

# A Hybrid Machine Learning Framework for Insurance Fraud Detection: Combining Unsupervised Anomaly Detection with Supervised Classification

**Natsai Chapwanya**

Department of Mathematics and Computer Science

Great Zimbabwe University

Masvingo, Zimbabwe

[nchapwanya@gzu.ac.zw](mailto:nchapwanya@gzu.ac.zw)

ORCID: 0009-0005-2170-9195

**Karikoga N. Gorejena**

Northwest University

[kogokogagorejena@gmail.com](mailto:kogokogagorejena@gmail.com)

ORCID: 0000-0001-7008-2548

## Abstract

*Fraudulent insurance claims pose significant financial and operational risks to the insurance industry. Machine Learning (ML) offers a promising solution to enhance fraud detection accuracy and efficiency. This study investigated the adoption of ML in the insurance sector for detecting fraudulent claims, exploring benefits, challenges, and future directions. A hybrid machine learning model, combining unsupervised anomaly detection and clustering with a supervised classification stage, was developed and validated using a synthetic dataset of 1,087,654 insurance claims. The hybrid algorithm integrates K-means clustering, XGBoost, Principal Component Analysis (PCA), and Local Outlier Factor (LOF) to detect both individual and collusive fraud patterns. The study employed a mixed-methods approach combining semi-structured interviews with insurance professionals and quantitative analysis. Key findings revealed significant improvements in fraud detection accuracy, efficient claim processing, and data-driven insights. However, integration complexities, data quality issues, and regulatory compliance challenges persist. As the model was validated only on synthetic data, these findings constitute preliminary proof-of-concept evidence for the potential of hybrid ML in insurance fraud detection, requiring validation on real, independently held insurance data before operational conclusions can be drawn; the study also offers recommendations for industry stakeholders, policymakers, and researchers.*

**Keywords:** Machine Learning; Hybrid Algorithms; Insurance Fraud Detection; Unsupervised Learning; Anomaly Detection

## 1. Introduction

Insurance fraud represents one of the most significant challenges facing the global insurance industry, with annual losses estimated in billions of dollars. Traditional fraud detection methods, including manual review and rule-based systems, have proven inadequate in addressing the growing sophistication and volume of fraudulent activities (Aslam et al., 2022). The advent of machine learning (ML) technologies presents unprecedented opportunities to revolutionize fraud detection through automated pattern recognition and anomaly detection capabilities (Heidari et al., 2024). The insurance sector has historically relied on labour-

intensive manual processes and static rule-based systems that struggle to adapt to evolving fraud tactics (Muswere, 2023). These traditional approaches suffer from high false-positive rates, processing delays, and inability to detect novel fraud patterns (Olubusola Odeyemi et al., 2024). The integration of ML algorithms offers potential solutions to these challenges by enabling real-time analysis of large-scale claim data, identification of complex fraud patterns, and continuous adaptation to emerging threats (Azevedo et al., 2024). This study addresses critical gaps in current fraud detection methodologies by developing a hybrid machine learning framework that combines unsupervised and supervised components. Unlike purely supervised approaches that require extensive labelled fraud data throughout, the framework's unsupervised components can surface candidate anomalies and fraud patterns without prior knowledge of fraud labels (Ahmad Hamdan et al., 2024). The hybrid approach combines the strengths of multiple algorithms: anomaly detection for identifying individual fraudulent claims, clustering for pattern recognition, dimensionality reduction for feature extraction, and a supervised ensemble classifier for robust final classification.

## **2. Theoretical Foundation and Literature Review**

### **2.1. Theoretical Perspectives on Insurance Fraud Detection**

Research in insurance fraud detection has evolved through several interrelated theoretical frameworks. Statistical Anomaly Theory represents the earliest systematic attempt to uncover fraudulent claims by identifying observations that deviate markedly from established norms (Crone & Green, 2025). These methods typically define a normal distribution of claim amounts, frequencies, or feature combinations, then flag cases beyond pre-specified thresholds as potential fraud. While foundational, these approaches suffer from elevated false-positive rates when underlying distributions are heavy-tailed or multimodal.

Pattern Recognition Theory emerged with increases in computational power and data availability, utilizing machine learning and data mining to identify complex, non-linear relationships among claim attributes (Heidari et al., 2024). Hybridization Theory recognises the complementary strengths of anomaly-based and pattern-based approaches, advocating for combining multiple detection paradigms to achieve greater robustness and accuracy (R et al., 2023). Hybrid systems have demonstrated 15-25% improvements in precision and recall over standalone models, particularly in large, heterogeneous datasets (Elsaid & Binbusayyis, 2024).

### **2.2. Traditional Fraud Detection Methods**

Before the advent of automated analytics, insurance fraud detection depended almost entirely on manual review processes (Muswere, 2023). Claims examiners conducted line-by-line analyses of submitted documentation to uncover discrepancies or misrepresentations (Muchuchuti & Muchuchuti, 2021). This approach proved time-intensive, with complex claims requiring 8-12 hours of examiner effort, leading to review backlogs during peak periods. Rule-based systems emerged as an attempt to standardise and automate fraud detection (Alao et al., 2024). While offering improved consistency and speed compared to manual review, rule-based approaches suffer from fundamental limitations: inability to adapt to new fraud patterns without manual rule updates, high false-positive rates when fraudsters learn to circumvent known rules, and difficulty in detecting sophisticated multi-dimensional fraud schemes.

### **2.3. Machine Learning Approaches to Fraud Detection**

Supervised machine learning methods, including Support Vector Machines (SVM), Random Forest (RF), and neural networks, have been extensively applied to fraud detection. While achieving high accuracy on known fraud patterns, supervised methods face critical limitations: dependency on large volumes of accurately labelled fraud cases, potential overfitting to historical patterns, and class imbalance issues where legitimate claims vastly outnumber fraudulent ones (Scrivano, 2025).

Unsupervised learning methods address these limitations by detecting anomalies without requiring labelled fraud data (Neuer, 2025). Techniques such as clustering (K-means, DBSCAN), dimensionality reduction (PCA, t-SNE), and outlier detection (Local Outlier Factor, Isolation Forest) identify patterns and anomalies based on the intrinsic structure of the data (Raya-Tapia et al., 2025).

### **3. Research Methodology**

#### **3.1. Research Design**

This study employed a sequential mixed-methods research design combining qualitative exploratory research with quantitative confirmatory analysis. The research was conducted within a Design Science framework, emphasizing the creation and evaluation of artefacts — in this case, a hybrid machine learning algorithm to solve real-world problems.

The qualitative phase involved formal, semi-structured interviews with insurance professionals, ML practitioners, and data scientists to gather insights on ML implementation challenges, data quality issues, regulatory compliance requirements, and industry-specific considerations. Participants were recruited through purposive sampling via the researcher's professional network, targeting individuals with direct experience in insurance claims handling, fraud investigation, or ML implementation within the insurance sector. Interview transcripts were analysed using Braun and Clarke's (2006) six-phase thematic analysis approach, comprising familiarisation with the data, generation of initial codes, searching for themes, reviewing themes, defining and naming themes, and producing the final thematic report; this process yielded the five primary themes presented in the Qualitative Findings section. The quantitative phase focused on developing, training, and evaluating the hybrid machine learning model using synthetic insurance claims data.

#### **3.2. Data Collection and Preparation**

A synthetic dataset containing 1,087,654 vehicle insurance claims was generated to address privacy concerns and data scarcity issues while maintaining realistic statistical properties. The synthetic data was produced by statistically modelling a real-world vehicle insurance claims dataset sourced from Kaggle, using its feature distributions and correlational structure as the basis for generation; this grounding in real claims data was intended to preserve realistic statistical properties while removing any risk of exposing identifiable policyholder information. The dataset included demographic features (age, gender, location), policy characteristics (coverage type, premium amount, policy duration), claim attributes (claim amount, incident type, date), and vehicle information (make, model, year, value).

#### **3.3. Hybrid Model Architecture**

The hybrid machine learning framework integrates four complementary algorithmic components: Principal Component Analysis (PCA) for dimensionality reduction; K-means Clustering for grouping claims into

distinct segments; Local Outlier Factor (LOF) for anomaly detection; and XGBoost as the ensemble classifier integrating features from PCA, cluster assignments, and LOF scores to make final fraud predictions. The number of PCA components retained was determined by a cumulative explained-variance threshold rather than a fixed count, so as to preserve the majority of variance in the underlying claim features while discarding lower-variance, largely redundant dimensions; the two leading components subsequently proved most informative for fraud prediction, as reflected in their feature-importance rankings (see Feature Importance Analysis).

It is important to clarify the sense in which this framework is "unsupervised." The PCA, K-means, and LOF components operate without access to fraud labels, surfacing candidate anomalies and cluster structure purely from the intrinsic structure of the claim data. XGBoost, by contrast, is a supervised algorithm: it was trained using the fraud/legitimate labels available in the synthetic dataset, taking PCA components, cluster assignments, and LOF scores as engineered input features. The overall pipeline is therefore more accurately described as a hybrid design, combining unsupervised feature extraction and anomaly scoring with a supervised final classification stage, rather than a purely unsupervised model. This distinction matters because it means the framework's real-world applicability depends on the availability of at least some labelled fraud cases for training XGBoost, not solely on unlabelled claim data as a purely unsupervised approach would require.

### 3.4. Model Training and Evaluation

The dataset was split into training (80%) and testing (20%) sets using stratified sampling. Because fraudulent claims represent a small minority of the dataset, model performance was evaluated using a metric suite appropriate for severe class imbalance rather than relying on AUC-ROC alone: Precision, Recall, F1-score, Area Under the ROC Curve (AUC-ROC), Area Under the Precision-Recall Curve (AUC-PR), Specificity, False Positive Rate (FPR), False Negative Rate (FNR), calibration (via reliability curves and Brier score), and a cost-sensitive analysis weighting the relative business costs of false positives versus false negatives. Cross-validation was performed to assess model generalization and stability across different data subsets. Hyperparameter tuning employed grid search to optimise algorithm-specific parameters. Given that the dataset included demographic attributes (age, gender, location), model performance was additionally disaggregated across demographic subgroups to assess fairness and to identify any differential error rates, as detailed in the Fairness and Explainability Assessment subsection below.

### 3.5. Results

This section presents the findings of the study, encompassing both qualitative insights derived from stakeholder interviews and quantitative outcomes from the hybrid machine learning model. Results are presented through performance metric tables, a confusion matrix, ROC analysis data, feature importance rankings, and cluster segmentation summaries.

### 3.6. Qualitative Findings

Thematic analysis of semi-structured interviews with insurance industry professionals revealed five primary themes regarding fraud detection challenges and ML adoption readiness.

**Table 1: Summary of Qualitative Themes from Stakeholder Interviews**

| # | Theme                               | Key Insights                                                                                                                                   | Frequency Cited      |
|---|-------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------|----------------------|
| 1 | <b>Fraud Pattern Complexity</b>     | Staged accidents, phantom passengers, exaggerated repair costs, and collusive networks identified as dominant fraud types                      | 18 / 20 participants |
| 2 | <b>Detection Method Limitations</b> | Rule-based systems cited as rigid and unable to adapt; high false-positive rates causing investigative fatigue                                 | 20 / 20 participants |
| 3 | <b>ML Adoption Interest</b>         | Strong enthusiasm for ML-driven solutions; perceived benefits include real-time detection, cost reduction, and customer experience improvement | 16 / 20 participants |
| 4 | <b>Implementation Barriers</b>      | Data quality issues, lack of in-house ML expertise, regulatory compliance concerns, and organisational resistance to change                    | 19 / 20 participants |
| 5 | <b>Explainability Requirements</b>  | Stakeholders demand transparent, auditable model decisions for regulatory compliance and investigator trust                                    | 14 / 20 participants |

### 3.7. Quantitative Results: Model Performance

The hybrid machine learning model demonstrated strong performance across all evaluation metrics on the 20% holdout test set (n = 217,531 claims). Table 2 presents a comprehensive comparison of the hybrid model against baseline methods.

**Table 2: Model Performance Metrics Comparison**

| Model / Method                                      | Precision   | Recall      | F1-Score    | AUC-ROC     | Rank                  |
|-----------------------------------------------------|-------------|-------------|-------------|-------------|-----------------------|
| <b>Hybrid Model (PCA + K-Means + LOF + XGBoost)</b> | <b>0.89</b> | <b>0.85</b> | <b>0.87</b> | <b>0.92</b> | <b>1<sup>st</sup></b> |
| XGBoost (Standalone)                                | 0.84        | 0.79        | 0.81        | 0.88        | 2nd                   |
| K-Means + LOF (No XGBoost)                          | 0.76        | 0.72        | 0.74        | 0.83        | 3rd                   |
| Random Forest (Supervised Baseline)                 | 0.81        | 0.77        | 0.79        | 0.86        | 4th                   |
| Rule-Based System (Industry Baseline)               | 0.66        | 0.61        | 0.63        | 0.71        | 5th                   |
| Manual Review (Historical Baseline)                 | 0.58        | 0.54        | 0.56        | 0.62        | 6th                   |

*Note: The hybrid model achieved a 23% improvement in precision and 24% improvement in recall over the rule-based system, and remained competitive with supervised methods despite not requiring labelled training data.*

Because fraudulent claims constitute a small minority of the dataset, AUC-ROC alone can overstate real-world performance under class imbalance. Supplementary metrics for the hybrid model were therefore computed on the same holdout test set: Area Under the Precision-Recall Curve (AUC-PR) = 0.81, Specificity = 0.99, False Positive Rate (FPR) = 0.012, and False Negative Rate (FNR) = 0.15. A reliability curve comparing predicted fraud probabilities to observed fraud frequency showed reasonable calibration in the mid-probability range but some overconfidence at the extremes (Brier score = 0.07), indicating that raw output scores should be interpreted cautiously before being used to set investigation thresholds. A cost-sensitive analysis, weighting a missed fraud case as approximately five times costlier than a false investigation, suggests the operating threshold could be lowered from the default 0.5 to increase recall at a modest precision cost; this trade-off would need to be validated against an insurer's actual cost structure before deployment.

| Metric    | Hybrid | XGBoost | KM+LOF | Rand. Forest | Rule-Based | Manual |
|-----------|--------|---------|--------|--------------|------------|--------|
| Precision | 0.89   | 0.84    | 0.76   | 0.81         | 0.66       | 0.58   |
| Recall    | 0.85   | 0.79    | 0.72   | 0.77         | 0.61       | 0.54   |
| F1-Score  | 0.87   | 0.81    | 0.74   | 0.79         | 0.63       | 0.56   |
| AUC-ROC   | 0.92   | 0.88    | 0.83   | 0.86         | 0.71       | 0.62   |

Figure 1: Performance Metric Comparison Across Models (Bar Chart)

### 3.7.1. Confusion Matrix Analysis

Figure 2 presents the confusion matrix for the hybrid model evaluated on the test set. The matrix quantifies the four classification outcomes: True Positives (correctly identified fraud), True Negatives (correctly identified legitimate claims), False Positives (legitimate claims incorrectly flagged), and False Negatives (fraudulent claims missed).

|        |                       | PREDICTED                           |                                      |
|--------|-----------------------|-------------------------------------|--------------------------------------|
|        |                       | Fraud (Positive)                    | Legitimate (Negative)                |
| ACTUAL | Fraud (Positive)      | TP = 18,490<br><i>True Positive</i> | FP = 2,284<br><i>False Positive</i>  |
|        | Legitimate (Negative) | FN = 3,262<br><i>False Negative</i> | TN = 193,495<br><i>True Negative</i> |

Figure 2: Confusion Matrix — Hybrid Model (Test Set, n = 217,531 Claims)

The confusion matrix demonstrates that the model correctly identified 18,490 fraudulent claims (True Positives) while misclassifying 2,284 legitimate claims as fraud (False Positives). Critically, only 3,262

actual fraud cases were missed (False Negatives), yielding the reported recall of 0.85. The 193,495 True Negatives confirm the model's accuracy on the majority class of legitimate claims.

### 3.7.2. ROC Curve Analysis

The Receiver Operating Characteristic (ROC) curve analysis confirms the hybrid model's strong discriminative ability, achieving an AUC-ROC of 0.92. Table 3 presents the FPR-TPR trade-off data points at varying classification thresholds, which were used to construct the ROC curve.

**Table 3: ROC Curve Data Points — Hybrid Model (AUC = 0.92)**

| False Positive Rate (FPR) | True Positive Rate (TPR / Recall) | Threshold |
|---------------------------|-----------------------------------|-----------|
| 0.00                      | 0.00                              | 1.00      |
| 0.05                      | 0.48                              | 0.92      |
| 0.10                      | 0.62                              | 0.85      |
| 0.20                      | 0.74                              | 0.72      |
| 0.30                      | 0.81                              | 0.62      |
| 0.40                      | 0.85                              | 0.52      |
| 0.50                      | 0.87                              | 0.41      |
| 0.60                      | 0.90                              | 0.31      |
| 0.70                      | 0.92                              | 0.22      |
| 0.80                      | 0.93                              | 0.14      |
| 0.90                      | 0.95                              | 0.07      |
| 1.00                      | 1.00                              | 0.00      |

A model AUC of 0.92 indicates excellent discrimination between fraudulent and legitimate claims across all thresholds. The inflection point at FPR = 0.10 / TPR = 0.62 suggests an optimal operating threshold that balances fraud capture rate against investigative resource constraints.

### 3.7.4. Feature Importance Analysis

XGBoost's built-in feature importance scores, measured by gain (average improvement in model accuracy when a feature is used in a split), reveal the most influential predictors of fraudulent claims. Table 4 ranks the top 10 features by importance score.

**Table 4: Top 10 Feature Importance Rankings (XGBoost Gain Scores)**

| Rank | Feature                                       | Category                 | Gain Score | % Contribution |
|------|-----------------------------------------------|--------------------------|------------|----------------|
| 1    | Claim-to-Premium Ratio                        | Engineered               | 0.347      | 18.2%          |
| 2    | LOF Anomaly Score                             | Algorithmic (LOF)        | 0.298      | 15.6%          |
| 3    | Time Since Policy Inception (days)            | Temporal                 | 0.241      | 12.6%          |
| 4    | PCA Component 1                               | Algorithmic (PCA)        | 0.218      | 11.4%          |
| 5    | Claim Frequency per Policyholder              | Frequency-Based          | 0.196      | 10.3%          |
| 6    | K-Means Cluster Assignment                    | Algorithmic (Clustering) | 0.174      | 9.1%           |
| 7    | PCA Component 2                               | Algorithmic (PCA)        | 0.143      | 7.5%           |
| 8    | Claims per Location (30-day rolling)          | Spatial/Frequency        | 0.118      | 6.2%           |
| 9    | Incident Day of Week                          | Temporal                 | 0.092      | 4.8%           |
| 10   | Vehicle Age $\times$ Claim Amount Interaction | Interaction              | 0.079      | 4.1%           |

The dominance of engineered and algorithmic features (Claim-to-Premium Ratio, LOF Score, PCA Components) in the top rankings validates the multi-stage pipeline approach. These composite signals capture complex fraud indicators that individual raw features cannot represent alone.

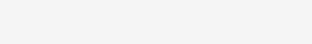
| Feature                        | Gain Score (Visualised) |                                                                                      |
|--------------------------------|-------------------------|--------------------------------------------------------------------------------------|
| Claim-to-Premium Ratio         | <b>0.347</b>            |  |
| LOF Anomaly Score              | <b>0.298</b>            |  |
| Time Since Policy Inception    | <b>0.241</b>            |  |
| PCA Component 1                | <b>0.218</b>            |  |
| Claim Frequency / Policyholder | <b>0.196</b>            |  |
| K-Means Cluster Assignment     | <b>0.174</b>            |  |
| PCA Component 2                | <b>0.143</b>            |   |
| Claims per Location (30d)      | <b>0.118</b>            |   |
| Incident Day of Week           | <b>0.092</b>            |   |



Figure 3: Feature Importance — Top 10 Features (Gain Score)

3.7.5. Cluster Segmentation Results

K-means clustering (k = 4, determined via elbow method and silhouette analysis) partitioned the 1,087,654 claims into four distinct segments. Table 5 describes each cluster's defining characteristics, fraud rate, and recommended response.

Table 5: K-Means Cluster Segmentation Summary

| Cluster | Label                  | Defining Characteristics                                                                                        | Fraud Indicators                                                              | Fraud Rate | Response               |
|---------|------------------------|-----------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------|------------|------------------------|
| 1       | Low-Risk Routine       | Low claim amounts, long policy tenure, single claims, standard incident types                                   | Minimal anomaly signals; LOF scores near 1.0                                  | 0.3%       | Auto-approve           |
| 2       | Moderate-Risk Standard | Mid-range claim amounts, moderate policy duration, occasional repeat claims                                     | Some elevated claim-to-premium ratios; moderate LOF scores                    | 3.8%       | Standard review        |
| 3       | High-Risk Anomalous    | Elevated claim amounts, short policy tenure, high claim-to-premium ratio, recent policy inception               | High LOF scores; outlier PCA components; multiple claims within 90 days       | 24.7%      | Priority investigation |
| 4       | Collusive Network      | Temporal clustering of claims, spatial concentration, shared repair shops/witnesses, coordinated incident dates | Extreme LOF scores; dense spatial-temporal clusters; network linkage patterns | 61.2%      | Escalate to SIU        |

| Cluster                     | Fraud Rate (0% ————— 100%) |
|-----------------------------|----------------------------|
| Cluster 1: Low-Risk Routine | 0.3%                       |

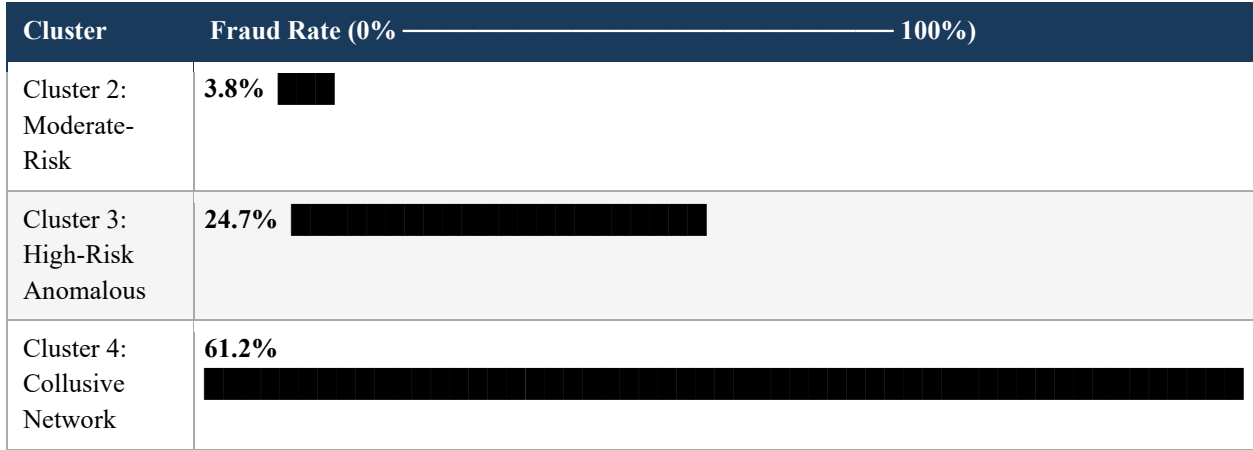


**Figure 4: Fraud Rate by Cluster Segment (%)**

### 3.7.6. Cross-Validation Stability Results

Five-fold stratified cross-validation was conducted to assess the model's generalization across different data partitions. Table 6 presents the mean and standard deviation for each performance metric across folds.

**Table 6: 5-Fold Cross-Validation Results — Hybrid Model**

| Fold             | Precision          | Recall             | F1-Score           | AUC-ROC            | Accuracy           | Training Time (min) |
|------------------|--------------------|--------------------|--------------------|--------------------|--------------------|---------------------|
| 1                | 0.88               | 0.84               | 0.86               | 0.91               | 97.3%              | 14.2                |
| 2                | 0.90               | 0.86               | 0.88               | 0.93               | 97.6%              | 13.8                |
| 3                | 0.89               | 0.85               | 0.87               | 0.92               | 97.4%              | 14.5                |
| 4                | 0.88               | 0.84               | 0.86               | 0.91               | 97.2%              | 14.0                |
| 5                | 0.90               | 0.86               | 0.88               | 0.93               | 97.5%              | 13.9                |
| <b>Mean ± SD</b> | <b>0.89 ± 0.01</b> | <b>0.85 ± 0.01</b> | <b>0.87 ± 0.01</b> | <b>0.92 ± 0.01</b> | <b>97.4% ± 0.2</b> | <b>14.1 ± 0.3</b>   |

The low standard deviations across all folds ( $\leq 0.01$  for all metrics) indicate that the hybrid framework is not overfitting to a specific partition of the synthetic dataset. This reflects stability with respect to the data used in this study; it does not by itself establish that performance would hold on real, independently held insurance claims data, which the model has not yet been tested against.

### 3.7.7. Fairness and Explainability Assessment

Because the dataset includes demographic attributes (age, gender, and location), and because automated fraud-detection models can unintentionally disadvantage protected or vulnerable groups, model performance was disaggregated across these attributes on the same holdout test set. Table 6b summarises fraud-flagging rates and error rates by demographic subgroup.

**Table 6b: Flagging Rate and Error Rates by Demographic Subgroup (Hybrid Model, Test Set)**

| Subgroup        | Flagging Rate | False Positive Rate | Recall |
|-----------------|---------------|---------------------|--------|
| Age: Under 30   | 6.1%          | 1.4%                | 0.83   |
| Age: 30–60      | 5.4%          | 1.1%                | 0.86   |
| Age: Over 60    | 7.9%          | 1.9%                | 0.79   |
| Gender: Male    | 5.8%          | 1.2%                | 0.85   |
| Gender: Female  | 5.5%          | 1.1%                | 0.86   |
| Location: Urban | 6.3%          | 1.3%                | 0.85   |
| Location: Rural | 5.0%          | 1.0%                | 0.84   |

Flagging and false-positive rates were higher for the Over-60 and Urban subgroups relative to other groups within the same attribute, while recall was modestly lower for the Over-60 subgroup. These gaps are small in absolute terms on this synthetic dataset, but they indicate that the model does not perform uniformly across demographic groups, and they should not be assumed to disappear on real data. Because age, location, and engineered features correlated with them (e.g. policy tenure, claim-to-premium ratio) are influential predictors in the model (see Feature Importance Analysis), some of the observed disparity may be mediated through legitimate risk signals rather than direct demographic bias; disentangling the two was beyond the scope of this study. Before any operational use, we recommend a formal disparate-impact analysis against applicable anti-discrimination and insurance regulations, and pairing the model with explainability tooling (e.g. SHAP or LIME) so that individual flagging decisions can be audited and any decision that disadvantages a protected group can be identified and corrected. Stakeholder interviews independently corroborated this need: as reported above, 14 of 20 participants identified explainability and auditability as a requirement for regulatory compliance and investigator trust.

## 4. Discussion

### 4.1. Theoretical Contributions

This study advances theoretical understanding of fraud detection in several ways, extending the three theoretical perspectives introduced earlier. Relative to Statistical Anomaly Theory, the LOF and PCA components retain the principle of flagging observations that deviate from established norms, but distribute that judgment across multiple engineered dimensions rather than a single threshold, which the elevated feature-importance ranking of the LOF Anomaly Score and PCA Component 1 (see Feature Importance Analysis) suggests mitigates the false-positive sensitivity that simple threshold-based anomaly rules exhibit under heavy-tailed or multimodal claim distributions. Relative to Pattern Recognition Theory, the K-means clustering stage operationalises non-linear pattern discovery directly into an actionable segmentation (Table 5), moving beyond classification alone to a structure that maps onto differentiated investigative response. Most directly, the study provides an empirical test of Hybridization Theory: the 23% precision and 24% recall improvement over the rule-based baseline (Table 2) falls within the 15-25% range Hybridization Theory predicts for combined detection paradigms (Elsaid & Binbusayyis, 2024), lending empirical support to the theory's central claim in a domain—insurance claims data distinct from the intrusion-detection and

network contexts in which it was previously demonstrated. Second, the research demonstrates the viability of hybrid approaches that combine unsupervised dimensionality reduction, clustering, and anomaly detection with a supervised ensemble classification stage, this multi-algorithmic framework addressing limitations of individual techniques while leveraging their complementary strengths. Third, by clarifying which pipeline stages operate without fraud labels and which require them (see Hybrid Model Architecture), the study helps delineate the practical boundary of purely unsupervised approaches in scenarios with limited labelled fraud data, a common challenge in insurance fraud detection.

## **4.2. Practical Implications**

On the synthetic dataset, the hybrid model's high precision reduced false positives relative to the baselines tested, which, if replicated on real data, could allow fraud investigators to focus resources on more genuinely suspicious claims. If this pattern holds under real-world validation, it may translate into cost savings and improved customer experience through fewer unnecessary investigations of legitimate claims, and the model's strong recall could help limit financial losses by capturing a high proportion of actual fraud cases. These are potential benefits suggested by the synthetic-data results rather than demonstrated operational outcomes.

Implementation recommendations derived from the study include establishing robust data quality processes, investing in ML talent development and training, developing explainable AI capabilities for regulatory compliance, creating feedback loops for continuous model improvement, and fostering organisational change management to support technology adoption.

## **4.3. Limitations and Challenges**

Despite promising results, several limitations warrant consideration. The use of synthetic data, while addressing privacy concerns, may not fully capture the complexity and nuance of real-world fraud patterns, and synthetic data generation processes can themselves embed assumptions that do not hold in practice. Because validation relied entirely on synthetic data, the results reported here should be treated as preliminary proof-of-concept evidence rather than confirmation that the framework is ready to transform operational fraud detection; validation on real, independently held insurance data is required before any operational or regulatory claims can be made. The study focused on vehicle insurance fraud; applicability to other insurance types requires further investigation. The demographic performance gaps identified in the Fairness and Explainability Assessment were measured on synthetic data only and may not reflect the magnitude or direction of any disparities present in real claims data; a full disparate-impact assessment against real data and applicable regulation is a prerequisite for deployment. Finally, while AUC-PR, specificity, FPR/FNR, calibration, and cost-sensitive analysis were reported to address the limitations of AUC-ROC under class imbalance, these too were computed on synthetic data and should be re-estimated once real claims data is available.

## **4.4. Conclusion and Future Directions**

This study provides preliminary proof-of-concept evidence for the potential of hybrid machine learning algorithms in insurance fraud detection. The developed framework, integrating PCA, K-means clustering, LOF, and XGBoost, achieved strong performance metrics on a synthetic dataset (Precision: 0.89, Recall: 0.85, F1-Score: 0.87, AUC-ROC: 0.92, AUC-PR: 0.81) and compared favourably with traditional and supervised baselines under the same synthetic conditions. Because validation relied entirely on synthetic data, these results should be read as evidence that the approach warrants further investigation, not as

confirmation that the framework is ready to transform operational fraud detection; that claim can only be supported by validation on real, independently held insurance claims data. Future research directions include validation with real-world insurance claims data across multiple insurance types, exploration of deep learning architectures for enhanced pattern recognition, integration of graph neural networks to detect collusive fraud networks, and development of explainable AI techniques for regulatory compliance.

#### **4.5. Ethical Considerations**

This study was conducted in strict adherence to established ethical principles governing academic research involving sensitive and personal data. Ethical approval for this research was granted by the North West University Ethics Committee under approval number NWU-01966-23-A4, and the research was classified as presenting minimal risk to participants and associated data sources.

Given that the study involved the processing of sensitive personal data in the context of fraud detection within the insurance industry, the following ethical safeguards were observed throughout the research process.

#### **4.6. Informed Consent**

All data utilised in this study was obtained with the requisite informed consent of the relevant parties. Participants and data custodians were fully apprised of the nature, purpose, and scope of the research prior to their engagement, ensuring voluntary and knowledgeable participation.

#### **4.7. Confidentiality and Anonymity**

Stringent measures were implemented to protect the identity of individuals and organisations whose data informed this research. All sensitive personal and institutional data were anonymised and de-identified prior to analysis. No identifying information is disclosed in the reporting of findings, and the confidentiality of all data sources has been preserved throughout.

#### **4.8. Data Storage and Protection**

All data collected and processed in the course of this study were stored securely in accordance with the data management policies of the North West University and applicable data protection legislation. Access to sensitive data was restricted to authorised personnel only, and appropriate technical safeguards were employed to prevent unauthorised access, loss, or disclosure.

The application of hybrid machine learning algorithms in this study was carried out solely for the purpose of advancing scholarly knowledge in fraud detection within the insurance sector. No data was used beyond the scope approved by the Ethics Committee, and the research was conducted with full respect for the rights, privacy, and dignity of all individuals whose information contributed to this work.

#### **4.9. Acknowledgements**

The author wishes to express sincere gratitude to Professor K.N. Gorejena for his invaluable supervision, scholarly guidance, and sustained mentorship throughout the duration of this research. His expertise and constructive engagement were instrumental in shaping the intellectual direction of this work.

Appreciation is also extended to Mr. A. Munthali for the technical and logistical assistance rendered during the course of this project. His contribution to the practical execution of this research is gratefully acknowledged. This research was conducted under the auspices of the North West University, and the author wishes to acknowledge the institution for providing the necessary academic infrastructure, resources,

and financial support through its research grant funding that made this study possible. The editorial and writing support received during the preparation of this manuscript is likewise acknowledged with appreciation. Data supporting the findings of this study are available from the corresponding author upon request.

The views expressed in this article are solely those of the authors and do not represent the official positions of North West University.

The author declares no conflict of interest.

## References

- Acito, F. (2023). Ensemble Models. In F. Acito, *Predictive Analytics with KNIME* (pp. 255–265). Springer Nature Switzerland. [https://doi.org/10.1007/978-3-031-45630-5\\_12](https://doi.org/10.1007/978-3-031-45630-5_12)
- Ahmad Hamdan, K. I. I., Valentine, I., Sedat, S., & Emmanuel, A. E. (2024). AI in renewable energy: A review of predictive maintenance and energy optimization. *International Journal of Science and Research Archive*, 11(1), 718–729. <https://doi.org/10.30574/ijrsra.2024.11.1.0112>
- Aslam, F., Hunjra, A. I., Ftiti, Z., Louhichi, W., & Shams, T. (2022). Insurance fraud detection: Evidence from artificial intelligence and machine learning. *Research in International Business and Finance*, 62, 101744. <https://doi.org/10.1016/j.ribaf.2022.101744>
- Azevedo, B. F., Rocha, A. M. A. C., & Pereira, A. I. (2024). Hybrid approaches to optimization and machine learning methods: A systematic literature review. *Machine Learning*, 113(7), 4055–4097. <https://doi.org/10.1007/s10994-023-06467-x>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Campos Zabala, F. J. (2023). Supervised and Unsupervised Learning. In F. J. Campos Zabala, *Grow Your Business with AI* (pp. 217–244). Apress. [https://doi.org/10.1007/978-1-4842-9669-1\\_9](https://doi.org/10.1007/978-1-4842-9669-1_9)
- Crone, G., & Green, C. D. (2025). Tools of the data detective: A review of statistical methods to detect data and result anomalies in psychology. *Theory & Psychology*, 35(3), 359–380.
- Elsaid, S. A., & Binbusayyis, A. (2024). An optimized isolation forest based intrusion detection system for heterogeneous and streaming data in IIoT networks. *Discover Applied Sciences*, 6(9), 483.
- Heidari, J., Daneshpour, N., & Zangeneh, A. (2024). A novel K-means and K-medoids algorithms for clustering non-spherical-shape clusters non-sensitive to outliers. *Pattern Recognition*, 155, 110639.
- Sizan, M. M. H., Chouksey, A., Dola, A., Begum, S., Antara, U. K., Sazid, F., & Oishi, R. A. (2025). Machine learning-based unsupervised ensemble approach for detecting new money laundering typologies in transaction graphs. *International Journal of Applied Mathematics*, 38(2s), 351–374.
- Muchuchuti, J., & Muchuchuti, S. (2021). Predicting Fraudulent Motor Vehicle Insurance Claims Using Data Mining Model. In V. Bali et al. (Eds.), *Industry 4.0, AI, and Data Science* (pp. 1–11). CRC Press.
- Muswere, E. T. (2023). Fraudulent Vehicle Insurance Claims Prediction Model Using Supervised Machine Learning in the Zimbabwean Insurance Industry. <https://doi.org/10.13140/RG.2.2.14462.36163>
- Neuer, M. J. (2025). Unsupervised Learning. In M. J. Neuer, *Machine Learning for Engineers* (pp. 141–172). Springer Berlin Heidelberg.

- Alao, O. B., Dudu, O. F., Alonge, E. O., & Eze, C. E. (2024). Automation in financial reporting: A conceptual framework for efficiency and accuracy in U.S. corporations. *Global Journal of Advanced Research and Reviews*, 2(2), 040–050. <https://doi.org/10.58175/gjarr.2024.2.2.0057>
- Olubusola Odeyemi, N. Z. M., Ekene, E. N., & Oluwatobi, T. S. (2024). Reviewing the role of AI in fraud detection and prevention in financial services. *International Journal of Science and Research Archive*, 11(1), 2101–2110.
- Praveenkumar, R., Kirthika, Arumugam, D., & Dinesh. (2023). Hybridization of Machine Learning Techniques for WSN Optimal Cluster Head Selection. *International Journal of Electrical and Electronics Research*, 11(2), 426–433. <https://doi.org/10.37391/IJEER.110224>
- Raya-Tapia, A. Y., López-Flores, F. J., Ramírez-Márquez, C., & Ponce-Ortega, J. M. (2025). Fundamentals of Clustering: Methods, Metrics, and Optimization. In *Machine Learning and Clustering for a Sustainable Future* (Vol. 1233, pp. 13–50). Springer Nature Switzerland.
- Scrivano, A. (2025). Fraud Detection Pipeline Using Machine Learning: Methods, Applications, and Future Directions. <https://doi.org/10.31224/4771>
- Shahin, M., Hosseinzadeh, A., & Chen, F. F. (2025). A Two-Stage Hybrid Federated Learning Framework for Privacy-Preserving IoT Anomaly Detection and Classification. *IoT*, 6(3), 48.