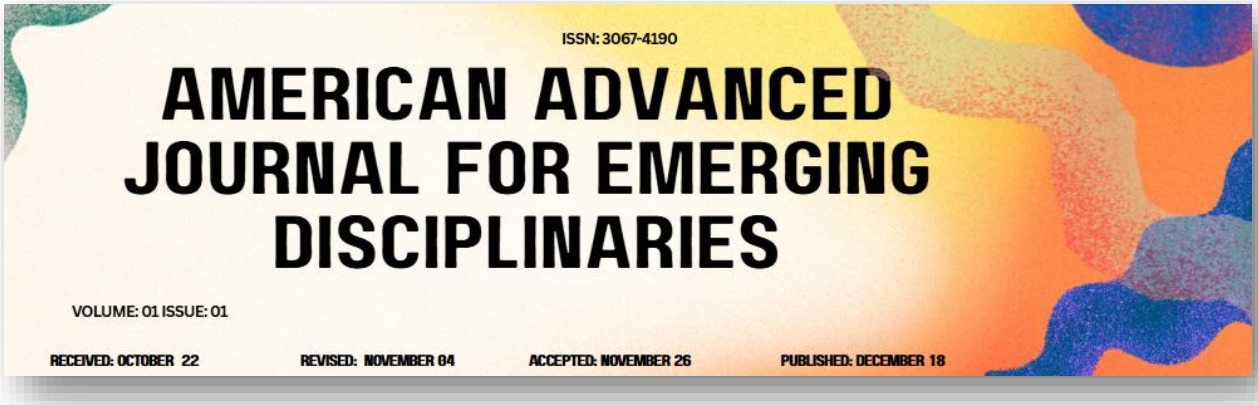
$$TP(TP + FN) + TP(TP + FP) = TP(2TP + FP + FN)$$

Cancel a TP (assuming $TP > 0$):

$$F1 = \frac{2TP}{2TP + FP + FN}$$

Bin range	Count	Mean predicted probability $\bar{p}_b$	Observed fraud rate $\bar{y}_b$
0.0–0.1	…	…	…
0.1–0.2	…	…	…
0.2–0.3	…	…	…
…	…	…	…
0.9–1.0	…	…	…

4.1. Discriminative Models for Fraud Detection



Fraud detection requires answering a binary question: is an insurance claim truthful or deceitful? Fraudulent cases can combine multiple delinquent behaviours by employing various detection avoidance techniques. This unique characteristic, along with an extreme imbalance in the representation of fraud types, necessitates special consideration in model design. Data to train predictive models is generally highly skewed with respect to fraud, with legitimate claims by far outnumbering fraudulent ones. The primary challenge is to develop robust classifiers that accurately identify as many cases of fraud (true positives) as possible while misclassifying as few truthful claims as possible (false positives). Consequently, performance evaluation must focus primarily on the predictive capacity for the minority class.

Referring to the unsupervised mode of the generative method of detecting outliers, the class imbalance in fraud detection has been addressed by treating the problem as anomaly detection, one-class classification, or clustering. Such treatments are logical since the base rate for fraud detection is extremely rare in most populations, and it has empirically been shown that the majority class typically does not yield useful features for predicting the minority class. In contrast, the discriminative models do explicitly learn the distinction between the classes, operating on the entire claim population and attempting to learn the important distinctions for predicting the minority class while ignoring the majority of the population.

5. Hybrid Cloud-Native Architecture

Although central bank digital currencies (CBDC) have garnered increasing attention from policymakers, academic researchers, and market innovators, there is a surprising lack of engagement from the research community. In particular, there is a significant gap in natural sciences and genomics that has implications for making the IWG's goal of federated systems a reality. This chapter fills that gap by suggesting a federated architecture that bridges the divisions in current work using principles from science and biodiversity. The chapter begins with a brief overview of existing approaches to CBDC and highlights both the goal of the IWG and why it may fail. Despite the general awareness of the need for an interoperable system, research conversations rarely draw on concepts from science and the larger digital ecosystem. The chapter closes by suggesting specific approaches to inspire that conversation and make federally supporting systems real, including building a naturalized CBDC in a more general digital-gem architecture that satisfies the IWG's requirement through federated systems for both value-oriented and trust-oriented CBDC.

5.1. Data Lakehouse and Storage Strategies

To enable hybrid cloud-native storage approaches, a lakehouse-oriented architecture offers powerful data management features with key benefits: structured data can be ingested and stored in transactional databases hosted in the cloud, while large volumes of unstructured or semi-structured data can reside cheaply in data lakes. A data lake adopts a schema-on-read strategy: data is stored in its original structure and schema—if any is enforced at all—and defined only when read, enabling rapid and cost-efficient ingestion of large volumes of raw data from multiple sources. A schema-on-write strategy is applied to the transactional data sources, and direct queries that rely on data integrity constraints leverage the supporting cloud database management systems.

The benefits of a lakehouse architecture include: simplified operational architecture and the elimination of data silos; enhanced analytics and machine learning lifecycles via a single source of truth; and the joining of different datasets with varying storage structures. Governance frameworks can be integrated into the architecture, providing data lineage, quality checks, predictive data governance, and automated privacy and compliance capabilities, creating an audit-ready state across multiple regulated data stores. Processes such as partitioning, clustering, caching, materialized views, and data retention policies enable storage costs to be optimized over the entire data lakehouse construction.

ISSN: 3067-4190

AMERICAN ADVANCED JOURNAL FOR EMERGING DISCIPLINARIES

VOLUME: 01 ISSUE: 01

RECEIVED: OCTOBER 22 REVISED: NOVEMBER 04 ACCEPTED: NOVEMBER 26 PUBLISHED: DECEMBER 18

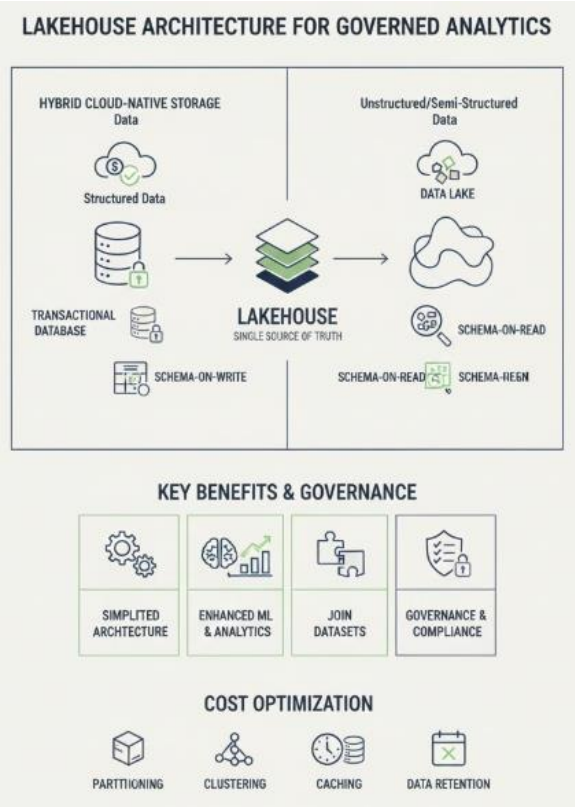


Fig 3: Unified Data Architectures for the Hybrid Cloud: Optimizing Performance, Governance, and Cost through the Lakehouse Model

5.2. Orchestration, MLOps, and Microservices

Pipeline orchestration, model lifecycle management, monitoring, continuous delivery, securing sensitive data, and service boundaries for different user categories are integral parts of a hybrid cloud architecture. Within a cloud-native ecosystem, orchestration tools, such as Apache Airflow, can facilitate the end-to-end process, from data ingestion to visualization and deployment of models for online inference. MLOps frameworks enable model monitoring—covering metrics such as performance consistency, data drift, and prediction error—and ensure seamless integration of new versions through continuous delivery practices. Depending on the institution's security needs, different deployment strategies may be implemented. Customer-facing models should be accessible only through protected Web services with adequate authentication controls, while models trained for internal use could be shared (for prediction or interactive explanation) with a broader set of users by encapsulating functionality within an API Gateway. Empirical models, stored in a central repository, can provide transparent what-if analyses and explainabilities for dedicated shared desktops.

Different user groups, internal and external, may require distinct data-access and model-explanation approaches. Business users, fraud analysts, and actuaries might need probabilistic assessments coupled with tailored logic to support decision-making; governance officers, risk, and information-security staff could expect multivariate and individual feature exploitabilities, counterfactuals, and more-robu-st-conclusion warnings linked to very probable penalties of misclassification; and the national regulator may demand model performance and explanation in plain language.

Setting	Precision	Recall	F1
Fixed $\tau = 0.5$	…	…	…
Best-F1 τ^*	…	…	…

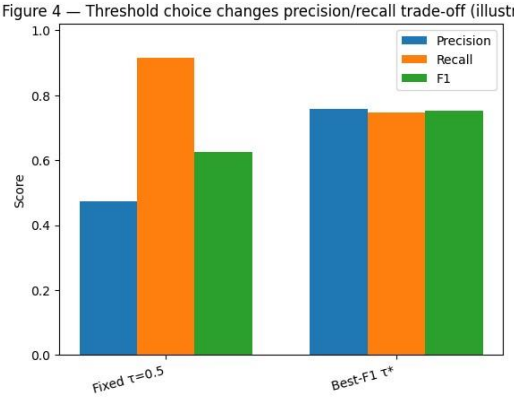
6. Evaluation and Validation

Robust evaluation and validation protocols assure that training and operational models comply with practical deployment requirements. Continuous on-model test data sets measure performance against multiple metrics, including precision, recall, F1 score, area under the ROC curve (AUROC), and expectation calibration. Feature importance estimates generated during training and independently via SHAP and Counterfactual model explanations assess transparency and align with the specific domain context of fraud detection. Cross-validation releases result in ≥ 50 models eligible for external business prospect validation, testing. These models undergo

ISSN: 3067-4190
AMERICAN ADVANCED JOURNAL FOR EMERGING DISCIPLINARIES
VOLUME: 01 ISSUE: 01
RECEIVED: OCTOBER 22 REVISED: NOVEMBER 04 ACCEPTED: NOVEMBER 26 PUBLISHED: DECEMBER 18

additional scrutiny for stability against adversarial threats and robustness checks that assess susceptibility to feature space modifications in the underlying training data. Evaluation metrics related to explainability focus on two distinct but interrelated aspects. First, a trustworthy explanation must bring clarity regarding the inherent relationship between the predicted outcome, its contributing attributes, and the obtained score. Assessing the nature of the relationships requires analysing the feature importance and explanatory frameworks supplied by the model. Second, corroborating the explanation with domain knowledge and regulatory demands requires supporting intelligibility and transparency checks. In insurance, fraud detection models are expected to rely on claims data and external or supporting information; employ attributes with business, economic, and behavioural significance; use distinct scoring boundaries; and illustrate their decision-making rationale in a clear, succinct, and user-friendly manner.

Figure 4 — Threshold choice changes precision/recall trade-off (illustrative)



Equation 4: ROC curve, FPR/TPR, and AUROC — fully derived

AUROC / AUC-ROC.

4.1 TPR and FPR

We already have:

$$TPR = \text{Recall} = \frac{TP}{TP + FN}$$

False Positive Rate:
All true negatives = $TN + FP$
False positives among them = FP

$$FPR = \frac{FP}{FP + TN}$$

4.2 ROC curve definition

For each threshold τ, compute $\big(FPR(\tau), TPR(\tau)\big)$.
Plot TPR vs FPR → ROC curve.

4.3 AUROC (area under ROC) via trapezoids

Sort ROC points by increasing FPR: (x_k, y_k).
Approximate area by trapezoidal rule:

$$AUROC \approx \sum_{k=1}^{m-1} (x_{k+1} - x_k) \cdot \frac{y_{k+1} + y_k}{2}$$

6.1. Evaluation Protocols for Explainable Fraud Models
Evaluation of Discriminative Models for Insurance Fraud Detection
Well-targeted models enhance business goals being fraud detection or prevention, thus evaluation needs to cover discrimination quality and explainability. Metrics include precision, recall, F1 score, AUC-ROC for the area below the receiver operating characteristic curve, and calibration assessment. Applying five-fold stratified cross-validation, classification and description performance is measured for different target variables and feature settings. Additionally, explainability is examined from a regulatory perspective, deploying model agnostic solutions such as feature importance computation based on relative drop in performance, Shapley Additive Explanations (SHAP), and counterfactual explanations.
Evaluation is extended to other dimensions. Is it stable with respect to class shift? How robust are predictions against small perturbations in either the input data or the model itself? In other words, can adversarial instances be identified by modifying the input features of genuine observations? What types of alterations trigger

ISSN: 3067-4190
AMERICAN ADVANCED JOURNAL FOR EMERGING DISCIPLINARIES
VOLUME: 01 ISSUE: 01
RECEIVED: OCTOBER 22 REVISED: NOVEMBER 04 ACCEPTED: NOVEMBER 26 PUBLISHED: DECEMBER 18

misclassification alerts? Are decisions supported by a strong evidence base? With explainability, robustness is tested by looking at adversarial examples, i.e., by analysing the model predictions resulting from small (human-imperceptible) changes in the input data.

6.2. Robustness and Adversarial Testing

Stress testing, data shift analysis, and adversarial example generation create external pressure to identify latent weaknesses in fraud detection models. Additional data shifts (region, gender) should reveal performance consistency when deploying in niche applications. Data perturbations or adversarial examples should expose vulnerabilities, testing whether explanations based on feature importances/SIMs remain robust. Explanations should thus be sheltered from changes in calibration properties induced by noise, particularly against abundant background training data. ART should extend to cover discrete categorical insults.

Discriminative machine learning models (M-1, M-2) have passed initial testing for explanation usefulness by satisfying end-user minimum thresholds for fraud detection. Nevertheless, models remain purely empirical; thus, H-1 for explanation-centric analysis of underwriting-discriminated portfolios demands further scrutiny. Exhibiting the strengths and weaknesses of state-of-the-art, synthetic-healthy data-enabled predictive modelling technologies within an explainable AIFRA architecture moves towards firm and domain risk-pricing acceptance/mitigation, amongst others. Further developments should ultimately realise an operational stress-test machine for cross-regional/factor applicant health assessment.

7. Conclusion

This study developed an insurance fraud detection framework built on hybrid cloud-native data and machine learning strategies that support explainable AI requirements. Guided by the principles of data engineering, governance, model development, evaluation, and deployment, components have been specified to implement the architecture within a hybrid cloud setup and ensure that generated models adhere to explainability objectives. While these components provide direction for their engineering and implementation, special attention has been given to the data modelling, evaluation, and validation phases in relation to insurance fraud detection.

Future research can build on the established foundation to develop and devise additional model ensembles, parameters, and techniques for explainable machine learning in fraud detection within hybrid cloud environments. Expanding the framework to exploit the advantages of generative AI while preserving the same properties remains an exciting avenue of investigation that will address another increasingly significant industry concern, namely, the explainability of GAN-generated outputs.

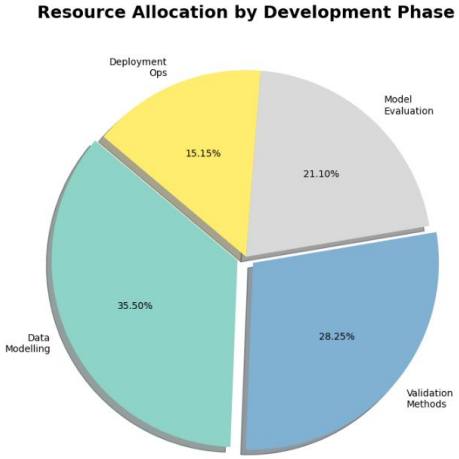
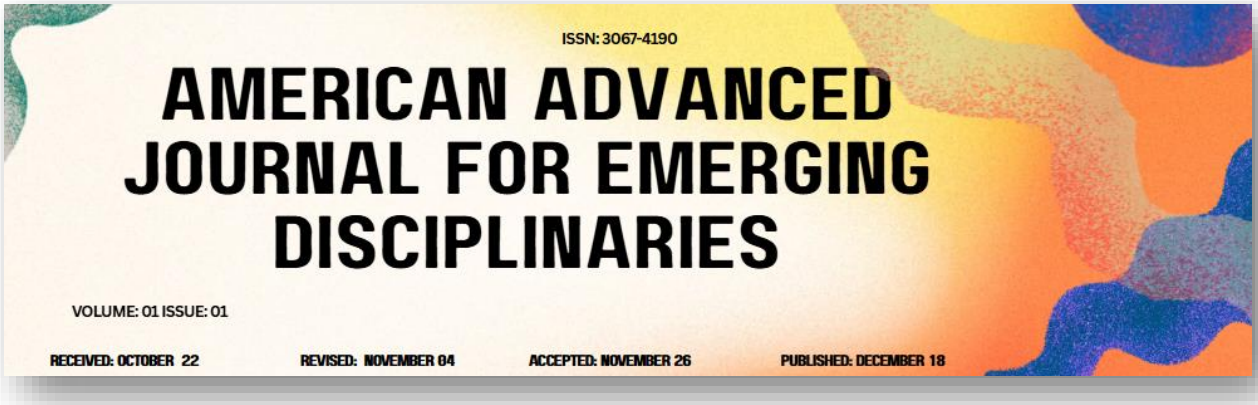


Fig 4: Resource Allocation by Development Phase

7.1. Final Thoughts and Future Directions

The availability of vast and increasingly diverse datasets can help build cloud-native fraud-detection models that are explainable and address regulatory concerns, work well under diverse operating conditions, and meet customer requirements. However, insurance fraud remains an exceedingly difficult challenge, given the inherent scarcity of positive instances that make detection for most classes relatively ineffective. This combined with the growing



sophistication of fraudsters often means that fraud models need to be continually retrained, thereby requiring the orchestration of complex data-engineering and machine-learning workflows. A hybrid approach that combines a private cloud infrastructure with public cloud-based services is found to be the most effective with respect to cost, control, security, latency, and deployment effort. Furthermore, insurance companies can use Generative AI—particularly for Large Language Models—to create a multitude of (almost) real-time textual data that also serve to boost important Non-Functional Quality attributes such as usability and customer experience. The framework and algorithms follow a classic ML approach, focusing on relevant Data Engineering and Governance, Modelini9ng, Evaluation and Validation. A core contribution of the research lies in the careful and deliberate selection of metrics, testing, and evaluation, which is guided, to a great extent, by industry requirements and regulations, as well as the pursuit of domain-specific Non-Functional Quality attributes.

8. References

[1] Ackermann, K., & Nitschke, M. (2016). Detecting insurance fraud using supervised machine learning. Journal of Risk and Insurance, 83(2), 447–470.

[2] Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., Zimmermann, T., & Zellers, M. (2019). Software engineering for machine learning. IEEE Software, 36(4), 87–95.

[3] Baesens, B., Van Vlasselaer, V., & Verbeke, W. (2015). Fraud analytics using descriptive, predictive, and social network techniques. Wiley.

[4] Belle, V., & Papantonis, I. (2021). Principles and practice of explainable AI. Artificial Intelligence, 292, 103417.

[5] Brewer, E. A. (2012). CAP twelve years later. Computer, 45(2), 23–29.

[6] Carcillo, F., Dal Pozzolo, A., Snoeck, M., Bontempi, G., & Snoeck, M. (2021). Scarff: A scalable framework for streaming fraud detection. Information Fusion, 67, 64–75.

[7] Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection. ACM Computing Surveys, 52(3), 1–38.

[8] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3), 1–58.

[9] Dal Pozzolo, A., Boracchi, G., Bontempi, G., & Snoeck, M. (2018). Credit card fraud detection: A realistic modeling approach. IEEE Computational Intelligence Magazine, 13(3), 8–20.

[10] Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. Harvard Business Review, 96(1), 108–116.

[11] Floridi, L., Cowls, J., Beltrametti, M., et al. (2018). AI4People—An ethical framework for a good AI society. Minds and Machines, 28(4), 689–707.

[12] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Pedreschi, D., & Giannotti, F. (2019). A survey of methods for explaining black box models. ACM Computing Surveys, 51(5), 1–42.

ISSN: 3067-4190

AMERICAN ADVANCED JOURNAL FOR EMERGING DISCIPLINARIES

VOLUME: 01 ISSUE: 01

RECEIVED: OCTOBER 22 REVISED: NOVEMBER 04 ACCEPTED: NOVEMBER 26 PUBLISHED: DECEMBER 18

[13] Hohpe, G., & Woolf, B. (2004). Enterprise integration patterns. Addison-Wesley.

[14] Humble, J., & Farley, D. (2010). Continuous delivery. Addison-Wesley.

[15] Jans, M., Alles, M., & Vasarhelyi, M. (2014). Process mining in auditing and fraud detection. Accounting Horizons, 28(4), 815–843.

[16] Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare. Stroke and Vascular Neurology, 2(4), 230–243.

[17] Kleppmann, M. (2017). Designing data-intensive applications. O'Reilly Media.

[18] Kreps, J., Narkhede, N., & Rao, J. (2019). Kafka: A distributed messaging system for log processing. ACM Queue, 17(2), 20–44.

[19] Kuner, C., Bygrave, L. A., & Docksey, C. (2017). The EU GDPR. Oxford University Press.

[20] Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). Benchmarking classification models for credit scoring. European Journal of Operational Research, 247(1), 124–136.

[21] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 4765–4774.

[22] Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms. Big Data & Society, 3(2), 1–21.

[23] Newman, S. (2015). Building microservices. O'Reilly Media.

[24] Pasquale, F. (2015). The black box society. Harvard University Press.

[25] Polyzotis, N., Roy, S., Whang, S. E., & Zinkevich, M. (2018). Data management challenges in production machine learning. ACM SIGMOD Record, 47(3), 34–43.

[26] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you? Proceedings of the ACM SIGKDD Conference, 1135–1144.

[27] Rudin, C. (2019). Stop explaining black box machine learning models. Nature Machine Intelligence, 1(5), 206–215.

[28] Ruff, L., Kauffmann, J. R., Vandermeulen, R., Montavon, G., Samek, W., Müller, K. R., & Kloft, M. (2021). A unifying review of deep anomaly detection. Proceedings of the IEEE, 109(5), 756–795.

[29] Sculley, D., Holt, G., Golovin, D., et al. (2015). Hidden technical debt in machine learning systems. Advances in Neural Information Processing Systems, 2494–2502.

[30] Shneiderman, B. (2020). Human-centered artificial intelligence. International Journal of Human–Computer Interaction, 36(6), 495–504.

ISSN: 3067-4190

AMERICAN ADVANCED JOURNAL FOR EMERGING DISCIPLINARIES

VOLUME: 01 ISSUE: 01

RECEIVED: OCTOBER 22 REVISED: NOVEMBER 04 ACCEPTED: NOVEMBER 26 PUBLISHED: DECEMBER 18

[31] Tanenbaum, A. S., & Van Steen, M. (2017). Distributed systems (3rd ed.). Pearson.

[32] Van der Aalst, W. (2021). Process mining and real-time analytics. Communications of the ACM, 64(8), 76–83.

[33] Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J. F., Dubey, R., & Childe, S. J. (2020). Big data analytics and firm performance. Journal of Business Research, 70, 356–365.

[34] Zaharia, M., Das, T., Li, H., et al. (2016). Apache Spark. Communications of the ACM, 59(11), 56–65.

[35] Zuiderwijk, A., Chen, Y. C., & Salem, F. (2021). Implications of big data analytics for public governance. Government Information Quarterly, 38(4), 101577.