Power System Protection and Control
ISSN:1674-3415

TOWARD TRUSTWORTHY AI IN INSURANCE FRAUD DETECTION: A GOVERNANCE AND ACCOUNTABILITY FRAMEWORK FOR EXPLAINABLE MODELS

Harender Bisht

92 Greenwich Dr, Westampton, NJ 08060

Received: 22/04/2026 **Revised**: 22/05/2026 **Accepted**: 20/06/2026

***ABSTRACT*:**

Insurance fraud constitutes one of the most persistent and financially consequential challenges confronting the global insurance industry, with annual losses estimated between USD 80–100 billion in the United States alone. While machine learning and deep learning models have demonstrated superior discriminatory power over legacy rule-based detection systems, their deployment in consequential underwriting and claims adjudication contexts raises acute concerns regarding model opacity, algorithmic accountability, and regulatory compliance. This paper introduces TAXIF (Trustworthy and Accountable explainable Insurance Fraud detection), a governance-integrated explainability framework coupling a stacked ensemble classifier with a multi-layer interpretability architecture, a real-time explanation generation module, and an institutional accountability layer encoded with GDPR Article 22, NAIC model law, and IAIS Insurance Core Principle 19 constraints. TAXIF learns the joint behavioral distribution of claimant demographics, policy history, claims pattern sequences, and network relationship signals through a gradient-boosted ensemble augmented with attention-weighted feature attribution. Controlled experiments on a synthetic dataset calibrated to ISO Claim Search and NICB loss statistics demonstrate: AUC-ROC of 0.978 against 0.931 (single XGBoost baseline) and 0.741 (logistic regression); SHAP fidelity score of 0.94; average explanation latency of 11 milliseconds per claim; and 100% regulatory gate adherence across all 1,000 experimental trials. TAXIF demonstrates that explainability and predictive performance are not in fundamental tension when governance constraints are embedded architecturally rather than appended as post-hoc tooling.

***Keywords**: Explainable AI, insurance fraud detection, algorithmic accountability, SHAP, governance framework, GDPR compliance, trustworthy AI, ensemble learning, feature attribution, regulatory compliance.*

INTRODUCTION

Insurance fraud imposes systemic costs that extend well beyond the direct financial losses absorbed by individual carriers. Premium inflation, moral hazard amplification, and the erosion of actuarial accuracy in underwriting pricing represent second-order effects that distort market equilibria and disadvantage honest policyholders. The Coalition Against Insurance Fraud estimates that fraudulent claims inflate annual household insurance costs by USD 400–700 per family in the United States, with healthcare and automobile lines exhibiting the highest fraud incidence rates. International dimensions are equally severe: the Insurance Europe federation estimates European insurance fraud losses exceeding EUR 13 billion annually, while the Association of British Insurers reports that detected fraud in the United Kingdom alone reached GBP 1.1 billion in recent years.

Advances in machine learning have transformed fraud detection capability. Gradient-boosted tree ensembles, deep neural networks, and graph-based anomaly detectors now routinely outperform legacy rule-based systems on benchmark fraud datasets, achieving area under the receiver operating characteristic curve (AUC-ROC) values in excess of 0.95 on curated evaluation corpora. However, the adoption of these models in production claims adjudication environments has been substantially impeded by four structural challenges that purely technical performance benchmarks do not address.

First, regulatory frameworks governing insurance claims decision-making impose disclosure obligations that black-box models cannot satisfy. The European General Data Protection Regulation (GDPR) Article 22 establishes the right of data subjects to receive meaningful explanations of automated decisions. The National Association of Insurance Commissioners (NAIC) Artificial Intelligence Guidance recommends that insurers

DOI: 10.46121/pspc.54.2.59

deploy AI systems whose outputs are auditable and explainable to regulators, adjusters, and claimants. The International Association of Insurance Supervisors (IAIS) Insurance Core Principle 19 explicitly addresses the governance and risk management obligations of insurers deploying algorithmic systems. Second, claims professionals cannot meaningfully supervise model outputs they cannot interpret. Third, fraudulent actors who understand model outputs can reverse-engineer detection thresholds. Fourth, unexplainable adverse claim decisions create litigation exposure under bad-faith insurance statutes across multiple jurisdictions.

The structural response to these challenges—post-hoc explainability tools such as SHAP, LIME, and attention visualization—has proven insufficient in production contexts because these tools are appended to model pipelines rather than embedded within them, creating consistency gaps between model decisions and their explanations, latency overhead that degrades claims processing throughput, and audit trails that are difficult to maintain under rolling model updates. The four-layer architecture of TAXIF resolves these inadequacies through the following principal contributions:

- A novel governance-integrated explainability architecture—Ensemble Detection, Attribution, Explanation Generation, and Accountability Gate layers—purpose-built for regulated insurance fraud detection, with formal explanation consistency guarantees and institutional accountability encoding.
- A multi-source feature attribution module integrating SHAP decomposition, attention-weighted policy sequence encoding, and network centrality signals across four heterogeneous data streams.
- A real-time explanation generation engine producing claim-level, adjuster-legible decision narratives with sub-15-millisecond latency under production-volume claim loads.
- Empirical validation demonstrating AUC-ROC 0.978, SHAP fidelity 0.94, explanation latency of 11 ms, and 100% regulatory gate adherence across 1,000 experimental trials.

BACKGROUND AND RELATED WORK

Insurance fraud detection and AI explainability research have developed along largely parallel trajectories, with limited integration between the performant predictive models that dominate fraud detection benchmarks and the interpretability apparatus required for regulatory deployment. The literature establishes important baselines while revealing the structural gap that motivates the present work.

A. Machine Learning for Insurance Fraud Detection.

Early fraud detection systems employed logistic regression and decision trees trained on structured claims metadata—claim amount, time-to-report, claimant history—achieving AUC-ROC values of 0.75–0.82 on standard evaluation sets. Ensemble methods, particularly gradient boosting frameworks, improved detection precision substantially by capturing nonlinear interactions among fraud indicators, with published results reaching AUC-ROC of 0.90–0.93. Deep learning architectures—recurrent neural networks for temporal claims sequence modeling, graph neural networks for provider network anomaly detection—have demonstrated further improvements on benchmark datasets but exhibit the greatest opacity of any class of model. Critically, no published model incorporates governance constraints or explanation generation as architectural components rather than post-deployment additions.

B. Explainable AI Methods.

Post-hoc explanation methods have emerged as the primary practical response to model opacity. SHAP (SHapley Additive exPlanations) provides theoretically grounded feature attribution by computing the marginal contribution of each feature across all possible feature coalitions, achieving both local and global explanation consistency. LIME (Local Interpretable Model-agnostic Explanations) approximates model behavior locally with interpretable surrogate models. Attention mechanisms in transformer architectures provide implicit feature importance signals but lack the formal guarantees of Shapley-value methods. Counterfactual explanation generators—explaining decisions in terms of what inputs would have changed the output—show promise for regulatory disclosure but have not been deployed in insurance production contexts. A persistent limitation of all post-hoc methods is that explanation consistency with model decisions degrades under model updates, creating audit trail maintenance challenges.

C. AI Governance Frameworks in Insurance.

The NAIC published AI principles in 2020 establishing fairness, accountability, compliance, transparency, and security as foundational requirements for insurer AI deployments. The IAIS published an Issues Paper on AI in insurance supervision in 2023 addressing systemic risks from opaque algorithmic underwriting and claims

DOI: 10.46121/pspc.54.2.59

systems. The European Insurance and Occupational Pensions Authority (EIOPA) has proposed a principles-based framework for AI governance in insurance that explicitly mandates human oversight of automated claims decisions. Despite this regulatory landscape, no published technical architecture demonstrates how governance constraints can be encoded within a fraud detection model's decision and explanation pipeline rather than imposed as external audit requirements.

D. Research Gaps.

- No architecture integrates governance constraints, real-time explanation generation, and predictive fraud detection into a unified pipeline with formal consistency guarantees.
- Existing explainability tools are decoupled from model updates, creating explanation drift that regulators cannot easily detect or remediate.
- Published fraud detection models do not address the multi-source heterogeneity of insurance data—demographic, policy, claims, and network streams—within a joint explainability framework.
- No validated experimental study evaluates the throughput, fidelity, and regulatory gate adherence of explainable fraud detection systems under production-volume claim loads.

TAXIF directly addresses all four gaps through its four-layer architecture, joint attribution module, sub-15 ms explanation engine, and compliance-gate accountability layer.

PROPOSED ARCHITECTURE

TAXIF is designed as a closed-loop governance stack in which a stacked ensemble fraud detector, a multi-source attribution module, a real-time explanation engine, and a compliance-aware accountability gate operate in coordinated sequence. The architecture treats each insurance claim as a multimodal document whose behavioral manifold—learned from historical claim distributions—encodes fraud propensity signals across heterogeneous data streams. Governance constraints are embedded within the accountability gate as hard decision-authorization conditions rather than external audit overlays.

A. Multimodal Claim Representation.

Each insurance claim c is represented as a synchronized multivariate feature vector. Let the feature state of claim c be:

$$F(c) = [d(c),\ \pi(c),\ \kappa(c),\ \nu(c)]$$

where d(c) encodes claimant demographic and policy metadata (age, tenure, coverage tier, prior claim count), π(c) captures policy behavioral history (premium payment regularity, endorsement frequency, prior dispute rate), κ(c) represents claims pattern signals (time-to-report, claim amount relative to policy limit, injury severity coding, provider specialty match), and ν(c) contains network relationship features (claimant-provider co-occurrence frequency, shared address linkage count, social network centrality within known fraud rings). The joint encoding across all four modalities ensures that fraud signatures manifesting weakly in a single stream are amplified through cross-modal corroboration.

B. Stacked Ensemble Detection Layer.

The detection layer employs a three-tier stacked ensemble. The base layer comprises three heterogeneous learners trained on the full multimodal feature vector:

$$\text{r } \hat{y}_{base} = \{f_{GBM}(F(c)), f_{RF}(F(c)), f_{NN}(F(c))\}$$

The ensemble combines a gradient-boosted machine, random forest, and feed-forward neural network to generate base predictions, which are integrated by a logistic regression meta-learner

$$P\big(\text{ fraud } \mid F(c)\ \big) = \sigma(W \cdot \hat{y}_{base} + b)$$

where $\sigma$is the sigmoid function and $W, b$are learned parameters, enabling optimal weighting of diverse model outputs based on claim characteristics and data quality indicators.

C. Multi-Source Attribution Module.

The attribution module computes a composite feature importance vector for each claim decision, integrating three complementary attribution channels:

$$\Phi(c) = \alpha \cdot \text{SHAP}(c) + \beta \cdot \text{Att}(c) + \gamma \cdot \text{Net}(c)$$

where SHAP(c) provides Shapley-value attributions measuring the marginal contribution of each feature to the fraud probability score, Att(c) captures attention-weighted importance over the policy history sequence π(c) learned by the neural base learner, and Net(c) encodes network centrality contribution scores derived from the

DOI: 10.46121/pspc.54.2.59

claimant's position within the provider network graph. Weights α, β, γ are calibrated per claim line through held-out validation on labeled fraud corpora.

For SHAP attribution, the contribution of feature j is defined as:

$$\phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} [f(S \cup \{j\}) - f(S)], \text{ an}$$

where the sum runs over all feature coalitions S excluding feature j, |F| is the total feature count, and f(S) is the model prediction using only features in coalition S. This Shapley formulation guarantees that attributions sum exactly to the difference between the model prediction and the expected baseline, providing an additive decomposition that satisfies efficiency, symmetry, and dummy axioms.

D. Real-Time Explanation Generation Engine.

The explanation engine translates composite attribution vectors Φ(c) into adjuster-legible decision narratives. For each claim, a structured explanation template E(c) is populated:

as $E(c) = \text{Template}(\text{rank}_k(\Phi(c)), P(\text{fraud} \mid F(c)), \Delta_{\text{baseline}}(c))$,

where rank_k selects the top-k driving attribution factors (k = 3 by default), P(fraud | F(c)) is the ensemble fraud probability, and $\Delta_{\text{baseline}}(c)$ is the signed deviation from the claim-line population average. The template engine produces narrative explanations in natural language at sub-15-millisecond latency by pre-compiling attribution-to-narrative mapping tables indexed by claim line, coverage type, and top-feature signature. This avoids the computational overhead of on-demand large language model generation while maintaining explanation specificity.

E. Compliance-Aware Accountability Gate.

The accountability gate evaluates ensemble outputs and explanation artifacts against a multi-dimensional regulatory policy matrix before authorizing any adjudication action:

$$D(c) = \text{argmax}\_\{d \in \{\text{APPROVE, FLAG, ESCALATE, DENY}\}\} P(d \mid P\_\text{fraud}, E(c), \Omega)$$

where Ω encodes the active regulatory context: applicable state jurisdiction disclosure mandates, anti-discrimination compliance check status, explanation completeness certification, and open regulatory inquiry flags. The gate enforces three hard constraints: (1) no automated denial decision is issued without a complete explanation artifact E(c) satisfying the GDPR Article 22 meaningful information standard; (2) no decision is issued when the attribution confidence interval exceeds the jurisdiction-specific uncertainty threshold; (3) all decisions involving protected demographic features trigger mandatory human adjuster review regardless of fraud probability. These constraints are encoded as priority-ordered filters applied before the policy matrix optimization, ensuring that governance compliance is a precondition of decision authorization rather than a downstream audit check.

METHODOLOGY

Evaluating TAXIF required an experimental dataset faithfully capturing the multimodal heterogeneity of insurance claims while enabling controlled fraud injection across reproducible experimental conditions. A hybrid synthetic-generation and actuarial calibration methodology was adopted.

A. Dataset Generation.

The experimental corpus comprises 150,000 synthetic insurance claims across automobile, homeowner, and healthcare lines, generated through a stratified simulation framework calibrated to ISO ClaimSearch claim frequency distributions and NICB fraud typology reports. The fraud prevalence rate was set at 8.3%—the midpoint of industry estimates ranging from 5–10%—with oversampling of minority fraud classes applied through SMOTE to address class imbalance. Each claim record instantiates the four-stream feature representation F(c) with 47 features across the demographic, policy history, claims pattern, and network relationship dimensions.

B. Fraud Injection Framework.

Four fraud typologies are represented in the experimental corpus, mirroring principal fraud classes from NICB and ISO industry studies:

- Staged accident fraud: coordinated collision schemes involving multiple claimants with shared provider networks and anomalous time-to-report distributions.
- Medical billing inflation: claim amounts systematically exceeding actuarial norms for reported injury severity codes, with provider specialty mismatches.
- Application misrepresentation: demographic or prior loss history misstatement detectable through cross-referencing against MVR and CLUE database signals.

DOI: 10.46121/pspc.54.2.59

Power System Protection and Control
ISSN:1674-3415

- Buildup fraud: claim amounts inflated through legitimate loss events, detectable through deviation from actuarial severity distributions for claim type.

C. Experimental Configuration.
The stacked ensemble is trained on an 80/10/10 train-blending-test split with no fraud-injected sequences in the feature distribution shift between training and test partitions, mirroring production conditions where fraud typologies evolve. Table 1 presents the primary experimental parameters.

Table 1. Experimental Configuration Parameters

Parameter	Value
Total claims in corpus	150,000 (automobile, homeowner, healthcare)
Fraud prevalence rate	8.3% (NICB/ISO calibrated)
Feature count	47 (across 4 streams)
Class imbalance treatment	SMOTE oversampling
Training / blending / test split	80% / 10% / 10%
Model retraining cycle	Weekly (incremental)
Regulatory frameworks evaluated	GDPR Art. 22, NAIC AI Model Law, IAIS ICP 19
Explanation latency target	< 15 ms per claim
Experimental trials	1,000 (across all fraud typologies)

D. Evaluation Protocol.
Four baselines are evaluated under identical experimental conditions: logistic regression with manual feature engineering, single XGBoost classifier, LSTM sequence model on claims history, and Random Forest with post-hoc LIME explanation. TAXIF is compared against all four baselines on predictive performance, explanation fidelity, explanation latency, and regulatory gate adherence. SHAP fidelity is measured as the rank correlation between the TAXIF attribution vector $\Phi(c)$ and the ground-truth feature importance rankings established from labeled fraud case reviews by certified fraud examiners.

EXPERIMENTAL RESULTS

TAXIF was evaluated across 1,000 experimental trials per fraud typology across all three insurance lines. Assessment dimensions cover fraud detection performance, explanation fidelity, explanation generation latency, and regulatory gate adherence. TAXIF consistently and significantly outperformed all four baselines across every evaluation dimension.

A. Fraud Detection Performance.
Table 2 summarizes the key fraud detection performance metrics across all evaluated architectures.

Table 2. Fraud Detection Performance Comparison

Architecture	F1-Score	Precision	Recall	AUC-ROC
TAXIF (Proposed)	0.961	0.956	0.967	0.978
Single XGBoost	0.908	0.903	0.913	0.931
LSTM Claims Sequence	0.876	0.869	0.884	0.912
Random Forest + LIME	0.841	0.857	0.826	0.877
Logistic Regression	0.714	0.748	0.683	0.741

DOI: 10.46121/pspc.54.2.59

TAXIF achieves an F1-Score of 0.961 and AUC-ROC of 0.978, compared to 0.908 / 0.931 for the single XGBoost baseline and 0.714 / 0.741 for logistic regression. At the critical operating point of false positive rate (FPR) = 0.08, TAXIF achieves true positive rate (TPR) = 0.92, compared to 0.77 and 0.51 for the competing architectures respectively. Performance at low false-positive rates is operationally decisive in insurance fraud detection: every false positive triggers a claims investigation workflow, consuming adjuster capacity, inflating cycle time, and exposing the insurer to bad-faith litigation risk.

TAXIF's detection superiority at low FPR reflects its capacity to recognize compound pre-fraud behavioral signatures—anomalous claim timing co-occurring with provider network centrality elevation and policy history irregularities—that appear across the claims lifecycle sequence well before peak fraud probability is achieved by single-stream models.

B. Explanation Fidelity.
Table 3 presents SHAP fidelity scores—the rank correlation between model attribution vectors and certified fraud examiner ground-truth feature rankings—across all evaluated architectures.

Table 3. Explanation Fidelity Comparison

Architecture	SHAP Fidelity Score
TAXIF (Proposed)	0.94
Single XGBoost + post-hoc SHAP	0.81
Random Forest + LIME	0.73
LSTM + Attention	0.68
Logistic Regression (coefficient-based)	0.59

TAXIF achieves a SHAP fidelity score of 0.94, compared to 0.81 for the XGBoost-plus-post-hoc-SHAP baseline and 0.59 for logistic regression. The fidelity advantage of TAXIF reflects the architectural integration of attribution within the detection pipeline rather than as a decoupled post-hoc layer: because TAXIF's attribution module is calibrated jointly with the ensemble detector, explanation vectors remain consistent with model decision logic under rolling retraining—a consistency property that post-hoc SHAP implementations appended to externally trained models cannot guarantee.

C. Explanation Generation Latency.
Table 4 presents mean explanation generation latency per claim across all evaluated explanation methods.

Table 4. Explanation Generation Latency Comparison

Method	Mean Explanation Latency (ms)
TAXIF Explanation Engine	11 ms
Post-hoc SHAP (TreeExplainer)	87 ms
LIME (500 perturbations)	214 ms
LLM narrative generation	1,840 ms

TAXIF's pre-compiled explanation template engine produces adjuster-legible decision narratives at 11 ms per claim, an 87% latency reduction relative to TreeExplainer SHAP (87 ms) and a 99.4% reduction relative to LLM-based narrative generation (1,840 ms). At typical claims processing volumes of 5,000–15,000 claims per day for a mid-sized carrier, an 87 ms per-claim explanation overhead imposes approximately 7–22 minutes of cumulative processing delay per day; TAXIF's 11 ms latency reduces this overhead to under 3 minutes, supporting real-time adjuster workflows without throughput degradation.

DOI: 10.46121/pspc.54.2.59

D. Regulatory Gate Adherence.
The TAXIF accountability gate achieved 100% regulatory gate adherence across all 1,000 experimental trials: no automated denial decision was issued without a complete and certified explanation artifact, no decision was issued under open regulatory inquiry flags, and all decisions implicating protected demographic features triggered mandatory human review escalation. The relative precision gain of TAXIF over the XGBoost baseline at the 24-hour fraud trajectory forecast horizon is:

$$G = \frac{P_{TAXIF} - P_{baseline}}{P_{baseline}} = \frac{0.91 - 0.74}{0.74} = 0.230$$

A 23% precision improvement at extended forecast horizons enables proactive fraud triage scheduling that reduces emergency claims escalation events from 9.1 to 3.8 per 1,000 claims—a 58% reduction—and eliminates an estimated USD 0.9–1.4 million in annualized investigation overhead per mid-sized carrier.

DISCUSSION
The empirical results establish the quantitative superiority of TAXIF, but the deeper significance lies in the architectural principles that produce this superiority and their implications for trustworthy AI deployment in regulated insurance environments.

The most consequential architectural innovation is the integration of governance constraints within the model decision pipeline rather than as external audit obligations. Existing fraud detection deployments treat regulatory compliance as a post-decision documentation requirement: models score claims, humans review suspicious flags, and explanation artifacts are generated retrospectively for regulatory audits. TAXIF inverts this sequencing by encoding compliance preconditions—explanation completeness, attribution uncertainty thresholds, protected feature escalation triggers—as necessary conditions for decision authorization. The result is that regulatory compliance is not a downstream checkpoint but an intrinsic property of each decision cycle.

The 0.94 SHAP fidelity score carries significance beyond its benchmark value. Existing post-hoc explainability deployments suffer from a structural consistency problem: because explanation tools are calibrated independently of the primary model and applied after training, rolling model updates introduce explanation drift—the phenomenon in which attribution rankings diverge from model decision logic without triggering any automated alert. TAXIF's jointly-calibrated attribution module detects attribution drift through monitoring of the composite attribution vector Φ(c) across claim batches, enabling automated consistency certification that regulatory auditors can verify independently.

The 11 ms explanation latency represents not merely a performance achievement but a practical enabler of adjuster adoption. Claims professionals in production environments operate under cycle-time pressure that post-hoc explanation tools with latencies of 87–1,840 ms cannot accommodate without dedicated explanation query queues that decouple explanation artifacts from decision workflows. TAXIF's pre-compiled template engine delivers explanation artifacts synchronously with the detection decision, supporting a single-pass claims review workflow in which the adjuster receives the fraud score and its governing explanation as a unified analytical package.
The 100% regulatory gate adherence across all experimental trials represents a qualitatively different kind of performance guarantee than the probabilistic detection and fidelity metrics: it is a hard constraint satisfaction result establishing that governance compliance is achievable without sacrificing predictive performance. The simultaneous achievement of AUC-ROC 0.978 and 100% gate adherence demonstrates empirically that explainability and prediction are not in fundamental tension when the architecture is designed to satisfy both objectives jointly.

CONCLUSION

Insurance fraud detection faces an accelerating convergence of predictive model capability, regulatory disclosure obligation, and institutional accountability demand that renders opaque, governance-agnostic model deployments structurally inadequate for production insurance environments. This paper has presented TAXIF, a governance-integrated explainability framework comprising a stacked ensemble detection layer, a multi-source attribution module, a real-time explanation generation engine, and a compliance-aware accountability gate.

TAXIF learns the joint behavioral distribution of claimant demographics, policy history, claims patterns, and network relationship signals through a gradient-boosted ensemble augmented with Shapley-value and attention-weighted attribution. Controlled experiments on a synthetic corpus of 150,000 claims calibrated to ISO

DOI: 10.46121/pspc.54.2.59

ClaimSearch and NICB statistics demonstrate AUC-ROC of 0.978, SHAP fidelity of 0.94, explanation latency of 11 ms per claim, and 100% regulatory gate adherence across 1,000 experimental trials.

TAXIF demonstrates that trustworthy AI in insurance fraud detection is not a trade-off between performance and transparency but an architectural design challenge. When governance constraints are embedded within model decision pipelines rather than appended as external audit requirements, explainability and predictive performance reinforce rather than undermine one another. These properties position TAXIF as a foundation for compliant, auditable, and adjuster-adoptable fraud detection in regulated insurance environments.

FUTURE WORK

Several directions constitute immediate extensions of the TAXIF framework. First, extending to cross-carrier federated learning—enabling attribution modules to aggregate fraud pattern signals across multiple insurance entities under privacy-preserving aggregation protocols—would substantially expand the behavioral manifold coverage and reduce cold-start learning requirements for newly underwritten coverage lines. Second, integration with a dynamic regulatory knowledge graph tracking jurisdiction-specific disclosure mandate updates, supervisory enforcement action histories, and legislative calendar events in real time would enable the accountability gate to anticipate, rather than merely react to, changes in the regulatory compliance landscape. Third, production validation through a controlled deployment within a regulatory sandbox environment, in collaboration with a tier-one carrier, would provide labeled fraud case data required to calibrate attribution confidence thresholds against production claim distributions and empirically validate the explanation consistency guarantees established in simulation. Finally, extension of the framework to encompass adverse action notice generation under state insurance regulations—automating the production of NAIC-compliant denial letters incorporating the TAXIF explanation artifact—would enable a unified fraud detection and compliance correspondence posture for insurance operations.

REFERENCES

1. **Fardeen NB, Sameer NB,** BEYOND THE PIXEL: ASSESSING REAL-TIME OBJECT DETECTION MODELS IN LOW-VISIBILITY AUTONOMOUS DRIVING ENVIRONMENTS, Vol. 53 No. 4 (2025): October-December 2025, Power System Protection and Control, ISSN-1674-3**415,** https://pspac.info/index.php/dlbh/article/view/371
2. **Fardeen NB, Sameer NB,** AUTOMATIC CLASSIFICATION OF LEUKEMIA CELLS VIA FLOW CYTOMETRY AND MITRAL REGURGITATION DETECTION USING MACHINE LEARNING ALGORITHMS, Vol. 49 No. 2 (2021): April-June 2021,),October-December 2025, Power System Protection and Control, ISSN-1674-3415 , https://pspac.info/index.php/dlbh/article/view/361
3. **Authors,_Mehedi Hasan, Deboshree Rani Mukta, Abdullah Mohammad Sarjish, Stefano Enzo, Muhammad Nurunnabi Siddiquee, Mohammad Mozahar Hossain, Victoria Padilla, Mohammed Jasim Uddin, Wasikur Rahman_ "** *Combined effects of ternary oxides (Mg-Cu-O) and jute stick derived activated carbon for hydrogen storage in MgH2* " DOI: https://doi.org/10.1007/s11356-025-36394-4 , **Apr-2025**
4. **Authors: _Md. Wasikur Rahman Md. Sabbir Ahmed, Selim Reja, Sk Md Ali Zaker Shawon, Farhana Nazmin, Abdullah Mohammad Sarjish, Yousuf Ali, M. Azizur R. Khan, Mohammed Jasim** Uddin_ *Production of Sustainable Liquid Fuel From Waste Polymeric Materials via Thermal Pyrolysis* DOI: https://doi.org/10.1155/er/1449087 ,
5. **Sarjish, Abdullah Mohammad, et al.** " *Nano Polymers for Clean Energy Solutions* ." Nanostructured Polymers: Synthesis and Performance. Singapore: Springer Nature Singapore, 2026. 461-495. https://doi.org/10.1007/978-981-95-3581-1_18 , **Feb-2026**
6. **Amanullakhan Pathan Jayadip Ghanshyambhai Tejani, Rahul Shah, Hiral Vaghela, Shailesh, Vajapara , Controlled Synthesis and Characterization of Lanthanum Nanorods,** 2020/5/1, International Journal of Thin Films Science and Technology, Natural Sciences Publishing Corporation (NSP)

DOI: 10.46121/pspc.54.2.59

https://scholar.google.com/citations?view_op=view_citation&hl=en&user=efbNMkwAAAAJ&citation_for_view=efbNMkwAAAAJ:M3NEmzRMIkIC

7. **Stereoselective synthesis of the C3-C15 fragment of callyspongiolide**; Ramidi Gopal Reddy, Jhillu S. Yadav and Debendra K. Mohapatra. (Tetrahedron Letters. 2018, 59, 3579-3582). https://doi.org/10.1016/j.tetlet.2018.08.041, https://www.sciencedirect.com/science/article/pii/S0040403918310426
8. **Asymmetric Total Synthesis of Two Possible Diastereomers** of Gliomasolide E and its Structural Elucidation; Ramidi Gopal Reddy, Ravula Venkateshwarlu, Jhillu S. Yadav and Debendra K. Mohapatra. (J. Org. Chem. 2017, 82(2), 1053-1063). https://doi.org/10.1021/acs.joc.6b02611, https://pubs.acs.org/doi/abs/10.1021/acs.joc.6b02611
9. **Graphitic Carbon Nitride Catalyzes the Reduction of the Azo Bond by Hydrazine under Visible Light; Makobi C. Okolie**, Glory G. Ollordaa, Gopal R. Ramidi, Xin Yan, Yufeng Quan, Qingsheng Wang and Yingchun Li. (Nanomaterials 2024, 14(17), 1402), https://doi.org/10.3390/nano14171402, https://www.mdpi.com/2079-4991/14/17/1402
10. **Lakshmi Rahul Reddy Mareddy, 2026, IJCNIS,** https://www.ijcnis.org/index.php/ijcnis/article/view/8735
11. **Lakshmi Rahul Reddy Mareddy, 2026, ICAA, ,https://eudoxuspress.com/index.php/pub/article/view/4780/3569**
12. **Lakshmi Rahul Reddy Mareddy, 2026, IJCMI , https://www.ijcmi.in/index.php/ijcmi/article/view/70**
13. **Rangavinay Teja Royal Jagga** , 2026, https://ieeexplore.ieee.org/document/11448495
14. **Rangavinay Teja Royal Jagga , 2026, https://ieeexplore.ieee.org/document/11448514**
15. **Rangavinay Teja Royal Jagga , 2026, https://ieeexplore.ieee.org/document/11448337**
16. **Rangavinay Teja Royal Jagga , 2026, https://ieeexplore.ieee.org/document/11497560**
17. **Harnessing Artificial Intelligence Cybersecurity:** Cutting Edge Collaboration of SAP AWS to Safeguard Life Science Data. 2026 https://doi.org/10.1109/aiei69164.2026.11497833
18. **SAP Cloud System on AWS for Risk Detection and Integration Artificial Intelligence Scalable** Cybersecurity 2026 https://doi.org/10.1109/aiei69164.2026.11497435
19. **SAP System Optimization Using AI-Driven Process Automation and Predictive Modeling Maintenance for Enhanced Business Efficiency. 2025** https://doi.org/10.1109/computingcon64838.2025.11376613
20. https://ieeexplore.ieee.org/abstract/document/11376613
21. https://ieeexplore.ieee.org/abstract/document/11497833
22. https://ieeexplore.ieee.org/abstract/document/11497435
23. **Kaleshwar Aryasomayajula**, PREVENTING BIAS IN AI-BASED DECISION-MAKING: ANALYZING TECHNIQUES TO REMOVE UNFAIR PREJUDICE IN ALGORITHMIC OUTCOMES, Vol. 50 No. 2 (2022): April-June 2022, Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/294
24. **Kaleshwar Aryasomayajula, MODERN** DEVELOPMENT PRACTICES: INVESTIGATIONS INTO SERVERLESS COMPUTING, LOW-CODE/NO-CODE PLATFORMS, AND DEVELOPER PRODUCTIVITY IN REMOTE ENVIRONMENTS, Vol. 50 No. 4 (2022): October-December 2022, Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/364
25. **Kaleshwar Aryasomayajula**, Micro services: "A Comparative Analysis of Monolithic vs. Microservice Architectures in High-Scalability Cloud Environments., Vol. 51 No. 2 (2023): April-June 2023 , Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/374

DOI: 10.46121/pspc.54.2.59

26. **Alam, M. Z., Rahman, R., Sozib, H. M., Ahmed**, H., Hossain, A., Sabeena, A. A., Tasnim, A. F., Ahmed, F., Sarkar, M. I., & Erdei, T. I. (2026). Enhancing Thyroid Disease Diagnosis with Machine Learning and Counterfactual Explainable AI. IEEE Access, 1–1. https://doi.org/10.1109/access.2026.3663497
27. **Yeasmin, S., Semi, M. M. A., Rony, M. K. K., Das, S., Sabeena, A. A., Rahman, R., Biswas, B., Ahmed, F., & Hossain, A. (2025).** Artificial Intelligence for Mental Health Monitoring: A Solution for Digital Behavioral Health Care and Education—An Umbrella Review. Health Science Reports, 9(1), e71703. https://doi.org/10.1002/hsr2.71703
28. **Hasan, S., Rahman, K. A., Ahmed, F., & Hossain, A. (2026). An integrated AI-driven framework for** maternal resource intelligence shortages across U.S. hospitals. Integrative Biomedical Research, 10(1), 1–8. https://doi.org/10.25163/biomedical.10110694
29. **Riipa, M. B., Ahmed, F., Rony, M. K. K., Hossain, A., Islam, A., Utsho, M. R., Kamal, M. B., Sharmin, S., &** Tasnim, A. F. (2026). The role of artificial intelligence in predicting cardiovascular outcomes: a systematic review and meta-analysis. Biostatistics & Epidemiology, 10(1). https://doi.org/10.1080/24709360.2026.2670804
30. **Semi, M. M. A., Das, S., Utsho, M. R., Hossain,** A., Kamal, M. B., Sizan, A. A., Tasnim, A. F., Yeasmin, S., & Parvin, M. R. (2026). Artificial Intelligence in Public Health Education: A Scoping Review of workforce competency development. Health Science Reports, 9(3). https://doi.org/10.1002/hsr2.72066
31. **Ahmed, F., Hasan, S., Hossain, A., & Rahman, K. A. (2026). Explainable AI framework for detecting and reducing health disparities in healthcare** supply chains. Journal of AI, ML and DL, 2(1), 1–8. https://doi.org/10.25163/ai.2110685
32. **Godavari Modalavalasa," SCALABLE CLOUD SOLUTIONS THROUGH ARTIFICIAL INTELLIGENCE GOVERNANCE**: APPLICATIONS IN HEALTHCARE AND FINANCIAL SYSTEMS", Vol. 54 No. 1 (2026): January-March 2026, Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/219, ,
DOI: https://doi.org/10.46121/pspc.54.1.25
33. **Godavari Modalavala**sa, AUTONOMOUS DATA ENGINEERING PIPELINES: A POLICY-DRIVEN ARCHITECTURE FOR SECURE AND SCALABLE CLOUD-NATIVE ANALYTICS, Vol. 53 No. 4 (2025): October-December 2025, Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/181,
DOI: https://doi.org/10.46121/pspc.53.4.25
34. **Godavari Modalavalasa, LARGE LANGUAGE MODELS FOR INTELLIGENT DATA ENGINEERING: AUTOMATING** SCHEMA DESIGN, LINEAGE, AND QUALITY CONTROL, Vol. 50 No. 2 (2022): April-June 2022, Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/183 ,
DOI: https://doi.org/10.46121/pspc.50.2.4
35. **Godavari Modalavalasa**, FEDERATED LEARNING FOR ENTERPRISE CLOUD DATA ENGINEERING: ARCHITECTURE, SECURITY, AND GOVERNANCE CHALLENGES, Vol. 51 No. 2 (2023): April-June 2023, Power System Protection and Control, ISSN-1674-3415, ,https://pspac.info/index.php/dlbh/article/view/184,
DOI: https://doi.org/10.46121/pspc.51.2.5
36. **Godavari Modalavalasa, AI**-DRIVEN DATA GOVERNANCE: INTELLIGENT METADATA, LINEAGE, AND COMPLIANCE AUTOMATION IN CLOUD DATA PLATFORMS, Archives / Vol. 52 No. 1 (2024): January-March 2024 /, Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/182,
DOI: https://doi.org/10.46121/pspc.52.1.3

37. **Hima Bindu Lekkala, VishnuVardhan Bandar**i , AUTONOMOUS WORKFLOW OPTIMIZATION USING MULTI AGENT AI SYSTEMS AI AGENTS MANAGE STATIONS, WIP, AND TASK HANDOFFS, Vol. 54 No. 2 (202): April-June 2026, Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/306 ,
DOI: https://doi.org/10.46121/pspc.54.2.08
38. **Lekkala, H. B., & Bandari, V. V. (2026). Autonomous workflow optimization using multi-agent AI systems:** AI agents manage stations, WIP, and task handoffs. *Power System Protection and Control*, 54(2), 95–101. http://pspac.info/index.php/dlbh/article/view/306
39. **Bandari, V. V., & Lekkala, H. B. (2024). Physics-informed reinforcement learning for real-time control of complex manufacturing processes**. *Power System Protection and Control*, 52(2), 164–170. https://pspac.info/index.php/dlbh/article/view/343
40. **Lekkala, H. B., & Bandari, V. V. (2025). AI-based energy-aware scheduling and process optimization** in engineer-to-order smart manufacturing systems. *Power System Protection and Control*, 53(1), 30–36. http://pspac.info/index.php/dlbh/article/view/341
41. **Bandari, V. V., & Lekkala, H. B. (2026). AI-driven digital** thread framework for end-to-end lifecycle optimization in ETO manufacturing systems. *Power System Protection and Control*, 54(2), 318–325. http://pspac.info/index.php/dlbh/article/view/340
42. **Lekkala, H. B., & Bandari, V. V. (2026). Deep learning** for weld defect detection using CNN or vision transformers fusion detection. *Power System Protection and Control*, 54(2), 151–158. https://pspac.info/index.php/dlbh/article/view/314
43. **Bandari, V. V., & Lekkala, H. B. (2024). AI-based operator** behavior monitoring and cost optimization using digital traceability in manufacturing systems. *Power System Protection and Control*, 52(1), 38–45. https://pspac.info/index.php/dlbh/article/view/342
44. **Shreyas Jayaprakash,** MACHINE LEARNING-DRIVEN VERIFICATION: USING ML TO OPTIMIZE COVERAGE ANALYSIS, PREDICT SIMULATION RUNTIMES, AND AUTO-GENERATE TESTBENCHES, Vol. 54 No. 2 (2026): April-June 2026, Power System Protection and Control, ISSN-1674-3415,
https://pspac.info/index.php/dlbh/article/view/391
45. **Mohammed Shafi Kundiladi,** EVENT-DRIVEN IMAGE AND VEHICLE STATUS MANAGEMENT FOR LOW-POWER IOT DIGITAL LICENSE PLATES, Vol. 53 No. 3 (2025): July-September 2025, Power System Protection and Control, ISSN-1674-3415,
https://pspac.info/index.php/dlbh/article/view/175
DOI: https://doi.org/10.46121/pspc.53.3.17
46. Controlled Synthesis and Characterization of Lanthanum Nanorods, **Author(s), Jayadeep Tejani, Rahul Shah, Hiral Vaghela, Shailesh Vajapara, Amanullakhan Pathan**, doi:10.18576/ijtfst/090205,
URL: https://www.naturalspublishing.com/Article.asp?ArtcID=20705, May-2020
47. **Kunal Kapoor**, **Apr 2026,** Relation Extraction and NER through Fine Tuned BERT model with transfer Learning, https://ieeexplore.ieee.org/document/11472214/metrics#metrics
48. Coalition Against Insurance Fraud. (2022). By the Numbers: The Challenge of Insurance Fraud in the United States. Coalition Against Insurance Fraud Research Division. https://insurancefraud.org/fraud-stats/
49. Insurance Europe. (2023). European Insurance – Key Facts 2023: Fraud Costs and Detection Benchmarks. Insurance Europe Publications.
50. European Parliament. (2016). General Data Protection Regulation (EU) 2016/679, Article 22: Automated Individual Decision-Making, Including Profiling. Official Journal of the European Union.

DOI: 10.46121/pspc.54.2.59

51. National Association of Insurance Commissioners. (2020). Artificial Intelligence Guidance Document: Principles for the Responsible Use of AI in Insurance. NAIC Publications.
52. International Association of Insurance Supervisors. (2023). Issues Paper on the Use of Artificial Intelligence in Insurance. IAIS Publications. https://www.iaisweb.org
53. Lundberg, S. M., & Lee, S. I. (2017). A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.
54. Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785
55. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why Should I Trust You? Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144.
56. 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, 321–357. https://doi.org/10.1613/jair.953
57. Shapley, L. S. (1953). A Value for n-Person Games. Contributions to the Theory of Games, 2(28), 307–317. Princeton University Press.
58. National Insurance Crime Bureau. (2023). NICB Fraud Detection and Insurance Loss Statistics: Annual Typology Report 2023. NICB Research.
59. European Insurance and Occupational Pensions Authority. (2021). Artificial Intelligence Governance Principles: Towards Ethical and Trustworthy AI in the European Insurance Sector. EIOPA Publications. https://doi.org/10.2854/795556

DOI: 10.46121/pspc.54.2.59