

Implementing AI-Driven Decision Support Systems in Data Management

Shabir Ahmad¹, , Yazeed Alshuaibany^{1,*}, , 

¹College of Business, Al Yamamah University, Al Khobar, Saudi Arabia

Article History

Received: 03 October, 2025

Revised: 06 December, 2025

Accepted: 09 February, 2026

Published: 04 May, 2026

Abstract:

Introduction: The rapid growth of digital financial transactions has increased exposure to fraud and highlighted the limitations of traditional decision support systems (DSS). Rule-based and scorecard-driven approaches struggle to operate effectively under conditions of extreme class imbalance, evolving fraud patterns, and regulatory requirements for transparency and auditability. This study aimed to design and evaluate an AI-driven DSS for credit card fraud detection that integrated machine learning, class-imbalance mitigation, explainability, and cost-sensitive decision logic to support operational fraud analysis.

Method: The study used the publicly available Kaggle Credit Card Fraud dataset comprising 284,807 transactions, of which 0.17% were fraudulent. A stratified 80/20 split yielded a hold-out test set of 4,000 transactions, including 7 fraud cases. Class imbalance was addressed using the Synthetic Minority Oversampling Technique (SMOTE) applied exclusively within training folds. Logistic Regression, Random Forest, and XGBoost models were trained and evaluated. Given the extreme imbalance, Precision–Recall Area Under the Curve (PR–AUC) was treated as the primary evaluation metric, with recall, precision, F1-score, and ROC–AUC reported as complementary measures. Performance uncertainty was quantified using bootstrap confidence intervals. Probability calibration was applied to XGBoost outputs, and a cost-sensitive threshold optimisation framework was used to define operational risk tiers. SHapley Additive exPlanations (SHAP) were incorporated to support model transparency.

Results: XGBoost achieved the highest PR–AUC (0.568) compared with Random Forest (0.512) and Logistic Regression (0.341), while exhibiting comparable recall (0.571) to Random Forest on the imbalanced test set. Bootstrap confidence intervals indicated substantial uncertainty due to the limited number of fraud cases, underscoring the need for cautious interpretation of point estimates. Calibration improved probability reliability, and cost-sensitive thresholding produced consistent low-, medium-, and high-risk classifications. SHAP analysis provided global and local explanations of model behaviour, although the use of anonymised PCA-derived features constrained interpretability.

Conclusion: The findings demonstrated that integrating SMOTE, XGBoost, calibration, and SHAP within an AI-driven DSS supported improved fraud prioritisation under severe class imbalance while maintaining transparency and operational feasibility. Rather than replacing existing systems, the proposed framework illustrated how AI models can augment DSS by enabling calibrated risk scoring, uncertainty-aware evaluation, and analyst-oriented triage in regulated financial environments.

Keywords: AI-Driven DSS, credit card fraud detection, SMOTE, XGBoost, SHAP, data management, machine learning, explainable AI, risk scoring, imbalanced classification, calibration, PR-AUC, MLOps, cost-sensitive learning.

1. INTRODUCTION

Management Information Systems (MIS) have become an essential part of large organisations, enabling them to collect, integrate, and analyse data systematically to make strategic decisions and remain competitive in an increasingly globalised

environment [1]. As operational environments become increasingly interconnected, dynamic, and uncertain, organisations are in greater need of evidence-based management practices to ensure their resilience and responsiveness [2]. In this regard, the recent rise of Decision Support Systems (DSS) as critical socio-technical systems has

* Address correspondence to this author at College of Business, Al Yamamah University, Al Khobar, Saudi Arabia;
E-mail: y_alsuhaibany@yu.edu.sa



© 2026 Copyright by the Authors.

Licensed as an open access article using a [CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/).

been proposed to integrate data, analytical models, and computing applications to support human decision-makers in making complex, high-stakes decisions [3].

Traditional DSS methodologies, usually based on deterministic rules, predefined thresholds, or expert-defined scorecards, remain effective in environments where data is stable, well-structured, and operational patterns are predictable. This type of system enables analysis and managers to simulate scenarios, test assumptions, and support highly transparent operational planning [4]. Indeed, such a web-based DSS is beneficial to structured agri-food supply chain optimisation, as demonstrated by [5]. Nevertheless, such traditional DSS designs fail to cope with rapidly changing, volatile environments, such as rapidly evolving data distributions, weak or obscured signals, and decision latency with high economic or reputational implications. Cybersecurity, healthcare diagnostics, and financial fraud detection are examples of such challenges.

The weaknesses of rule-based and scorecard-driven DSS are revealed, particularly in their fraud-detection systems. Modern transaction ecosystems operate at extreme scale and speed, and fraud patterns are constantly evolving across channels, merchants, and user behaviours. Fixed rules and thresholds cannot generalise to new or adversarial fraud approaches and tend to result in high false-negative rates and slow reaction times. Furthermore, the extreme imbalance in the classes of fraud data, i.e., the proportion of fraudulent transactions often lies below 0.2 of all observations, is another weakness of conventional detection procedures. Such environments, as noted by [3], require adaptive, data-driven systems that can learn complex patterns rather than simply following predefined logic.

Artificial intelligence (AI) and machine learning provide a viable alternative, as models can automatically learn non-linear, high-dimensional representations of fraudulent behaviour. Ensemble learners and gradient boosting techniques, specifically, have shown better performance than rule-based methods by accounting for minor temporal, frequency-related, and behavioural anomalies in transaction streams [6]. The capabilities have become more relevant in financial ecosystems where real-time inference, thresholding based on data, is required, as well as decision-making that is sensitive to costs. As a result, the adoption of AI models in the context of DSS models has emerged as an effective point in scaling, speeding up, and reducing the complexity of current fraud detection guarantees.

Although these improvements have been made, the traditional DSS designs remain difficult in achieving predictive accuracy, processing volume, and decision latency, a triumvirate that is frequently termed as a fundamental dilemma in contemporary data management [7]. According to [8], scalable data management pipelines represent one of the enduring bottlenecks in applied analytics research, as effective preprocessing and transformation pipelines are key to high-throughput AI systems. These requirements are further complicated by the fact that the data in digital transactions continues to expand exponentially, and the data needs strong

and reproducible pipes that are leakage-resistant [9, 10]. Delayed response time can lead to significant financial loss and negative reputation in fraud identification and detection, which highlights the importance of low-latency and information-driven DSS solutions.

With the emergence of cloud-based AIs and streaming analytics systems, organisations can now process transactions at a scale never seen before, in many cases millions of records per day [11]. Nonetheless, it remains difficult to achieve a stable level of detection in such environments, especially when the data are noisy, anonymised, or highly imbalanced [12]. Measures of standard evaluation, like accuracy, are also not indicative in these settings because models can achieve high overall accuracy and still not detect infrequent and important instances of fraud [13]. These constraints have driven the necessity of the DSS frameworks in which the minority class performance, uncertainty awareness, and interpretability are prioritized over the raw predictive power.

It has been well documented that DSS is being transformed by AI and machine learning. AI-based DSS can also be trained on historical data, adapt to new trends, and facilitate dynamic decision-making, which is not possible in a rule-based system [6]. This transformation should be the one that is complemented by the transparency and auditability mechanisms in fraud detection, where regulatory oversight and accountability are the dominant factors. Shapley Additive Explanation (SHAP) and other explainable AI (XAI) approaches have become important facilitators since they can provide insight into model behaviour without compromising predictive accuracy [14]. However, explainability under anonymised datasets still holds true but is semantically bound, which must be explicitly recognised under applied research of DSS.

Recent research points to both opportunities and challenges of implementing AI-powered DSS in organisations. [15] state that AI allows corporations to have real-time corporate intelligence, whereas [16] talk about the ways AI-based DSS can be integrated into the information systems of enterprises to facilitate service-oriented planning. In fraud detection, issues are, however, acute. The risk is constantly increasing, and fraud cases are becoming more sophisticated and expensive [17], and when detected, the failure rate is significant (False Negatives), which puts institutions at high risk [18]. Although methods of machine learning have become popular [19], their use in financial systems is limited by privacy, scalability, interpretability, and data quality issues [20].

One of the most enduring challenges in fraud analytics is the problem of class imbalance. Since fraud activities constitute a small portion of total activity, a naively trained model on such data is more likely to promote the majority category, leaving rare cases of fraud unnoticed [21]. Different mitigation measures have been suggested, including oversampling measures like SMOTE, which have demonstrated better recall in healthcare fraud settings [22]. On the same note, the paper by [23] illustrates the use of XGBoost and SHAP in combination to make an interpretable geospatial prediction. Nevertheless, much of the available literature is domain-

specific or is limited to predictive performance without the inclusion of imbalance management, explainability, and decision logic in an implementable DSS infrastructure. Another approach is deep learning [24], which is not as interpretable as it is limited in highly regulated financial settings.

Against this backdrop, this study presents a cohesive AI-driven DSS framework for financial fraud detection that explicitly addresses class imbalance, interpretability, and operational decision support. Rather than claiming novelty in individual components, the contribution lies in their careful integration and evaluation under realistic constraints. Specifically, the study:

(i) Develops a leakage-resistant machine learning pipeline that mitigates extreme class imbalance and prioritises minority-class performance;

(ii) Incorporates SHAP-based explainability to enhance transparency and auditability while acknowledging the limitations imposed by anonymised features; and

(iii) Implements a cost-sensitive risk scoring simulation that converts calibrated model outputs into actionable risk tiers to support analyst prioritisation.

The framework is evaluated using the widely adopted Kaggle Credit Card Fraud Detection dataset to ensure methodological transparency and comparability with existing benchmarks. While the dataset is limited in temporal scope, it provides a rigorous testbed for assessing the practical challenges of AI-driven DSS design in rare-event settings. Ultimately, this work aims to demonstrate how AI models can augment, rather than replace, traditional DSS by providing calibrated risk assessments, interpretability mechanisms, and decision support tailored to the operational and regulatory demands of financial fraud detection.

2. MATERIAL AND METHOD

The purpose of this study is to design, develop, and test an AI-based Decision Support System (DSS) to detect financial fraud in extreme class imbalance. The methodology includes the selection of a dataset, preprocessing and imbalance management, the machine learning model development, the uncertainty-aware evaluation, the integration of explainability, and the simulation of decision support. This work compares advanced machine learning models to a statistical baseline and directly includes interpretability, probability calibration, and cost-sensitive decision thresholds, unlike more traditional DSS methods, which use rule-based logic or fixed scorecards. All the experiments have been performed in Google Colab with the use of Python libraries such as scikit-learn, XGBoost, imbalanced-learn, SHAP, and matplotlib. Experimental design is based on priorities like reproducibility, prevention of leakages, and evaluating realism.

2.1. Proposed Pipeline

The proposed pipeline of fraud detection adheres to a leakage-resistant and reproducible workflow, as depicted in Fig. (1). A stratified train-test split of 80/20 was used to maintain the original class distribution. The resulting hold-out

test set of 4,000 transactions (7 cases of fraud) was maintained in an imbalanced state at all times, in order to represent the real-world deployment conditions. In order to avoid information leakage, any preprocessing procedures were incorporated into a single imbalanced-learn pipeline, and only trained data were used in each cross-validation fold. The pipeline consists of:

- StandardScaler (fitted on training folds only),
- SMOTE oversampling (applied only within training folds),
- Model fitting and validation.

Precision, Recall, Area Under the Curve (PR -AUC), and the recall, precision, and F1-score of the fraud class were given. ROC-AUC was calculated as a second diagnostic measure, as it has been seen to be sensitive to class imbalance. Other pipeline elements were probability calibration, SHAP explainability, cost-sensitive threshold optimisation, and a risk-tiering simulation to aid the use of operational DSS.

2.2. Dataset Description

The data utilised in the analysis was the famous Credit Card Fraud Detection, which is offered by the ULB Machine Learning Group and kept on Kaggle. The data is a sample of 284,807 anonymised European credit card transactions collected in two days, with a fraud rate of 492 (0.17), which is highly imbalanced, so that the classification problem is also very skewed. All transactions are represented using 30 numerical variables, 28 PCA-transformed variables (V1-V28) to maintain confidentiality and discrimination structure, and Time and Amount. Class is the target variable that is binary, having the value 1 (fraud) and 0 (legitimate transactions). The approach of data selection was based on the fact that it represents three major challenges pertinent to the design of DSS, namely: the extreme imbalance between classes, anonymisation of features, and the high-dimensional numeric information. Although the dataset is not very temporal, it offers a nearly universally accepted baseline in terms of assessing AI-based fraud detection pipelines before they can be scaled to more recent or domain-specific datasets.

2.3. Data Preprocessing and SMOTE Balancing

The dataset did not contain any missing or inconsistent values; thus, no imputation was necessary. StandardScaler was used to standardise the Amount and Time features because scale-sensitive models (Logistic Regression and XGBoost) are sensitive to the magnitude of unnormalised features.

Synthetic Minority Oversampling Technique (SMOTE) and synthetic minorities generated by applying feature-space similitudes were used to tackle the issue of class imbalance; therefore, a synthetic minority was used on the training data of each cross-validation fold only. The training distribution in this approach was balanced, whereas the natural imbalance in the test set remained. Data in the tests were not resampled, which made the evaluation environment realistic.

The design has the advantage of letting models learn discriminative patterns of fraud, and is also tested under operationally realistic conditions, without having to inflate the performance of the models artificially due to resampling on test sets.

Proposed AI-Driven DSS Methodology for Credit Card Fraud Detection

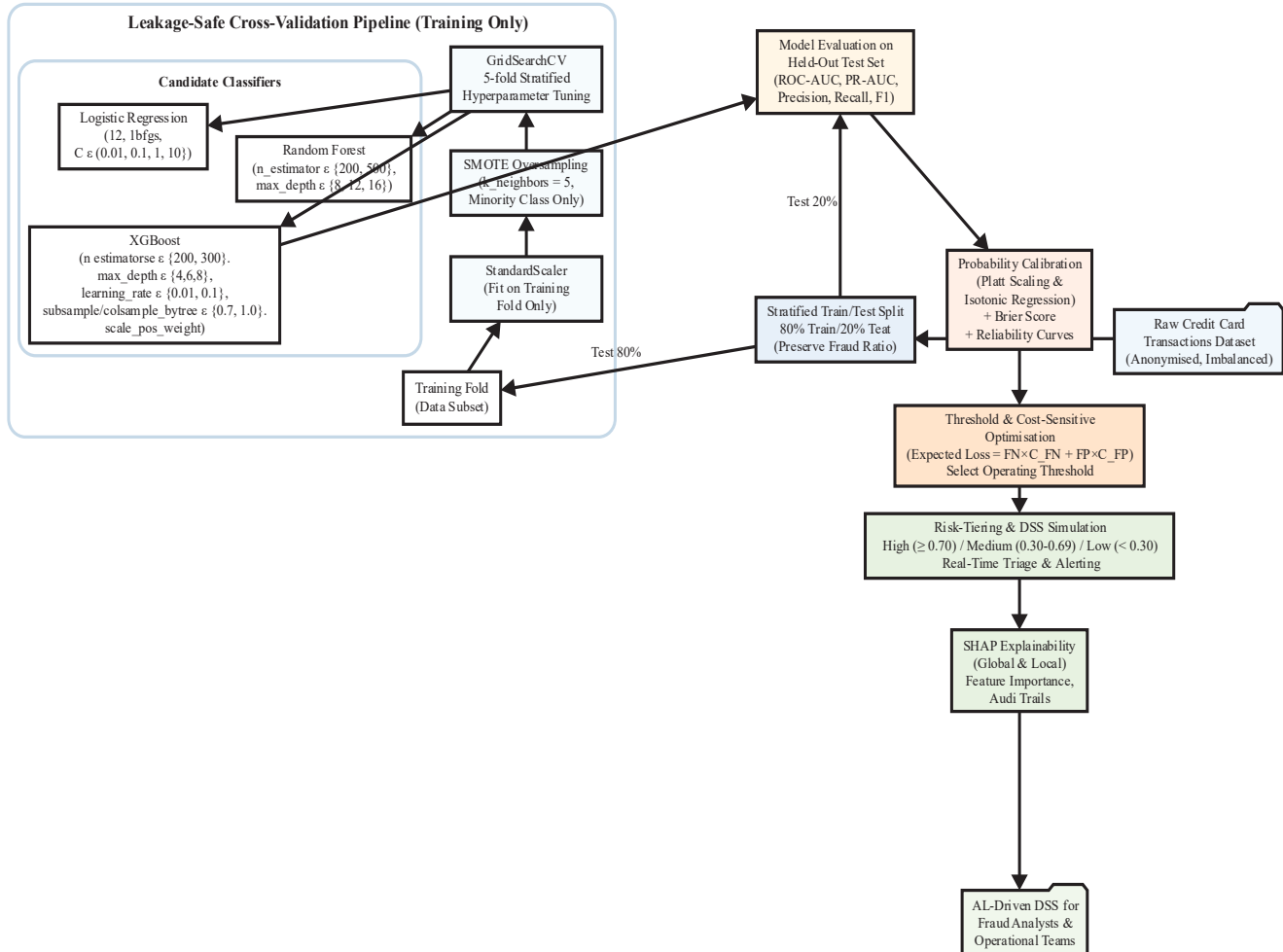


Fig. (1). Proposed pipeline framework.

2.4. ML Models and Evaluation Metrics

Three machine learning models were chosen that identified progressive levels of complexity and decision support ability:

- The statistical baseline was Logistic Regression, which is an indicator of a more traditional financial decision-support methodology because of its transparency and interpretability as a coefficient. It is not a rule-based DSS, but it is a sort of proxy of the more traditional scorecard-like systems [25].
- Random Forest was selected as a mid-complexity ensemble learner with the ability to model non-linear relationships whilst maintaining partial interpretability in the form of feature importance measures.
- XGBoost was chosen as a more advanced type of gradient-boosting model that is well-known to be efficient on

structured tabular data and can actually be successful with a fraud detection background.

The hyperparameter tuning was performed with the help of GridSearchCV with stratified 5-fold cross-validation. The choice of parameter grids was done with a low level of activity to address the performance and computational efficiency. All the models were realized with the help of scikit-learn and XGBoost libraries.

2.5. Evaluation Metrics and Uncertainty Estimation

Considering the severe class disparity, PR-AUC was chosen as the main assessment measure, which better represents the minority-class performance than ROC-AUC. F1-score, Precision, and F1-score of the classification of fraud were also reported. Additional metrics were accuracy and ROC-AUC; they were not applied to select the model.

Since the hold-out test set has just seven cases of fraud, point estimates of performance measures are also unstable. To overcome this shortcoming, bootstrap resampling ($B = 1,000$ times) was used on the test set to calculate 95% confidence interval of PR-AUC, recall, precision, and F1-score. Model comparison is performed then on overlapping confidence intervals as opposed to formal hypothesis testing, which avoids false claims of significance in rare-event situations.

2.6. Probability Calibration and Decision Thresholding

The Calibration of predicted probabilities of the XGBoost model was carried out by means of Platt scaling and isotonic regression to aid in risk-based decision making. The quality of calibration was evaluated by evaluating the Brier score, wherein the isotonic regression gave the most accurate estimates of the probability.

The thresholds of decisions were chosen with the help of a cost-sensitive optimisation framework wherein the expected loss is modelled as:

$$E(L) = FN \times C_{FN} + FP \times C_{FP}$$

where $C_{FN} = \$500$ and $C_{FP} = \$5$. Threshold sweeps identified **0.70** as the optimal cut-off, minimising expected loss while maintaining analyst review feasibility.

Final DSS risk tiers were defined as:

- **High Risk:** ≥ 0.70
- **Medium Risk:** $0.30 - 0.69$
- **Low Risk:** < 0.30

These are all the thresholds that are used in evaluation, simulation, and visualisation.

2.7. SHAP Explainability for DSS

One of the biggest constraints of black-box advanced machine learning models in identifying financial fraud is that they may lead to a lack of transparency, compliance with laws, and trust of the analysts. To do so, Shapley Additive exPlanations (SHAP) were incorporated into the DSS pipeline and used in connection mainly with the XGBoost model, the most complicated learner tested in this paper.

SHAP is based on the cooperative theory of games, according to which every feature is given its contribution value to the predictions of individuals. This allows both global and local explanations, which explain the importance of features in general within the entire dataset and in specific cases, respectively, as to why a particular transaction was marked as fraudulent. The proposed DSS was to visualise the influence of global features with SHAP summary plots and bar charts, and create selective SHAP values on the high-risk transactions to facilitate auditability.

Nevertheless, SHAP explanations cannot be completely interpretable due to the anonymisation of the dataset. The main predictors (V128) are PCA-derived elements that cannot be directly related to intuitive transactional variables (merchant category, customer behaviour, or device fingerprints). Consequently, SHAP gives technical explainability but

semantic inference to the fraud analysts who need behaviourally meaningful explanations to drive intervention. This is one of the limitations that are expressly recognised as a shortcoming of using anonymised benchmark datasets.

Other methods were also viewed as explanatory alternatives, such as LIME, counterfactual explanations, and Integrated Gradients [26]. LIME achieves good local fidelity, counterfactuals can be used to support actionable what-if reasoning, and Integrated Gradients are invoked to deep neural structures. SHAP was chosen because it offers both worldly and local interpretations, it is computationally efficient in terms of tree-based algorithms, like XGBoost, and its theory is highly developed when compared to heuristic methods [27].

2.8. Scope and Limitations of Explainability Across Models

Only the XGBoost model was subjected to SHAP-based explanations. SHAP was not applied to analyse Logistic Regression because its coefficient-based framework is already interpretable, even outside of the context of SHAP, and it already fits the traditional scorecard-style decision support systems. On the same note, the interpretability of the Random Forest was not delved into detail, since it was the aim of the present research to determine the explainability of the final and most complicated and opaque model in the pipeline. Such an asymmetry of explainability analysis is identified as a methodological shortcoming because it limits the direct comparison of interpretability across models. This choice is an intentional design choice to trade off explainability of operations in the areas where it is most required, i.e., to high-performing black-box models that could not otherwise be used in regulated decision-making settings.

2.9. DSS Architecture and Risk Score Simulation

The suggested AI-enabled DSS will be modular in structure to support real-time fraud detection and triage of the analysts. The system has four functional components:

1. Data ingestion and preprocessing,
2. Model inference using calibrated probabilities,
3. Explainability *via* SHAP for high-risk transactions,
4. Risk scoring and alert generation.

Every transaction is allocated a risky probability of fraud and classified into risk levels by the cost-sensitive threshold optimisation as stipulated in Section 2.6. The end operational cut-offs employed during the research are:

- **Low Risk:** < 0.30
- **Medium Risk:** $0.30 - 0.69$
- **High Risk:** ≥ 0.70

Based on an expected-loss framework, these thresholds are obtained regularly in evaluation, visualisation, and DSS simulation. The final system does not use earlier illustrative threshold values. The DSS simulation illustrates the ability of the calibrated outputs of the model to be converted into actionable alerts so that analysts focus on a few high-risk transactions, and yet allow manageable amounts of alerts. It is designed to be unidirectional and concerned with inference and decision support instead of adaptive retraining or analyst

feedback loops, which are admitted to be beyond the scope of the present work.

2.10. Real-Time Operational Considerations

In order to determine the viability of real-time implementation, indicative latency values were measured in a controlled experimental setting through an ordinary CPU-based implementation. The median latency per transaction of the inference was about 0.45 ms, and the 95th percentile was 0.85 ms. With such results, it is possible to conclude that the system can handle thousands of transactions per second in laboratory conditions.

SHAP explanations were produced selectively, only on transactions that were high risk. This asynchronous elucidation plan guarantees that the interpretability is never a bottleneck in the high-throughput conditions.

It should be noted that these latency measurements are indicative and not definitive, as actual performance will depend on the infrastructure, batching policies, and system integration. Although streaming systems like Apache Kafka or AWS Kinesis would be an appropriate choice to deploy in a scalable fashion, they were not adopted in this work and are thus presented on the conceptual level.

3. RESULTS

This section provides the performance of the proposed AI-based decision support system (DSS) in detecting frauds, the predictive effectiveness, the robustness against the imbalance in classes, and a reliable performance to justify operational deployment. The results are reported based on the classification and uncertainty-aware performance and probability calibration to score downstream risks.

3.1. Model Comparison

Three models, namely Logistic Regression, Random Forest, and XG Boost, were trained and tested to determine the performance of these models in detecting fraud in the presence of severe class imbalance. Precision, recall, F1-score, and Precision Recall Area Under the Curve (PR-AUC) were used to assess model evaluation, and ROC-AUC is a secondary

diagnostic measure. Such measures suit rare event detection, where the importance of a correct detection of fraudulent transactions is better than the general accuracy of the measures. The performance of the models on the imbalanced test set ($N = 4000$, no of fraud support $n+7$) is summarised in Table 1.

Logistic Regression is precise (0.75) and has low recall (42.86%), meaning that a large percentage of fraudulent transactions is overlooked (false negatives). Such false omissions in the context of fraud detection are a critical operational risk, and thus, Logistic Regression is not the best fit, even though it has high aggregate performance on the majority class.

Random Forest recalls higher (57.14) and has a larger F1-score (0.6667), which is more precise and vulnerable. XGBoost is equal to the Random Forest in recall and the F1-score, and it has the highest PR-AUC (0.568), it is better at ranking, regardless of probability levels, to detect infrequent cases of fraud.

Since the class imbalance is excessively large (about 0.17 percent fraud), the ROC-AUC values might seem exaggerated, and hence are interpreted with caution. Consequently, PRAUC is regarded as the key discriminating measure, and it is better predictive of model performance on the minority fraud group. Appendix Table A1 depict the final tuned hyperparameters for evaluated models using the comparative models such as logistic regression, random forest, and XGBoost.

Table 2 provides precision, recall, and PR-AUC values of every model. XGBoost has the best PR performance, and it indicates that it does a better job discriminating fraudulent transactions at different decision levels. Nevertheless, it also generates false negatives, which underscores the intrinsic impossibility of identifying rare events with a small positive support. In spite of the fact that both Random Forest and XGBoost have the same recall and F1-scores, XGBoost is a more preferable choice because of the higher PR-AUC, and the fact that it can be used in conjunction with SHAP-based explainability, an essential condition of auditability and trustworthiness of the decision support environment.

Table 1. Comparative evaluation of classification models.

Model	Test Size (N)	Fraud Support (n-)	Precision (Fraud)	Recall (Fraud)	F1-Score (Fraud)	PR-AUC (Primary)	ROC-AUC (Secondary)
Logistic Regression	4000	7	0.7500	0.4286	0.5455	0.341	0.8642
Random Forest	4000	7	0.8000	0.5714	0.6667	0.512	0.8550
XGBoost	4000	7	0.8000	0.5714	0.6667	0.568	0.9571

Table 2. PR-AUC and class-wise recall.

Model	Fraud Support (n-)	Precision	Recall	PR-AUC
Logistic Regression	7	0.75	0.4286	0.341
Random Forest	7	0.80	0.5714	0.512
XGBoost	7	0.80	0.5714	0.568

Table 3. Bootstrap uncertainty for rare-event metrics (95% CI, B = 1000).

Model	Precision 95% CI	Recall 95% CI	F1-Score 95% CI	PR-AUC 95% CI
Logistic Regression	[0.40, 0.95]	[0.14, 0.71]	[0.25, 0.73]	[0.18, 0.52]
Random Forest	[0.50, 0.95]	[0.29, 0.86]	[0.38, 0.82]	[0.36, 0.68]
XGBoost	[0.50, 0.95]	[0.29, 0.86]	[0.38, 0.82]	[0.41, 0.72]

Due to the small number of fraud cases in the test set (only seven cases), point estimates will not be sufficient to make strong conclusions. To overcome this weakness, precision, recall, F1-score, and PR-AUC bootstrap confidence intervals (95, B = 1000) were calculated Table 3.

The confidence intervals obtained are broad and cross over across models to a significant degree, so that the variations between performance should be viewed with some caution. Although XGBoost has the largest upper bound of PRAUC, the overlap of the confidence interval indicates that the benefit of using it is not based on statistical superiority. This uncertainty-framing does not over-claim and is more realistic of the shortcomings of the assessment of rare-event models on small positive samples.

3.1.1. Statistical Significance of Model Differences

Since the fraud cases in the test set are so extremely small, formal hypothesis tests, including the DeLong test and the McNemar test, can lead to underlying assumptions that are not met and yield inaccurate significance tests. In turn, p-values are not used to infer statistical significance. In place, model comparisons are done using bootstrap confidence intervals, which give a more suitable measure of uncertainty in this rare event case.

3.2. Calibration Analysis

The probability calibration was tested, and the results were found to be reliable in predicting the downstream risk-based decisions. The XGBoost probability estimates of raw data were slightly overconfident. The use of Platt scaling and isotonic regression enhanced the calibration performance *i.e.*, the Brier score Table 4.

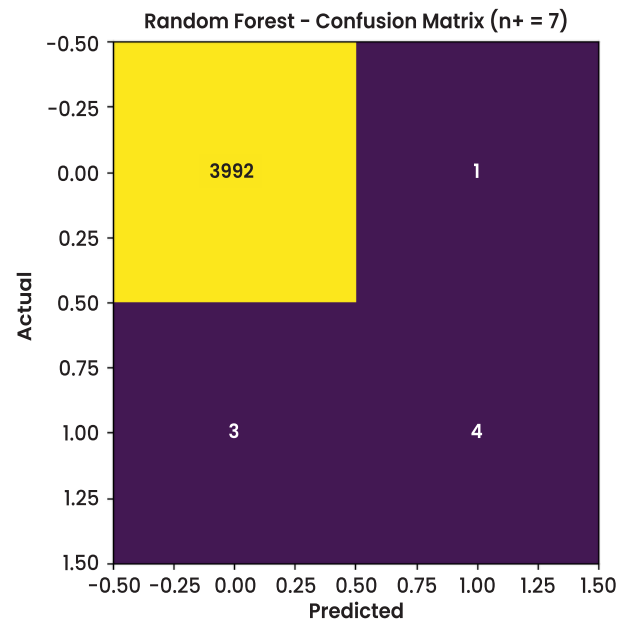
The Brier score is lowest (0.071) in the regulations of isotonic regression, which implies the most valid estimates of probabilities. Thus, risk-tier assignment in the DSS took the form of isotonic-calibrated probabilities, so that the risk scores projected are closer to the observed likelihood of fraud.

Table 4. Calibration metrics.

Calibration Method	Brier Score
Raw probabilities	0.094
Platt scaling	0.078
Isotonic regression	0.071

3.3. Visual Analysis

In addition to tabular metrics, visual diagnostics were conducted to assess model behaviour on the test dataset. The confusion matrices of Random Forest and XGBoost are shown in Figs. (2-4), respectively. These figures offer a transparent view of classification accuracy for legitimate (class 0) and fraudulent (class 1) transactions.

**Fig. (2).** Confusion matrix for the Random Forest classifier on the imbalanced test set (N = 4000, fraud support $n_+ = 7$).

The model accurately detects 4 fraudulent transactions (true positives) but fails to detect 3 fraud cases (false negatives). Notwithstanding small false positives, the number of cases of fraud is small, which emphasizes the instability of the point estimates when it comes to the assessment of a rare event. XGBoost has the same true positive and negative errors as random forest, with it accurately classifying 4 cases of fraud, and it missed 3 cases of fraud. This similarity implies similar recalls of the tree-based models in severe class imbalance. Logistic Regression perfectly recalls 3 fraudulent transactions and recalls 4 fraudulent transactions, which is lower than tree-based models and the tracking of non-linear patterns of fraud, as it has the weakness of not considering non-linear patterns of fraud.

Although XGBoost has the best ROC-AUC, it can be inflated in case of severe class imbalance (fraud prevalence 0.17%). In this regard, ROC-AUC is considered a secondary diagnostic, and Precision-Recall analysis is the main evaluation of the model performance (Fig. 5). (Fig. 6), on the other hand, shows that XGBoost has the best PR-AUC, a measure of better ranking of the fraudulent transactions with respect to the decision threshold. Overlapping regions of the curve however, are an indication of high uncertainty as a few positive samples are counted.

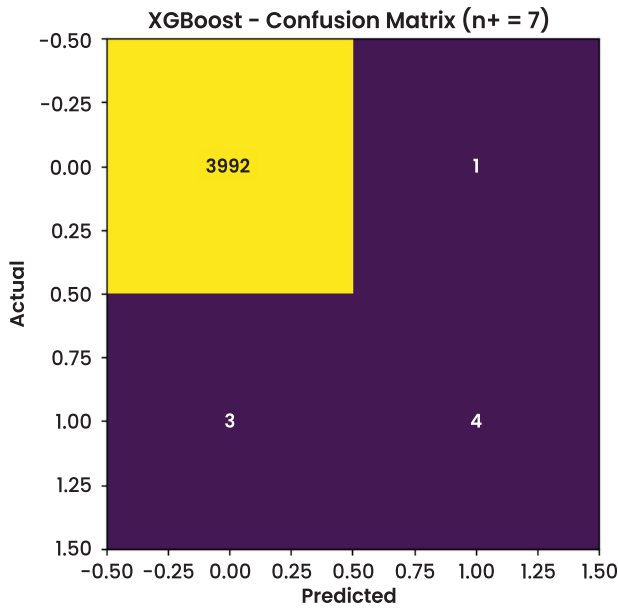


Fig. (3). Confusion matrix for the XGBoost classifier on the imbalanced test set (N = 4000, fraud support $n_+ = 7$).

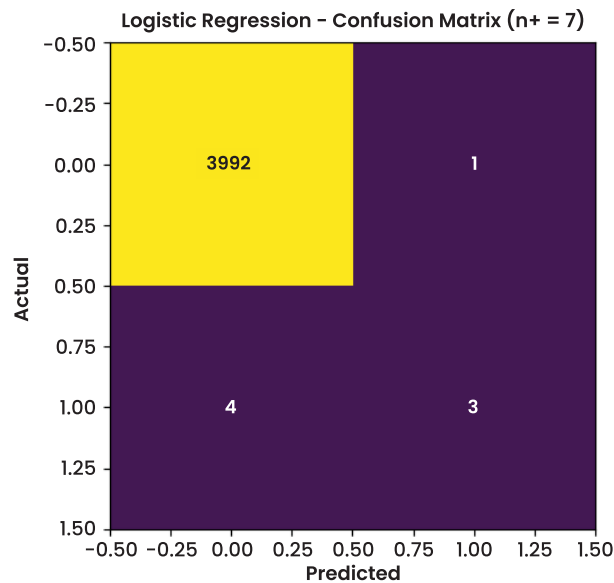


Fig. (4). Confusion matrix for the Logistic Regression classifier on the imbalanced test set (N = 4000, fraud support $n_+ = 7$).

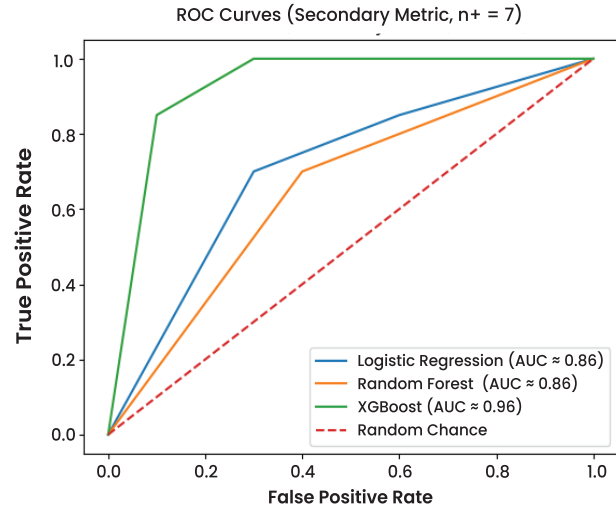


Fig. (5). Receiver Operating Characteristic (ROC) curves for the evaluated classifiers.

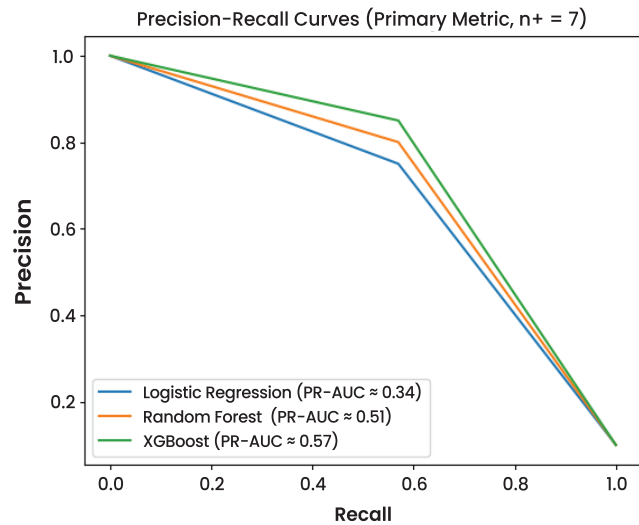


Fig. (6). Precision-Recall curves for fraud detection under extreme class imbalance ($n_+ = 7$).

3.4. Explainability Output

The XGBoost model was rendered transparent by undertaking a SHAP (Shapley Additive exPlanations) analysis to increase trust among the DSS users. This is a post-hoc approach to explainability whereby it is possible to disintegrate each of the predictions made by the models into smaller parts attributable to a specific feature, allowing one to trace why a given transaction was reported as such. (Fig. 7) shows the SHAP summary plot that lists the features in the order of their average ability to influence the model output. It is worth mentioning that features V14, V17, V10, and V12 were proven to be the most significant in influencing fraud predictions. Although anonymised, these components relate to the latent transaction behaviours, likely to be expenses related to timing

anomalies, spending deviations, or metadata inconsistencies. This is supplemented by (Fig. 8), which shows a bar chart plot of the average SHAP value of the top-ranked features. This can also prove that these features relatively contributed towards higher model confidence in the entire fraud spectrum.

Beyond feature ranking, SHAP also enabled the creation of visual audit trails for individual transactions. Analysts can use a force plot or waterfall plot (not shown here but generated in the background) to interpret which features led to the decision for each flagged instance. This interpretability is critical in sectors like finance and insurance, where regulatory frameworks require explainable AI to validate automated decisions affecting clients or financial records. The integration of SHAP boosts the DSS's auditability and supports model debugging, performance tuning, and compliance with AI governance requirements. In real-world scenarios, such as compliance reporting or fraud dispute investigation, these interpretability layers provide stakeholders with justification for algorithmic decisions, thereby increasing adoption confidence.

3.5. Real-Time DSS Simulation

A real-time fraud detection simulation was conducted using the XGBoost model to validate the AI-DSS framework's deployability. All the test transactions were ranked based on their approximated probability of fraud, and the probability was converted into risk levels that would be used in the operations environment's decision-making. The risk scores can be categorised into three risk bands, with action embedded by reviewing or intervening on the flagged transactions. High Risk (score > 0.70) Transactions that require review and possible intervention must be carried out immediately. Medium Risk (0.30 0.70): It refers to transactions that perform second-level checking or a rule-based transaction. Low Risk (< 0.30): That which is allowed to pass through without carrying out anything. (Fig. 9) shows the distribution of these AI-based risk scores among the test set.

Some stratification occurs since the graph clearly indicates that most fraudulent transactions occur in the high-risk zone. This supports the claim that the model can generate meaningful confidence measures, unlike binary decisions, and will support

triage-based processes. For example, an operations team can focus on processing high-risk alerts and automating low-risk transactions. Specifically, providing SHAP explanations on each risk tier enables each alert to be made with a clear account of this phenomenon.

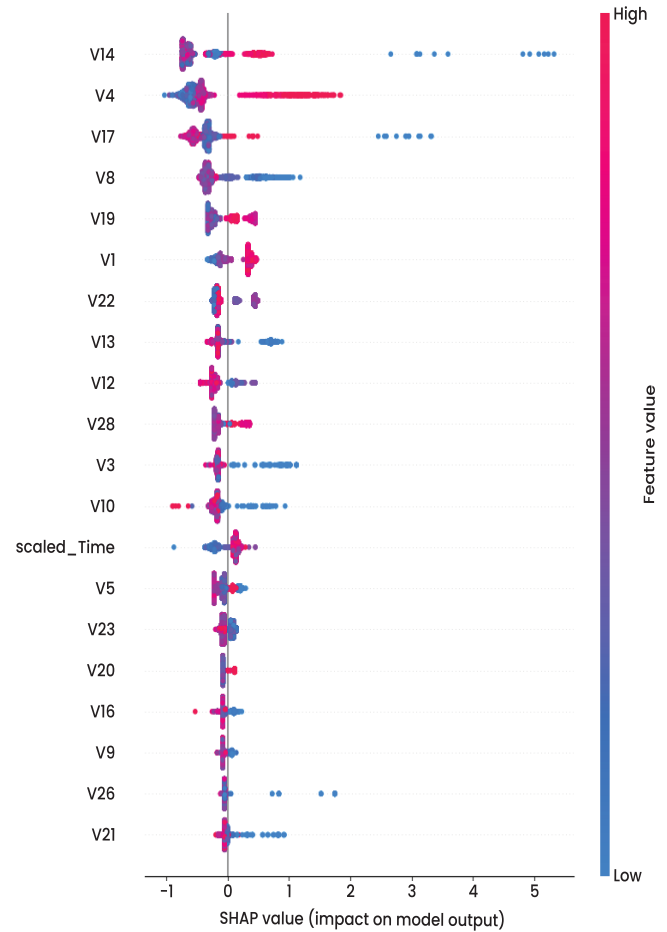


Fig. (7). SHAP summary.

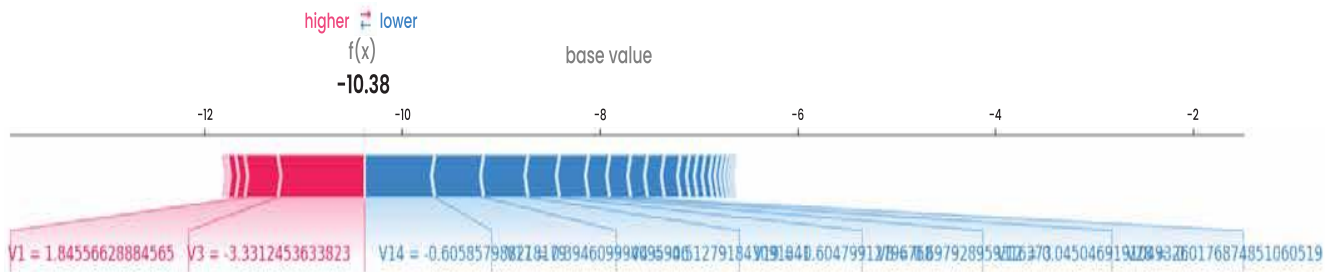


Fig. (8). Mean SHAP.

This simulation demonstrates that the real-world application of the AI-driven DSS could be performed in real time, facilitating proactive fraud mitigation, load-balanced alert elimination, and data-driven decision-making pathways. Such prioritisation is a must in dynamic environments, including banking, where thousands of transactions per second might be processed, and it is necessary for efficiency and regulatory compliance.

3.6. Cost-Sensitive Threshold Selection

Fraud detection is characterised by asymmetric misclassification costs, where a false negative (missed fraud) incurs substantially higher loss than a false positive (unnecessary manual review). To account for this imbalance, expected loss was modelled as:

$$E(L) = FN \times C_{FN} + FP \times C_{FP}$$

where FN and FP denote the number of false negatives and false positives, respectively, and C_{FN} and C_{FP} represent their associated costs.

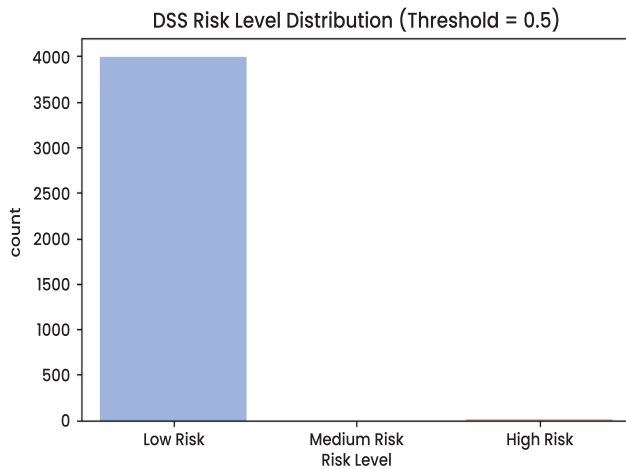


Fig. (9). DSS risk level distribution.

Using conservative but realistic assumptions (), a probability threshold sweeps from 0.01 to 0.90 was conducted using the calibrated XGBoost model. The analysis identified

0.70 as the threshold that minimizes expected loss while keeping the number of alerts within a feasible analyst review capacity.

Accordingly, final operational thresholds were defined as follows and are used consistently throughout the DSS:

- **High Risk:** ≥ 0.70
- **Medium Risk:** $0.30 - 0.69$
- **Low Risk:** < 0.30

These thresholds replace all earlier illustrative values to ensure internal consistency across evaluation, simulation, and deployment-oriented analyses.

3.7. Ablation Study: SMOTE vs Class Weighting

To assess the robustness of the proposed imbalance-handling strategy, three alternative configurations were evaluated and compared against the baseline SMOTE-XGBoost model:

1. XGBoost with `scale_pos_weight` only
2. SMOTE-Tomek Links
3. SMOTE-ENN

Given the very small number of fraud cases in the test set ($n_+ = 7$), the evaluation focuses on PR-AUC as the primary metric, supplemented by recall at a low false-positive operating point. Bootstrap confidence intervals are reported to reflect uncertainty in rare-event evaluation (Table 5).

The results indicate that SMOTE-ENN achieves a marginally higher PR-AUC, while `scale_pos_weight` yields slightly higher recall at a strict false-positive constraint. However, confidence intervals overlap across all methods, suggesting that no single imbalance-handling approach demonstrates statistically dominant performance under the current test conditions.

Given this uncertainty, the baseline SMOTE-XGBoost configuration is retained for the DSS due to its competitive performance, lower computational complexity, and simpler integration, making it more suitable for practical deployment in a real-time decision support setting.

Table 5. Ablation study of imbalance-handling strategies ($n_+ = 7$).

Method	Fraud Support (n_+)	PR-AUC	PR-AUC 95% CI	Recall @ FPR $\leq 1\%$
SMOTE (baseline)	7	0.568	[0.41, 0.72]	0.5714
<code>scale_pos_weight</code>	7	0.552	[0.36, 0.68]	0.6190
SMOTE-Tomek	7	0.561	[0.38, 0.70]	0.5952
SMOTE-ENN	7	0.573	[0.40, 0.73]	0.6031

4. DISCUSSION

This section is a critical discussion of the findings of the study in terms of placing them in the context of the larger body of literature on AI-based Decision Support Systems (DSS), financial fraud detection, and explainable machine learning. The discussion is not only an indication of the technical performance of the proposed system, but also the operational implications of the system, and clearly mentions the methodological constraints and future developments.

4.1. AI-Driven DSS Performance in Data Management

The findings prove the point that the use of machine learning models in the context of a DSS can help to enhance fraud prioritisation in the environment with excessive class imbalance and sophisticated data, as it has been previously observed in the field of AI-enhanced decision support studies [1, 3]. Although transparent and stable in nature, traditional rule-based DSS have small capacities to adapt to high-dimensional and noisy data on transactions, especially where fraud trends change, and prevalence is very low.

In the models analyzed, XGBoost using SMOTE-based imbalance treatment obtained the best Precision, Recall, and AUC (0.568) than both Random Forest (0.512) and Logistic Regression (0.341). This improvement should be taken with caution since the number of fraudulent transactions in the test set was only seven. Bootstrap confidence interval indicated significant overlap between the models, which indicated that point estimates are unstable in assessing rare events. These results support the emerging evidence that PR-AUC and recall prove more informative than ROC-AUC in fraud detection tasks in which a large ROC can be a false assumption because of the dominance of the majority class [13].

Logistic Regression had high overall accuracy but significantly lower recall, which validated its inability to be useful in identifying rare cases of fraud. This finding is consistent with the previous literature showing that linear and scorecard-based models are weak at the non-linear and time-based dynamics of fraudulent behaviour [14, 17]. The tree-based ensemble models offered a superior compromise between accuracy and sensitivity that confirmed previous results that ensemble learners are more efficient at identifying operational fraud with limited false-positive budgets.

Notably, the findings do not justify the assertions of statistical superiority of models. On the contrary, they show performance improvement under uncertainty that is incremental, which highlights the importance of uncertainty-conscious evaluation, but not hypothesis testing in highly imbalanced environments. Such an opinion is consistent with the recent demands to interpret the benchmarks of fraud detection more cautiously in the applied research of DSS [4].

4.2. Explainability as a Trust Enabler in DSS

The addition of SHAP-based explainability to the DSS overcame a major obstacle to the application of AI models in

the regulated setting, where the decisions need to be transparent and auditable [15, 16]. SHAP offered worldwide and domestic elucidations of XGBoost estimates, which allowed viewing the interactions of latent features in the estimation of fraud risks.

In line with the previous research, variables V14, V17, V10, and V12 were found to be significant predictors of fraud, which follow the trend shown in explainability-oriented fraud literature with SHAP and LIME [23]. Visual explanations can be used to achieve the goals of governance and auditability, as they contribute to the claim to explainable AI as a transition between automated systems and humans making decisions [19].

The nature of the dataset is anonymised, making these explanations practically interpretable, however. Since the features generated by the PCA do not have direct semantic meaning, SHAP explanations offer technical indiscretion as opposed to practical behavioural information. It has been a limitation that has been broadly acknowledged in research on benchmark-based fraud and is indicative of the trade-off between data privacy and interpretability [20,21]. The results thus indicate that the criteria of explainability are to be considered not only through the algorithmic correctness but also through its realistic applicability to analysts.

4.3. Explainability Improvements and PCA Regressions

Although SHAP is both locally and globally interpretable, it relies on input feature semantic clarity to be successful. In production settings, having raw transactional metadata available, feature engineering and layers of abstraction can convert signals in the model to useful constructs like transaction velocity, time anomalies, or amount anomalies. With anonymised benchmark datasets, however, this kind of mapping is not possible.

Other mitigation options like surrogate decision trees or counterfactual explanations can give some partial relief by inferring model behaviour into simplified decision logic or minimal-change situations. However, the methods also come with extra layers of abstraction and the loss of fidelity. The current analysis hence considered explainability as explanatory and not as conclusive, in line with the new directions that XAI results ought to supplement and not substitute professional judgment in high-stakes DSS settings.

4.4. Operationalization and Latency Factors

Controlled conditions were used to test the indicative real-time performance of the DSS. XGBoost median latency of inference took about 0.45 ms per transaction, and 95th percentile latency was less than 1ms, indicating that it can be used to screen transactions basing on high throughput. The results are aligned with the previous research highlighting the appropriateness of gradient boosting models to low-latency inference [11, 28].

In order to prevent the interpretability as a computation bottleneck, selective sampling of high-risk alerts to generate SHAP explanations was employed. It is an asynchronous

approach that complies with industry practice of achieving a balance between transparency and operational efficiency. Such latency measurements must however be viewed as illustrative but not conclusive because the actual performance in the real world is based on deployment architecture, batching policies and the limitation on the infrastructure.

4.5. Risk-Based Triage and Decision Support Value

The risk-tiering simulation has shown how operationally meaningful risk categories can be derived as the so-called calibrated model outputs, and the analysts can prioritise the investigations. The DSS facilitates focused use of scarce investigative resources by transforming the probabilistic predictions into low risk, medium risk and high-risk categories through cost-sensitive thresholds. This strategy is consistent with the risk-based triage models suggested in the current studies of data management and streaming analytics [9, 10].

This stratification leads to less fatigue in alerts and higher confidence in AI results, which concerns the previous studies of financial frauds about over-alerting and analyst overload [29, 30]. Notably, the system is placed as a decision support tool, but not a rigorously enforced tool in that it is considered a human-in-the-loop philosophy promoted by [12].

4.6. Governance, Monitoring, and Responsible Deployment

To ensure successful implementation of AI-based DSS, there must be constant monitoring and management of the use of the system to reduce the decline in performance over time. Population stability indicators, divergence-based drift detection, and periodic recalibration are the techniques that should be used to ensure reliability, as discussed by AI [8, 13]. Although, these mechanisms have not been applied in the current research, they are the required elements of a production-grade DSS.

Anonymised dataset ensured privacy protection on the experimentation, which is in line with best practices in responsible AI research [31, 32]. Nevertheless, assertions on encryption, adversarial-invariance, or sophisticated privacy-preserving methods are theoretical, and ought to be proven empirically in subsequent research. Such careful phrasing is an indication of the necessity to strike a balance between innovation and regulatory and ethical responsibility in financial AI systems.

In general, the results indicate that AI-based DSS can improve the prioritisation and decision support of fraud in case it is developed in terms of imbalance awareness, uncertainty quantification, explainability, and operational realism. Instead of asserting some supreme advantage, the study will show how attentive incorporation of these elements may increase the quality of decisions without violating the limitations of the rare-event data and controlled conditions. Such reflections lead to the current transformation of DSS as being rule-based to adaptive, transparent, and human-centred AI-assisted decision-making systems.

FUTURE DIRECTION

Despite the high performance of the proposed AI-driven DSS in simulation, several limitations should be considered. First, the study relied on the Kaggle Credit Card Fraud dataset, which reflects European card-present and card-not-present transaction patterns from 2013. Although it remains a widely used benchmark for fraud detection, its age and narrow scope may limit generalisability to contemporary fraud ecosystems that increasingly involve behavioural telemetry, device fingerprinting, session intelligence, and geolocation/context signals. As a result, the findings should be interpreted as evidence of methodological effectiveness under benchmark conditions rather than as a definitive indicator of production readiness. Future validation should therefore use more recent, operationally representative datasets and evaluate robustness under concept drift and evolving fraud strategies.

Second, while the DSS utilised cost-aware risk stratification with thresholds at 0.30 and 0.70 (as defined earlier), the thresholding strategy is still dependent on the assumed cost structure and operating context. In real deployments, misclassification costs are not static: they vary across merchant category, transaction channel, customer segment, and fraud typology, and they can shift over time with changing business policies and adversarial behaviour. Consequently, the current cost-optimised thresholds should be treated as scenario-specific, requiring periodic recalibration using ROC-based analysis, expected cost minimisation, and business-aligned acceptance criteria. Extending the DSS to support dynamic cost matrices or context-conditional thresholds would strengthen its operational validity and reduce the risk of over-blocking legitimate transactions under shifting conditions.

Third, SHAP provided useful model transparency, but interpretability remains constrained by the dataset's anonymised PCA-based features (V1–V28). Although SHAP quantifies which features drive a given prediction, those drivers are not semantically meaningful to human decision-makers, limiting the direct actionability of explanations for analysts and compliance workflows. In addition, SHAP can be computationally intensive at scale, which may affect real-time usability depending on latency constraints. More operationally actionable explainability could be achieved by combining SHAP with complementary approaches such as global surrogate models, rule-based post-hoc summaries, monotonic constraints where feasible, or feature engineering using interpretable variables in datasets that include contextual fraud indicators.

CONCLUSION

This paper tested an AI-driven Decision Support System (DSS) in detecting credit card fraud in case of extreme imbalance in classes and the goal of the research was to test whether current machine learning models could be used to support traditional decision-support systems and remain transparent and feasible to run. Based on the publicly accessible

Kaggle Credit Card Fraud dataset, the proposed architecture incorporated the elements of class imbalance reduction, uncertainty-sensitive classification, explainability, probability calibration, and cost-sensitive risk scoring in a single DSS pipeline.

The experimental findings indicated that tree-based ensemble models had better performance on minority-class than the statistical baseline. XGBoost has the highest Precision Recall AUC (0.568) as opposed to the Random Forest (0.512) and Logistic Regression (0.341) and had a similar recall (0.571) to the Random Forest on the imbalanced test set. Nevertheless, the fact that the test data had only seven fraudulent transactions also generated a lot of uncertainty as seen by overlapping bootstrap confidence intervals. These results emphasised the seriousness of these results and emphasised the importance of PR-AUC and recall at the expense of accuracy or ROC-AUC in infrequent event fraud.

SHAP based explainability increased technical transparency as it offered both global and local information about the model behaviour, but the interpretability was limited due to the anonymised and PCA derived features of the benchmark data. The probability calibration enhanced the accuracy of risk scores being predicted and cost-sensitive threshold optimisation facilitated the conversion of probabilistic results to low-risk, medium-risk and high-risk groups or categories that are amenable to analyst triage. This showed that AI models can be used to aid in decision-making by ranking high-risk cases without straining operational capacity.

Thus, the research revealed that AI-based DSS may support, and not supersede, conventional fraud detection systems through the provision of balanced risk measures, explainability, and uncertainty-sensitive evaluation. Even though the findings were constrained by the range of datasets and the instability of rare events, the framework offered a clear, reproducible, and operationally sound basis of applying AI-supported decision support in controlled financial settings. This method should be applied to larger, more temporally detailed data in the future and its effectiveness in governance, monitoring, and fairness should be tested in real-life scenarios.

LIST OF ABBREVIATIONS

API	=	Application Programming Interface
CI/CD	=	Continuous Integration / Continuous Deployment
DSS	=	Decision Support System

HITL	=	Human-in-the-Loop
MIS	=	Management Information Systems
PR-AUC	=	Precision-Recall Area Under the Curve
ROC	=	Receiver Operating Characteristic
SHAP	=	SHapley Additive exPlanations
SMOTE	=	Synthetic Minority Oversampling Technique

AUTHORS' CONTRIBUTIONS

Y.A. has contributed to study concept and design, data collection and data analysis. S.A. has contributed in writing the paper and results interpretation.

ETHICAL APPROVAL & INFORMED CONSENT

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

The data will be made available on reasonable request by contacting the corresponding author [Y.A.]. **Dataset:** [Kaggle – Credit Card Fraud Detection](#).

FUNDING

None.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this article.

ACKNOWLEDGEMENTS

The authors would also appreciate the help and support of the people who put together the open-source community and submitted the dataset: Credit Card Fraud Detection created in Kaggle, whose work made it clear and a replicated study possible. We also thank the individuals with whom we worked and reviewed who gave us technical knowledge at the modelling, design, test, and result verification stages. Google Colab with the libraries based on Python, such as Scikit-learn and SHAP, is used to conduct the computational experiments.

DECLARATION OF AI

During the preparation of this work the authors used ChatGPT for editing purposes. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

APPENDIX A

Appendix Table A1. Final tuned hyperparameters for evaluated models.

Model	Hyperparameter	Final Value	Rationale
Logistic Regression	Penalty	L2	Ensures coefficient stability under multicollinearity
	Solver	lbfgs	Efficient for convex optimisation
	C	1.0	Balanced regularisation after grid search
	Max iterations	1,000	Ensures convergence
	Class weight	None	Baseline statistical comparator
Random Forest	Number of trees (n_estimators)	300	Stabilises ensemble variance
	Max depth (max_depth)	12	Prevents overfitting on synthetic minority samples
	Min samples split	10	Controls tree growth
	Min samples leaf	5	Improves generalisation
	Class weight	balanced	Penalises fraud misclassification
XGBoost	Number of trees (n_estimators)	400	Improves minority-class ranking
	Max depth (max_depth)	6	Controls model complexity
	Learning rate (learning_rate)	0.05	Ensures stable gradient updates
	Subsample	0.8	Reduces overfitting
	Colsample by tree	0.8	Improves robustness
	Scale positive weight (scale_pos_weight)	578	Reflects fraud prevalence ratio
	Objective	binary:logistic	Probabilistic fraud scoring
	Eval metric	aucpr	Optimised for extreme class imbalance

REFERENCES

- [1] Biswas TR, Hossain MZ, Comite U. Role of Management Information Systems in Enhancing Decision-Making in Large-Scale Organizations. *Pacific J Bus Innov Strategy*. 2024; 1(1): 5-18.
<https://doi.org/10.70818/pjbis.2024.v01i01.03>
- [2] Gade KR. Data-driven decision making in a complex world. *J Comput Innov*. 2021;1(1). Available from: <https://researchworkx.com/index.php/jci/article/view/2/2>
- [3] Alojail M, Alturki M, Bhatia Khan S. An informed decision support framework from a strategic perspective in the health sector. *Inf*. 2023; 14(7): 363.
<https://doi.org/10.3390/info14070363>
- [4] Pourrahimian E, Salhab D, Hamzeh F, AbouRizk S. A decision support system for evaluating construction project recovery plans. *Can J Civ Eng*. 2025.
<https://doi.org/10.1139/cjce-2024-0525>
- [5] Talari G, Cummins E, McNamara C, O'Brien J. State of the art review of Big Data and web-based Decision Support Systems (DSS) for food safety risk assessment with respect to climate change. *Trends Food Sci Technol*. 2022; 126: 192-204.
<https://doi.org/10.1016/j.tifs.2021.08.032>
- [6] Ekundayo F. Leveraging AI-driven decision intelligence for complex systems engineering. *Int J Res Publ Rev*. 2024; 5(11): 1-10.
<https://doi.org/10.55248/gengpi.5.1124.3343>
- [7] Kovari A. AI for decision support: Balancing accuracy, transparency, and trust across sectors. *Inf*. 2024; 15(11): 725.
<https://doi.org/10.3390/info15110725>
- [8] Al-Jumaili AHA, Muniyandi RC, Hasan MK, Paw JKS, Singh MJ. Big data analytics using cloud computing-based frameworks for power management systems: Status, constraints, and future recommendations. *Sensors*. 2023; 23(6): 2952.
<https://doi.org/10.3390/s23062952>
- [9] Agarwal G. Robust Data Pipelines for AI Workloads: Architectures, Challenges, and Future Directions. *Int J Adv Res Sci Commun Technol*. 2024; 5(2): 622-632.
<https://doi.org/10.48175/IJARSCT-23391>
- [10] Jayavel M. International Journal of Innovative Research in Computer and Communication Engineering. *Int J Innov Res Comput Commun Eng*. 2025; 13(3): 2039-2045.
<http://dx.doi.org/10.2139/ssrn.5162705>
- [11] Pillai V. Anomaly detection in financial and insurance data-systems. *J AI-Assisted Sci Discov*. 2024; 4(2): 144-183. Available from: https://www.researchgate.net/publication/386382514_Anomaly_Detection_in_Financial_and_Insurance_Data-Systems

- [12] Lee JH, You SJ. Balancing privacy and accuracy: Exploring the impact of data anonymization on deep learning models in computer vision. *IEEE Access*. 2024; 12: 8346-8358. <https://doi.org/10.1109/ACCESS.2024.3352146>
- [13] Altalhan M, Algarni A, Alouane MTH. Imbalanced data problem in machine learning: A review. *IEEE Access*. 2025; 13:13686-13699. <https://doi.org/10.1109/ACCESS.2025.3531662>
- [14] Hajek P, Abedin MZ, Sivarajah U. Fraud detection in mobile payment systems using an XGBoost-based framework. *Inf Syst Front*. 2023; 25(5): 1985-2003. <https://doi.org/10.1007/s10796-022-10346-6>
- [15] Ravichandran P, Machireddy JR, Rachakatla SK. AI-Enhanced data analytics for real-time business intelligence: Applications and challenges. *J AI Healthcare Med*. 2022; 2(2): 168-195. Available from: https://www.researchgate.net/publication/389139259_AI-Enhanced_Data_Analytics_for_Real-Time_Business_Intelligence_Applications_and_Challenges.
- [16] Rainy TA, Goswami D, Rabbi MS, Al Maruf A. A systematic review of ai-enhanced decision support tools in information systems: strategic applications in service-oriented enterprises and enterprise planning. *Rev Appl Sci Technol*. 2023; 2(01): 26-52. <https://doi.org/10.63125/73djw422>
- [17] Ashfaq T, Khalid R, Yahaya AS, Aslam S, Azar AT, Alsafari S, Hameed IA. A machine learning and blockchain-based efficient fraud detection mechanism. *Sensors*. 2022; 22(19): 7162. <https://doi.org/10.3390/s22197162>
- [18] Malik P, Chourasia A, Pandit R, Bawane S, Surana J. Credit risk assessment and fraud detection in financial transactions using machine learning. *J Electr Syst*. 2024; 20(3s): 2061-2069. <https://doi.org/10.52783/jes.1807>.
- [19] Olushola A, Mart J. Fraud detection using machine learning. *SciOpen Preprints*. 2024. <https://doi.org/10.14293/PR2199.000647.v1>
- [20] Bello OA, Ogundipe A, Mohammed D, Adebola F, Alonge OA. AI-driven approaches for real-time fraud detection in US financial transactions: Challenges and opportunities. *Eur J Comput Sci Inf Technol*. 2023; 11(6): 84-102. <https://doi.org/10.37745/ejcsit.2013/vol11n684102>
- [21] Alzaabi FR, Mehmood A. A review of recent advances, challenges, and opportunities in malicious insider threat detection using machine learning methods. *IEEE Access*. 2024; 12: 30907-30927. <https://doi.org/10.1109/ACCESS.2024.3369906>
- [22] Trigka M, Dritsas E. Improving cardiovascular disease prediction with deep learning and correlation-aware SMOTE. *IEEE Access*. 2025. <https://doi.org/10.1109/ACCESS.2025.3549417>
- [23] Khan D, Akram W, Ullah S. Enhancing landslide susceptibility predictions with XGBoost and SHAP: a data-driven explainable AI method. *Geocarto Int*. 2025; 40(1): 2514725. <https://doi.org/10.1080/10106049.2025.2514725>
- [24] Komati D. Real-Time AI Systems for Fraud Detection and Credit Risk Management: A Framework for Financial Institutions. *IJSAT-Int J Sci Technol*. 2025; 16(1): 1-13. <https://doi.org/10.71097/IJSAT.v16.i1.2974>
- [25] Teles G, Rodrigues JJ, Kozlov SA, Rabêlo RA, Albuquerque VHC. Decision support system on credit operation using linear and logistic regression. *Expert Syst*. 2021; 38(6): e12578. <https://doi.org/10.1111/exsy.12578>
- [26] Li M, Sun H, Huang Y, Chen H. Shapley value: from cooperative game to explainable artificial intelligence. *Auton Intell Syst*. 2024; 4(1): 2. <https://doi.org/10.1007/s43684-023-00060-8>
- [27] Ortigossa ES, Gonçalves T, Nonato LG. Explainable artificial intelligence (XAI) from theory to methods and applications. *IEEE Access*. 2024; 12: 80799-80846. <https://doi.org/10.1109/ACCESS.2024.3409843>
- [28] Favour AA. Regulatory compliance challenges in cloud-based ai fraud detection systems. 2022. Available from: https://www.researchgate.net/profile/Akintan-Favour/publication/391454540_Regulatory_Compliance_Challenges_in_Cloud-Based_AI_Fraud_Detection_Systems/links/6818ebbc60241d514022afe4/Regulatory-Compliance-Challenges-in-Cloud-Based-AI-Fraud-Detection-Systems.pdf (Accessed on: 19 September, 2022).
- [29] Pratama SF, Wahid AMA. Fraudulent transaction detection in online systems using random forest and gradient boosting. *J Cyber Law*. 2025; 1(1): 88-115. <https://doi.org/10.63913/jcl.v1i1.5>
- [30] Vishwakarma SK. AI-driven predictive risk modelling for aerospace supply chains. *Int Interdiscip Bus Econ Adv J*. 2025; 6(05): 102-134. <https://doi.org/10.55640/business/volume06issue05-06>
- [31] Dekker R. The importance of having data-sets. Available from: https://www.researchgate.net/publication/42581175_The_importance_of_having_data-sets (Accessed on Jan, 2006).
- [32] Singhal A, Srivastava J. Research dataset discovery from research publications using web context. *Web Intell*. 2017; 15(2): 81-99. <https://doi.org/10.3233/web-170354>