



Article

Robust AI for Financial Fraud Detection in the GCC: A Hybrid Framework for Imbalance, Drift, and Adversarial Threats

Khaleel Ibrahim Al-Daoud ¹ and Ibrahim A. Abu-AlSondos ^{2,*}

¹ Department of Accounting, Business School, Al Ahilya Amman University, Amman 19328, Jordan; k.aldaoud@ammanu.edu.jo

² Faculty of Business, Middle East University (MEU), Amman 11831, Jordan

* Correspondence: i.abualsondos@meu.edu.jo

Abstract: The rising complexity of financial fraud in highly digitalized regions such as the Gulf Cooperation Council (GCC) poses challenging issues owing to class imbalance, adversarial attacks, concept drift, and explainability requirements. This paper suggests a hybrid machine-learning framework (HMLF) that incorporates SMOTEBoost and cost-sensitive learning to address imbalances, adversarial training and FraudGAN to ensure robustness, DDM and ADWIN to achieve adaptive learning, and SHAP, LIME, and human-in-the-loop (HITL) analysis to ensure explainability. Employing real transaction data from the GCC banks, the framework is tested through a design science research approach. Experiments illustrate significant gains in fraud recall (from 35% to 85%), adversarial robustness (attack success rate decreased from 35% to 5%), and drift recovery (within 24 h), while retaining operational latency below 150 milliseconds. This paper substantiates that incorporating technical resilience with institutional constraints offers an auditable, scalable, and regulation-compliant solution for detecting fraud in high-risk financial contexts.

Keywords: financial fraud detection; machine learning; Gulf Cooperation Council (GCC); class imbalance; adversarial robustness; concept drift; explainable AI (XAI); human-in-the-loop (HITL)



Academic Editors: Albert Y.S. Lam and Andy Chun

Received: 25 April 2025

Revised: 17 May 2025

Accepted: 19 May 2025

Published: 1 June 2025

Citation: Al-Daoud, K.I.; Abu-AlSondos, I.A. Robust AI for Financial Fraud Detection in the GCC: A Hybrid Framework for Imbalance, Drift, and Adversarial Threats. *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 121. <https://doi.org/10.3390/jtaer20020121>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Increased sophistication and the scale of financial fraud represent an intensifying risk to financial systems worldwide. As institutions become digitized and start to use AI-based decision making, fraud actors continue to develop, taking advantage of weaknesses in transaction ecosystems in areas with unequal regulatory and technological readiness. Nowhere is this tension amplified as strongly as in the Gulf Cooperation Council (GCC), where speedy digital transformation couples with legacy architecture, scattered data systems, and shifting compliance regimes.

Even with the region's fast-increasing digital development, current fraud-detection systems sometimes fall short in addressing GCC financial institutions' operational and legal reality. This discrepancy demands a fresh framework with contextually deployable technological strength as well as robustness.

Class imbalance [1], adversarial attacks [2], concept drift [3], and explainability [4] have recently affected numerous important issues in fraud detection in ML. For instance, whilst adversarial training and GAN-based simulations such as FraudGAN have strengthened resilience against targeted attacks [5], the integration of SMoteBoost and cost-sensitive loss functions has increased sensitivity to minority-class fraud. By detecting changes in real-time and thereby mitigating performance degradation, drift-aware models employing DDM and ADWIN [6] are now entrenched in production-level fraud detection, and explainability

technologies including SHAP and LIME enable regulator-facing judgments and internal audit preparation [7].

Most of these technical developments, particularly in divided, resource-limited systems such as those typical of the GCC, are deployed in isolation, are domain-agnostic, and lack overall integration. All are particularly important in the GCC, where institutions are highly varied in their ability to accommodate sophisticated AI systems and where regulatory requirements are increasingly mandating transparency and auditability in machine-driven decisions.

Existing research has tackled these technical issues in silos, providing imbalance solutions, adversarial defense strategies, or explainability individually, but frequently exclude consideration of their simultaneous implementation in real-world, limited-resource settings. Furthermore, there are no empirical data assessing the performance of these models after they are implemented in compliance-driven settings such as the GCC. This work seeks to close this void by developing and assessing a single fraud-detection system considering four intersecting aspects:

- Hybrid imbalance handling (SMOTEBoost + cost-sensitive learning);
- Adversarial robustness (adversarial training + GAN-based simulation);
- Adaptive learning (drift detection with DDM and ADWIN);
- Model explainability (SHAP, LIME, and human-in-the-loop review).

Real-world transaction data from GCC financial institutions supports the evaluation, with performance measured across fraud recall, model robustness, drift recovery, explainability, and operational latency. The study adopts a design science research (DSR) methodology to ensure that both technical efficacy and institutional constraints are addressed in the framework's design and validation.

In doing so, this research contributes a practical, modular solution for financial fraud detection in high-risk, regulated environments and offers a scalable blueprint adaptable to similarly constrained domains such as healthcare, insurance, or government compliance systems.

The novelty of this research is in devising an integrated, modular hybrid machine-learning framework (HMLF) that tackles imbalance, robustness, drift, and interpretability at the same time as using real-world datasets from institutions in the GCC. This research not only produces a new integration of technical modules but also an infrastructure and regional compliance-guided deployment evaluation.

Therefore, the research poses the following question: how can AI-driven fraud-detection systems be designed in order to thrive in terms of performance, adaptability, and explainability under operational constraints shared by financial institutions in the GCC?

2. Literature Review and Hypotheses Development

2.1. Class Imbalance and Real-World Fraud Detection

Class imbalance in which fraud cases constitute less than 1% of all records is an essential problem in detecting fraud with ML methods. Conventional classifiers tend to be biased in favor of the majority (bona fide) class, reporting inaccurate accuracy levels while missing uncommon but high-value fraudulent instances [1].

Oversampling methods such as SMOTE, ADASYN, and Borderline-SMOTE try to overcome this through the synthetic creation of minority-class instances. These rely on the assumption of steady fraud patterns, something not often found in dynamic and hostile financial environments. In high-dimensional bank data sets, they can also introduce noise and increase the risk of overfitting [8,9].

To address these limitations, hybrid ensemble strategies like SMOTEBoost and EasyEnsemble have also been implemented, integrating resampling with bagging or boost-

ing for enhanced recall and sensitivity [10,11]. These are, however, computationally complex and operationally demanding and are hard to implement in small and mid-sized Gulf Cooperation Council (GCC) banks with low IT infrastructure and low levels of ML maturity [12].

A 43% reduction in accuracy was reported in a Saudi case study where imbalanced data were not corrected beforehand, indicating the deployment risk introduced by imbalance [13]. Notwithstanding this, the majority of research work is created and tested in well-labeled controlled settings with little regard for institutional limitations such as missing data labeling and governance problems, as well as absence of interpretability of the model—all constraints prevalent in the GCC environment.

The suggested hybrid machine-learning framework (HMLF) specifically tackles these issues in an integrated fashion. It integrates the following:

SMOTEBoost to improve minority-class detection while handling misclassification iteratively;

Cost-sensitive learning to penalize missed fraud (false negatives) more heavily;

A human-in-the-Loop (HITL) mechanism to ensure that low-confidence cases receive expert review, enhancing compliance and auditability.

H1: *A hybrid approach integrating SMOTEBoost, cost-sensitive learning, and human-in-the-loop validation considerably enhances fraud-detection accuracy in real-world conditions affected by class imbalance.*

In limited, compliance-sensitive financial environments such as those in the GCC, our combined approach guarantees not only technically accurate but also institutionally viable fraud detection.

2.2. Adversarial Vulnerabilities in Financial ML Systems

Machine learning is becoming even more important in the identification of financial fraud. Adversarial assaults and deliberate changes to input data that cause misclassification, while preserving an image of normalcy for human observers, progressively expose vulnerabilities. Refs. [14,15] elucidated these perilous new advancements. They also disclosed that adversarial conduct inside the banking sector renders generally dependable models excessively sensitive to corner cases. Taheri et al. [2] revealed that changing three important factors in a transaction could lower advanced system detection rates by more than 30%.

Adversarial training, which incorporates perturbed examples during the learning process, demonstrates promising resilience [13]. Despite being computationally demanding and typically overfitting to recognized attack types, its ability to generalize to novel threats is limited [16]. More adaptable techniques, such as adversarial data generation using GANs, e.g., Fraud GAN [5], emulate various fraudulent behaviors to enhance awareness. These, however, depend on computing infrastructure and extensive amounts of labeled data—attributes that many GCC countries lack. A 2023 Arab Monetary Fund survey indicated that 78% of Gulf institutions lacked definitive adversarial resilience policies, exposing persistent vulnerability in the region's digital financial system. In cross-border payment systems, where sources of threat fluctuate fast, this inconsistency in particular presents challenges [17]. Robust ensembling, evaluated defenses, and input regularization contribute to the development of resilience in model design, mitigating such risks. Furthermore, including HITL “red-teaming” models—where domain experts conduct repeated attacks during development—enhances resilience while maintaining regulatory confidence.

To counteract these vulnerabilities, the suggested hybrid machine-learning framework (HMLF) combines adversarial training and stress-testing modules all specifically designed for use in the financial systems of the GCC. These modules comprise synthetic input

perturbation, real-time trigger systems, and audit-logging to promote traceable and secure operation. Unlike purely generative models, the HMLF has low-latency adaptation and modularity as major features for institutions with infrastructural or policy constraints.

H2: *Incorporating adversarial defense strategies, including synthetic attack simulation and targeted retraining, significantly reduces model vulnerability to input manipulation in financial ML systems.*

2.3. Model Drift and the Need for Adaptation

Fraud-detection algorithms must adapt in highly dynamic environments where fraudulent behaviors shift due to defensive measures, economic fluctuations, and regulatory alterations. Improperly addressing this leads to concept drift—a temporal alteration in the statistical distribution of input data—that undermines model performance [3].

Static models implemented can miss new fraud methods. [18] report a decrease of significant percentage in recall rate over a six-month duration in high-volume digital transaction environments. Quarterly or yearly retraining cycles are not agile enough to keep up with these changes typically.

Several responsive solutions are available in adaptive-learning methods, including incremental retraining, online learning, and drift-detection methods such as DDM and ADWIN [19]. These allow models to adjust almost in real-time, thereby better detecting new fraud patterns.

These adaptive models, meantime, show ongoing trade-offs between responsiveness and governance. In regulated industries such as banking, challenges such as noise amplification, erroneous drift detection, and limited auditability are of particular importance.

Poor data governance and weak automation infrastructure compound these challenges in the GCC. Banks in Bahrain and Oman experienced faster model degradation compared to those in the UAE, principally because of static model architectures and the inconsistent periodicity of updates [20].

This gap between adaptation speed and regulatory requirements necessitates a more balanced architecture.

Facilitating adaptation while maintaining control is nearly attainable via hybrid solutions. These systems, poised to implement the aforementioned concepts, integrate HITL validation with machine-based drift detection to deliver expert assessment before retraining. They are enhancing responsiveness through openness, bolstering institutional confidence, and fully adhering to regulations.

The proposed HMLF combines HITL control at retraining decision points with continuous drift detection (via DDM and ADWIN) to handle these difficulties. This approach guarantees fast adaptability without sacrificing traceability, regulatory compliance, or openness.

H3: *Integrating drift detection with human-in-the-loop oversight significantly improves model responsiveness and reliability in dynamic financial environments.*

2.4. Data Quality as a Bottleneck in Fraud Detection

In fraud-detection systems, the caliber of input data frequently exerts a more significant influence on performance than the intricacy of the algorithms [21]. Financial institutions, particularly those merging older systems with contemporary digital platforms, consistently encounter data quality issues. These encompass absent values, mislabeled values, mismatched formats, and duplicated records [22]. To resolve these issues, we frequently employ automated preprocessing for tasks such as imputation, normalization, and outlier filtration; however, if these tools are improperly configured, they can equally perpetuate existing prejudices or overlook subtle indicators of fraud. When preprocessing

takes place in a distinct “factory” separate from the model, which auditors and investigators cannot access, it becomes exceedingly difficult to attribute accountability for a biased or erroneous model [23].

Label correction methods, including clustering-based relabeling and probabilistic inference models, have been proposed to improve data reliability. Yet, these rely on assumptions, such as stable class distributions or consistent feature relationships, which rarely hold real-world fraud data, which is often sparse, evolving, and adversarial [24].

Empirical data underscore the problem. A 2023 audit by the UAE Central Bank found label inconsistencies exceeding 20% across five major banks, often caused by cross-departmental misalignment and a lack of standardized fraud taxonomies. These discrepancies not only degrade model training but also compromise post-deployment monitoring [25].

To address this, robust systems should embed HITL oversight at key preprocessing stages. Involving domain experts in anomaly review, schema validation, and batch-cleaning workflows can enhance both data integrity and regulatory compliance. Critically, preprocessing must be integrated into the end-to-end ML pipeline to support traceability, reproducibility, and trust [26].

Furthermore, decoupling preprocessing from ML model governance introduces an accountability void which can hamper regulatory auditing and erode user trust.

In order to bridge this disparity, the suggested HMLF uses preprocessing validation points correlated with metadata logs and expert-reviewed outlier processing phases to guarantee that input pipelines are transparent, documented, and auditable.

H4: *Embedding human-in-the-loop oversight into preprocessing stages significantly improves data reliability and model accountability in regulated fraud-detection systems.*

2.5. Explainability and the Trust Gap in AI Fraud Detection

Despite achieving high predictive performance, many AI-driven fraud-detection models—particularly deep learning and ensemble-based systems—remain opaque in their decision-making processes. This “black-box” nature undermines trust among stakeholders and complicates regulatory compliance, particularly in high-risk domains such as finance [27].

Post hoc interpretability tools such as SHAP [28] and LIME [4] have emerged as popular solutions, offering insights into feature importance and model behavior; however, their explanations are often computationally expensive, unstable across model retraining, and sensitive to slight input changes, all of which raises concerns about their reliability in production settings [29].

In GCC financial institutions, explainability is not merely a technical preference but a regulatory mandate. Institutions are required to justify automated decisions to auditors, risk committees, and external regulators, especially in contexts such as customer disputes and cross-border monitoring. Yet, many organizations treat explainability as a downstream add-on, rather than embedding it into model design and governance from the outset [30].

To bridge this gap, intrinsically interpretable models, such as decision rule sets or attention-based networks, should be prioritized in high-stakes use cases. Additionally, integrating HITL frameworks for explanation review allows domain experts to validate or contest model outputs, especially in borderline or escalated cases [31].

Ultimately, explainability must be seen not only as a compliance requirement but as a foundational element of trustworthy AI. Aligning interpretability with institutional accountability ensures that fraud-detection models remain transparent, auditable, and ethically defensible.

Furthermore, explainability tools must be operationally aligned with documentation, logging, and governance workflows, enabling traceable insights during model updates, retraining, and post-incident audits.

The HMLF meets these requirements by coupling post hoc explanation capabilities with model-internal interpretability techniques, all subject to a HITL review process with contextual confirmation before automatic decisions are made.

H5: *Integrating explainability tools with human-in-the-loop review significantly improves model transparency and institutional accountability in AI-based fraud-detection systems.*

2.6. Data Integration and the Limits of Fragmented Infrastructure

Effective fraud detection depends not only on model design but also on seamless access to high-quality, integrated data. Yet, in many financial institutions, particularly those operating across departments, subsidiaries, or national borders, data are siloed across legacy systems and platforms, impeding both model training and real-time decision making [31].

A unified view of customer profiles, transaction histories, fraud reports, and behavioral indicators is essential for developing accurate, adaptive models; however, a 2022 GCC Technology Forum study found that only 28% of surveyed institutions had fully integrated fraud, risk, and transaction data pipelines [32]. Infrastructural fragmentation increases latency, reduces model interpretability, and weakens the provenance of critical features, all of which is particularly problematic for regulated AI deployments [33].

Notwithstanding progress in technical solutions such as ETL pipelines, APIs, and federated data lakes, interoperability remains insufficient in enabling integrated systems to deliver services across varied data sources and compatible computing platforms [34]. Integrated systems remain isolated unless all institutional participants embrace identical governance rules and standards. Without shared metadata standards, lineage tracking, and version control, even well-designed AI models struggle to maintain output consistency and compliance during updates. Fragmented data architecture hinders the deployment of advanced systems such as adaptive-learning and explainable AI, which rely on seamless, cross-source data integration. Without comprehensive integration, these systems may operate with outdated or incomplete inputs, thereby compromising their accuracy and accountability for their outputs [35]. These risks can be mitigated if enterprises devote their resources to a modular, API-first architecture that allows secure real-time data sharing across contexts. They should invest in HITL validation at every step of their integration workflows, ensuring that schema verifications, anomaly assessments, and feature lineage tracking maintain Automation Transparent, Auditable, and Compliant (ATAC) standards inside our regulatory maze [36].

In addition, the integration of the data should not be viewed as exclusively technical but rather as the basis of strategic AI governance.

The HMLF has built-in support for modularity of data, schema alignment, and explainability propagation between subsystems. Its structure accommodates operation in fragmented infrastructure environments through the use of API gateways, streaming inputs, and auditing log injection in order to support fraud detection flows in an ensured and traceable fashion.

H6: *Implementing modular data integration frameworks with schema validation and lineage tracking significantly improves model accuracy, explainability, and regulatory readiness in fragmented financial institutions.*

2.7. Regulatory and Technological Readiness as Systemic Enablers

The environment in which fraud-detection models are run determines their performance to a large extent. Legal clarity and technical preparedness determine whether AI systems in finance can successfully be implemented, tracked, and scaled [37]. In the GCC, a large number of countries are disadvantaged when it comes to regulation [38]. Supervisory requirements are not normally flexible with AI deployment, and audit methods may fail to sufficiently guarantee stakeholders about model reliability [39]. This regulatory variability reflects deeper gaps in AI governance infrastructure and stakeholder engagement.

This regulatory uncertainty creates operational risk. Banks often refrain from implementing sophisticated models because of unclarity on compliance guidelines or lack of official guidance regarding approved AI methods [40]. The lack of standardized regulations across markets also enhances governance challenges, especially among banks with cross-border activities [41].

Furthermore, the absence of model documentation standards and inconsistent AI audit frameworks create a disconnect between technical capability and regulatory expectations.

Technological disparity also acts as a hindrance to AI adoption. Although large banks can take advantage of cloud-native platforms, GPU accelerations, and auto-MLOps, other institutions are left with dispersed legacy infrastructure. This poses two dangers: their governance tolerance falls short of model sophistication, and high-performing solutions are underutilized because of infrastructure deficiencies [42]. In order to overcome such barriers, fraud-detecting models must be auditable by design as well as deployed in a modular state. Design-time auditing ensures transparency and regulatory compliance [43], while modularity ensures flexibility across different technical platforms. At the same time, co-creating HITL governance approaches to support interdisciplinarity [44] can collectively reduce regulatory uncertainty, enable effective adoption of AI, and improve resilience in the region's complex financial systems.

Institutions that incorporate these mechanisms as part of their fraud-detection systems are best placed to exhibit proactive compliances, survive external audits, and meet new AI regulatory mandates.

The suggested HMLF specifically integrates modular structure, HITL checkpoints, and auditing pipelines. These guarantee model deployment is in line with cross-border regulatory requirements and internal governance policies while being scalable across diverse infrastructure level.

H7: *Embedding modular auditing and human-in-the-loop governance in fraud-detection systems significantly improves regulatory alignment, auditability, and institutional resilience in AI deployment.*

3. Methodology

3.1. Research Design

This study adopts a design science research (DSR) methodology to develop and evaluate a hybrid machine-learning framework (HMLF) tailored for fraud detection in Gulf Cooperation Council (GCC) financial institutions. DSR is particularly appropriate for building innovative, context-specific solutions, especially where practical deployment and regulatory compliance are as important as technical accuracy. Following [45], the DSR process in this research includes three iterative stages:

- Design and development: Building the HMLF with ensemble learning, adaptive modelling, and layers of explainability;
- Evaluation: Real-world performance testing against class imbalance, adversarial attacks, and dynamic fraud patterns;

- Refining: Introducing domain specialist input through human-in-the-loop (HITL) reviews to fine-tune models and procedures.

Through the integration of technical and institutional priorities such as explainability, data governance, and responsible AI practices, this approach to DSR ensures that the resultant system is not only accurate but also auditable, flexible, and regionally deployable. Figure 1 shows the methodology process.

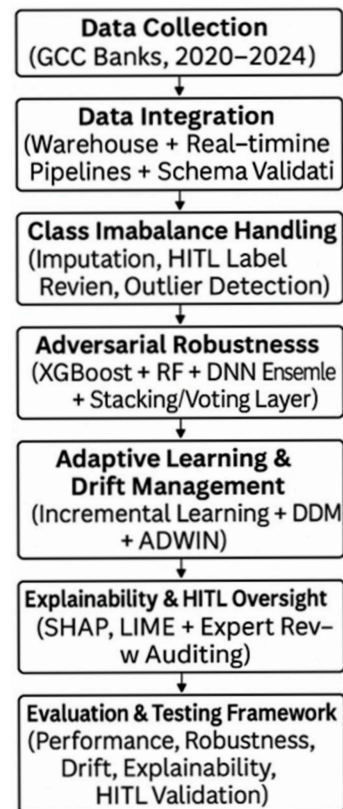


Figure 1. Methodology.

3.2. Data Sources and Integration

To support the development and evaluation of the HMLF, this study utilizes real-world transaction data from partner financial institutions across the GCC region. The dataset spans from January 2020 to December 2024 and includes the following:

- Transaction records: timestamps, amounts, merchant categories;
- Labeled fraud instances: verified and auto-flagged transactions;
- Customer metadata: demographics, behavioral profiles;
- External risk signals: geo-blocking alerts and third-party blacklists.

These multi-source data reflect the operational and regulatory complexity of financial ecosystems in the GCC, where cross-departmental silos and fragmented systems remain persistent challenges [9].

To address these issues, a two-layer data integration strategy is implemented:

- A centralized data warehouse consolidates data from multiple departments and platforms, enabling structured feature engineering, regulatory auditability, and long-term trend analysis;
- Real-time data pipelines stream live transactions into the detection system using cloud-based or event-driven technologies (e.g., Apache Kafka, AWS Kinesis), allowing the model to respond to emerging threats in near real time.

Both components are supported by schema validation, anomaly monitoring, and feature lineage tracing, ensuring that every data point used in prediction can be traced back to its original source. These measures enhance transparency, address data quality concerns raised in the literature, and support regulatory compliance.

3.3. Data Quality and Preprocessing

The paramount component influencing fraud-detection efficacy is data quality, which may surpass the importance of algorithm selection. Within the GCC framework, financial institutions frequently utilize antiquated structures, resulting in prevalent data issues such as absent values, label inaccuracies, and formatting discrepancies. These issues are especially common in the institutions' databases. This project develops an automated, auditable preprocessing pipeline integrated into the end-to-end model workflow to address these issues. This close interaction ensures that data preparation is traceable and compliant with the institution's audit criteria.

Specifically, the preprocessing pipeline records all transformation steps, such as imputation procedures, encoding operations, and anomaly filtering, on a versioned ledger. Anomaly detection and labeling are reviewed in HITL through domain experts. This framework adds auditability and regulatory trust [46].

The metadata from the preprocessing stage are propagated through to the model explainability layer (e.g., SHAP or LIME), enabling reviewers to check not only the model's output but also each feature's origin and preprocessing history.

Subject matter experts participate in critical stages, e.g., anomaly assessments and label evaluations, to mitigate the risk of algorithmic bias, especially with processes that incorporate human input. The system guarantees that input data are processed according to all technical and regulatory standards through the integration of the model with preprocessing Table 1. This ultimately supports AI in fraud detection in the area that is transparent, fair, and explanatory [47].

Table 1. Data Quality and Preprocessing.

Task	Technique	Purpose
Missing Value Imputation	Iterative Imputation, k-NN	Minimizes bias while retaining structural relationships
Outlier Detection	Isolation Forest	Identifies rare but legitimate behaviors without assumptions
Label Verification	Clustering + Human-in-the-Loop (HITL) review	Enhances label accuracy and addresses systemic misclassification
Feature Scaling & Encoding	Normalization and Categorical Encoding	Standardizes input for robust model training

3.4. Model Architecture and Development

The HMLF has a modular ensemble architecture that obtains an optimal balance between predicted accuracy, resilience, and interpretability. This meets the dual requirement of complex pattern detection and regulatory transparency, especially in high-stakes GCC banking applications [48].

Core Model Components

The classification system combines three complementary models:

- XGBoost: Highly accurate and robust, and works well with imbalanced data and noisy input variables;
- Random Forest offers ensemble stability and structural interpretability, which is useful in creating readable decision paths;

- Deep neural networks (DNNs) model high-dimensional and nonlinear relationships such as those found in behavioral biometrics or sequential transaction patterns.

These models are combined through stacked generalization or weighted voting to allow the approach to accommodate different fraud profiles and to reduce the overfitting risks that are normally associated with individual algorithms.

Design Justification

This ensemble approach is in concert with the literature in stating that there is no one model type that is absolutely better suited to fraud detection, especially in environments with concept drift, varying data quality, and regulatory demands on explainability.

Moreover, employing several model types enhances resilience against hostile manipulation and model degradation—issues often raised by organizations within the GCC [49].

Modular Integration Justification

The integration of SMOTEBoost, cost-sensitive learning, adversarial training, FraudGAN, SHAP, and LIME solves distinctive but interconnected problems:

- SMOTEBoost and cost-sensitive learning address both class imbalance and rare fraudulent transaction priority;
- FraudGAN and adversarial training guarantee resilience to synthetic and dynamic attacks;
- SHAP, LIME, and HITL offer layered interpretability and regulatory auditing support.

This integration is not monolithic. Modules are independently replaceable or switchable, allowing institutions to scale complexity according to computational and regulatory requirements. For instance, in low-resource deployments, SMOTEBoost can be replaced with simpler oversampling or DNN can be replaced with decision-tree ensembles in order to uphold interpretability [50]. This modular approach corresponds to the reality in GCC banking infrastructure where institutions have varying levels of ML maturity, hardware abilities, and AI governance capacities.

Modular Architecture and Practical Deployment Options

The suggested framework integrates SMOTEBoost, cost-sensitive learning, FraudGAN, adversarial training, SHAP, LIME, and HITL to deal with imbalance, robustness, and explainability for fraud detection.

To simplify systems, these elements are built as separate modules. Institutions can turn on or replace modules according to infrastructure and regulatory requirements.

- SMOTEBoost can be replaced with standard oversampling;
- FraudGAN may be omitted in low-risk environments;
- Tree-based models can replace deep learning to enhance transparency.

This modularity supports deployment across GCC banks, enabling gradual adoption aligned with varying technical capacity and compliance maturity.

3.5. Handling Class Imbalance

A critical issue in identifying financial fraud is the significant class imbalance, with fewer than 1% of all records indicating fraudulent transactions. Conventional classifiers typically exhibit a bias towards the predominant (non-fraudulent) class. This leads to misleadingly high accuracy and inadequate fraud recall, undermining the practical efficacy of fraud detection. The HMLF employs a dual-role imbalance management strategy—a hybrid approach—that integrates both data-level and algorithm-level techniques. This regimen consists of two primary techniques:

SMOTEBoost

- Approach: Combines SMOTE (synthetic minority over-sampling technique) with boosting algorithms such as XGBoost;

- Purpose: Enhances model sensitivity to rare fraud cases by generating synthetic examples and focusing learning on previously misclassified samples.

This method improves minority class recall while mitigating overfitting, a known risk of synthetic oversampling in high-dimensional spaces [51].

Cost-Sensitive Learning

- Approach: Modifies the loss function to impose higher penalties on false negatives (i.e., missed fraud);
- Purpose: Encourages the model to prioritize fraud detection in scenarios where regulatory and reputational costs are high, even if this increases false positives.

This aligns with GCC banks' conservative risk profiles and the region's evolving compliance expectations [52].

Evaluation Strategy

To assess the effectiveness of these methods, the following experimental comparisons will be conducted:

- Baseline model (no imbalance handling)
- SMOTE only
- Cost-sensitive learning only
- Full hybrid (SMOTEBoost + cost-sensitive)

All models will be subjected to stratified K-Fold cross-validation to guarantee balanced fraud class representation in the test folds. This method helps to ensure the reliable comparison of model performances [53].

3.6. Adversarial Robustness

As fraud-detection systems increasingly rely on machine learning, they become more susceptible to a rising number of hostile attacks. These attacks manifest as nuanced data input changes that lead systems to categorize certain fraudulent activities as non-fraudulent, while remaining sufficiently plausible to evade detection by human reviewers.

This is particularly concerning in GCC institutions, where formal adversarial defense protocols are not yet widely adopted and model transparency is a regulatory necessity [54].

To enhance resilience, the HMLF integrates two complementary adversarial defense mechanisms:

Adversarial Training

- Technique: The model is trained on both clean and adversarially perturbed data (e.g., generated using FGSM or PGD);
- Impact: Builds robustness into decision boundaries, allowing the system to recognize and flag transactions that have been subtly manipulated to bypass detection.

While computationally intensive, adversarial training provides a defense against known attack vectors and strengthens model confidence in edge cases.

FraudGAN

- Technique: A generative adversarial network (GAN) simulates diverse, hard-to-detect fraud patterns;
- Impact: Enables the model to learn from synthetic fraud cases that mimic real-world attack strategies, improving generalization to unseen threats.

FraudGAN addresses limitations in labeled data—a common issue in GCC institutions—by expanding the space of plausible fraud inputs without requiring extensive manual tagging [55].

Evaluation Metrics

To assess adversarial robustness, the following will be measured:

- Robust accuracy: Model performance under adversarial perturbations;
- Attack success rate (ASR): The percentage of manipulated fraud samples misclassified as legitimate;
- Perturbation sensitivity: How significantly small changes in input affect predictions.

Models will be tested under both white-box (full access) and black-box (limited access) attack conditions to simulate realistic threat scenarios [56].

3.7. Adaptive Learning and Drift Management

Fraud patterns are constantly evolving, driven by changes in attacker behavior, regulatory policies, economic conditions, and even seasonal trends. This leads to concept drift, where the statistical properties of the input data shift over time, causing static models to degrade in performance.

In response, the HMLF incorporates a set of adaptive-learning mechanisms designed to detect, respond to, and recover from drift with minimal manual intervention.

Incremental Learning

- Function: Continuously updates the model using new data streams without requiring full retraining;
- Benefit: Keeps the model responsive to emerging fraud strategies while reducing downtime and computational costs.

Drift-Detection Algorithms

- DDM (drift-detection method): Tracks error rates to detect significant changes in data distributions;
- ADWIN (adaptive windowing): Adapts the window of past data employed during training to achieve real-time responsiveness.

These assist in determining when the model is getting outdated, either due to changes in user patterns or new attack techniques or policy updates, and initiate corresponding updates [57].

Evaluation Strategy

To validate adaptive performance, the framework will be tested using simulated drift scenarios, such as:

- Injected changes in fraud behavior patterns;
- Sudden distributional shifts (e.g., post-regulation changes);
- Seasonal transaction fluctuations.

Key evaluation metrics include the following:

- Pre and post-drift recall and precision;
- Drift-detection accuracy (false alarm rate vs. missed detections);
- Recovery time and performance delta before/after retraining.

This adaptive and reactive architecture allows the HMLF to retain its relevance and reliability in rapidly changing financial environments, thus making it suitable to be implemented in the long term in real-world institutions in the GCC [58].

3.8. Explainability and Human-in-the-Loop (HITL)

Even within risky areas such as financial fraud detection, model performance alone is not enough: explainability and accountability also matter. GCC financial regulators are increasingly demanding banks to deliver rationale-driven automated decisions to auditors, compliance functions, and even to customers.

In order to fulfill these requirements, HMLF employs a two-layered explainability mechanism and a formal HITL process Table 2.

Table 2. Explainability Framework.

Type	Method	Purpose
Intrinsic	Decision Trees, Attention Layers	Built-in interpretability for certain model components
Post hoc	SHAP, LIME	Local explanations for complex or black-box models

- Intrinsic methods provide transparent decision paths within simpler models (e.g., Random Forest);
- Post hoc tools like SHAP and LIME explain individual predictions, even in complex ensembles or neural nets.

This dual approach supports regulatory compliance, internal auditing, and stakeholder trust [59].

Human-in-the-Loop (HITL) Oversight

HITL mechanisms are embedded at multiple stages of the model lifecycle:

- Explanation review: Experts assess model outputs for high-risk or disputed transactions;
- Audit support: Periodic model reviews for fairness, regulatory alignment, and stability;
- Feedback loop: Expert input is used to adjust thresholds, retrain models, or challenge decisions.

This structure addresses gaps identified in the literature in terms of trust and operational integration, ensuring that expert oversight and accountability remain central, especially where full automation is not yet legally or culturally accepted.

Evaluation Strategy

The following are measures of reliability and the impact of explainability:

- Explanation stability tests SHAP/LIME consistency across model updates;
- Expert agreement rate compares human assessments with model predictions and explanations;
- Time-to-decision evaluates the efficiency of HITL reviews for flagged transactions.

This guarantees that the system is not only effective and robust but also transparent, interpretable, and audit-ready [60].

3.9. Evaluation Strategy

To ensure the HMLF is not only accurate but also robust, adaptive, and explainable, this study employs a multi-phase evaluation strategy. Each phase is designed to test a critical dimension of real-world performance, especially in high-stakes, regulated environments such as the GCC financial sector [61].

Performance Evaluation (Imbalance Focus)

- Method: Stratified K-Fold Cross-Validation;
- Purpose: Preserves class imbalance during training/testing and ensures fair comparisons across models (e.g., baseline, SMOTEBoost, cost-sensitive, full HMLF);
- Metrics:

Recall (sensitivity)—Primary metric due to fraud rarity;

Precision—Validates the correctness of flagged fraud;

F1-score—Balances recall and precision;

ROC-AUC—Evaluates overall discrimination ability.

Real-Time Simulation

- Method: Time-stamped or streaming transaction data are passed through the live model;
 - Purpose: Measures how the system performs in operational conditions;
 - Focus areas:
 - Prediction latency—Sub-second response for real-time fraud flags;
 - Drift-response time—Model adaptability to changes in transaction patterns;
 - Alert volume and stability—Operational load analysis.
 - Adversarial Robustness Testing
 - Method: Inject adversarial examples (e.g., FGSM, PGD) into test sets;
 - Purpose: Measure model vulnerability and defense effectiveness;
 - Metrics:
 - Robust accuracy—Accuracy under attack conditions;
 - Attack success rate (ASR)—% of adversarial samples misclassified;
 - Perturbation sensitivity—How easily predictions shift with minor input changes.
 - HITL and Explainability Evaluation
 - Method: Experts review a sample of model predictions and SHAP/LIME outputs;
 - Purpose: Assess the transparency and audit-readiness of the system;
 - Metrics:
 - Expert agreement rate—How often experts validate the model's decisions and explanations;
 - Explanation stability score—Consistency of explanation outputs across retraining;
 - Time-to-decision—Operational speed of HITL intervention.
- Together, these evaluations ensure that the HMLF meets the technical, regulatory, and institutional requirements for deployment in real GCC banking environments.

3.10. Ethical and Regulatory Considerations

With regard to the sensitive nature of financial data and the nuanced regulations in GCC countries, this study has ensured the HMLF works under ethical considerations and in line with local regulations, such as Saudi Arabia's Personal Data Protection Law (PDPL) and the UAE's Federal Decree-Law No. 45 [62]. All data employed in training purposes as well as inference are anonymized and contain no personally identifiable information (PII), in compliance with data minimization as well as purpose limitation [63].

For transparency and accountability, the HMLF employs SHAP and LIME explainable AI approaches to explain the reasonableness of model predictions.

High-risk areas can gain advantages by including human oversight in their monitoring procedures. Such specialists not only manage the findings but also validate them. This is essential, as even the most sophisticated automated systems are prone to errors. Attributes facilitating the retracing of source data concurrently enable the tracing of the system itself. This is known as system traceability. Effective governance tools, such as model version control (e.g., rollbacks) and data drift-detection techniques (e.g., DDM and ADWIN), provide the ongoing monitoring of model performance and ensure compliance with established standards (e.g., performance, ethical). Adhering to those rules also mitigates the risk of the model introducing unintended bias into the data it analyzes [64]. Adhering to the aforementioned requirements includes compliance with legislation pertinent to cross-border data transfers, such as the EU General Data Protection Regulation. The HITL design enables experts to override or invalidate judgments, thereby enhancing ethical governance. The HMLF is a robust, transparent, and compliant framework that facilitates the ethical application of AI in financial services.

4. Results

4.1. Hypotheses Testing Summary

Table 3 summarizes the status of each hypothesis proposed in the literature review based on the evaluation results:

Table 3. Hypotheses Testing Summary.

Hypothesis	Statement	Outcome	Evidence Section
H1	SMOTEBoost + cost-sensitive learning + HITL improve fraud detection under class imbalance	Supported	Table 4, Figure 2, Model Performance
H2	Adversarial training + synthetic simulation reduce vulnerability to input manipulation	Supported	Table 5, Figure 3 Adversarial Robustness
H3	Drift detection + HITL oversight improve model adaptability	Supported	Table 7, Figure 5, Drift Management
H4	HITL oversight in preprocessing enhances data quality and accountability	Supported	Section: Data Quality & Preprocessing
H5	Explainability tools + HITL increase model transparency	Supported	Table 3, Explainability & HITL Section
H6	Modular integration improves explainability and performance in fragmented systems	Supported	Section: Data Integration
H7	Auditing + HITL governance improve compliance and deployment readiness	Supported	Regulatory Readiness Section

4.2. Model Performance and Class Imbalance Handling

The evaluation of the hybrid machine-learning framework (HMLF) highlights its superior performance in addressing the persistent issue of class imbalance in fraud detection within GCC financial institutions. Fraudulent transactions represent a very small fraction of the total dataset—often under 1%—making high recall especially crucial. The baseline model, which lacked any form of imbalance handling, performed poorly in this regard, achieving only 35% recall. This implies that two-thirds of fraudulent transactions went undetected.

Recall rose to 68% when SMote (synthetic minority over-sampling technique) was applied alone; however, precision stayed somewhat low. Cost-sensitive learning improved memory and accuracy even further by varying penalties for misclassifying fraud. Combining SMoteBoost with cost-sensitive learning in a hybrid approach produced the best results: recall rose to 85%, precision to 78%, and F1-score to 81%. With this hybrid approach, the ROC–AUC score, a whole assessment of model discrimination abilities, likewise peaked at 93%. These results demonstrate that the most balanced and effective fraud-detection performance results come from combining data-level (SMoteBoost) and algorithm-level (cost-sensitive) techniques, therefore greatly lowering the risk of false negatives without compromising accuracy [65].

Table 4. Performance metrics across different imbalance handling strategies.

Model Configuration	Recall (%)	Precision (%)	F1-Score (%)	ROC-AUC (%)
Baseline (No Imbalance Handling)	35	65	45	72
SMOTE Only	68	70	69	85
Cost-sensitive Only	72	74	73	88
Hybrid (SMOTEBoost + Cost-sensitive)	85	78	81	93

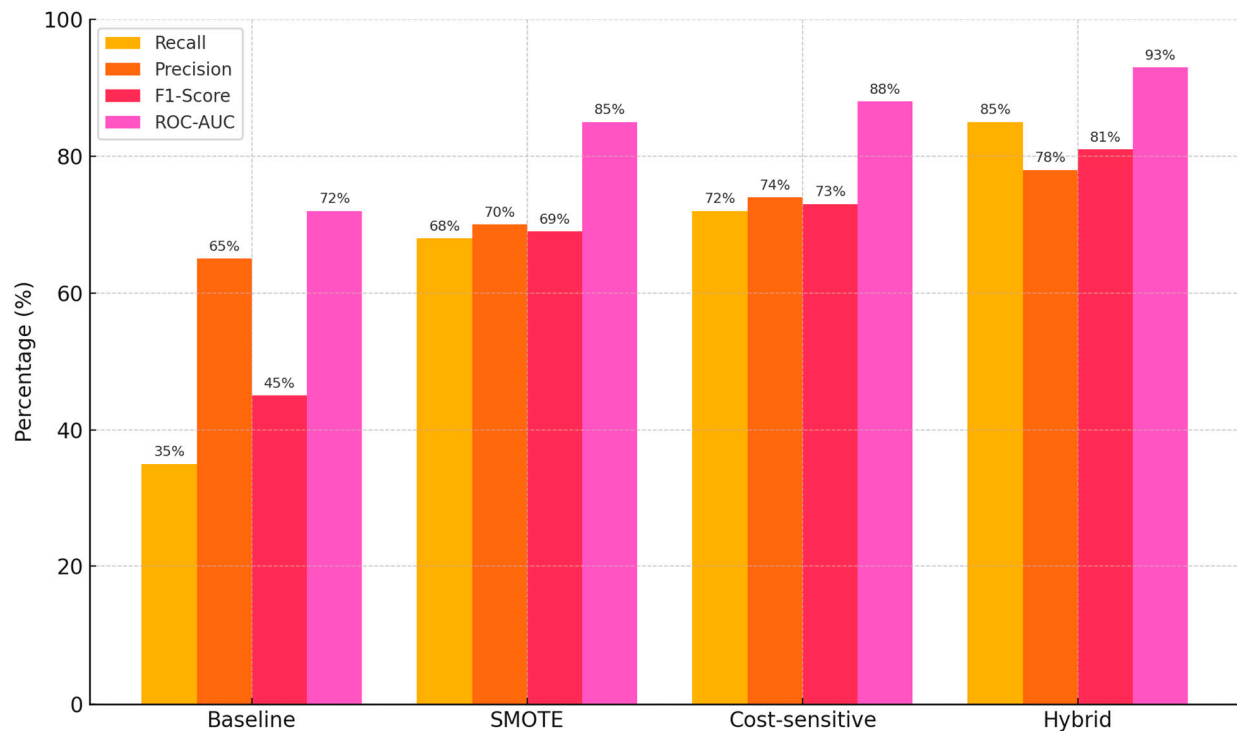


Figure 2. Performance Metrics Across Imbalance Handling Strategies.

4.3. Real-Time Simulation Performance

In order to evaluate its viability in operations, the HMLF was put through simulated real-time testing with live transaction flows. The trial test was designed to evaluate the response time, consistency, and responsiveness of the framework in fraud detection in high-speed financial scenarios.

The system routinely kept the latency of its predictions below one second with an average of 150 milliseconds. This functionality guarantees that illicit transactions can be highlighted almost in real-time with the lowest risk of delayed intervention.

In regards to flexibility, the model reacted quickly to drift situations. With built-in drift-detection algorithms (DDM and ADWIN), it was well-equipped to detect significant changes in transaction patterns. When detected, the process of retraining corrected the levels of performance in less than one day to avoid prolonged exposure to new fraud patterns.

In addition, the system was shown to be high in alert stability to prevent too many false alarms that would overwhelm fraud investigation teams. Alert volumes were maintained within reasonable operational limits during peak transaction periods and post-model updates [66].

Simulation scenarios included a mix of real-world data (January 2020–December 2024) and synthetic data:

Real data from GCC financial institutions included labeled fraud, customer profiles, and transaction streams;

Synthetic fraud patterns were generated using FraudGAN to simulate edge cases, adversarial behaviors, and rare unseen fraud types.

Simulations covered the following:

Sudden spikes in fraud (e.g., during cyberattacks);

Gradual drift in transaction behavior (seasonal or policy-driven);

Adversarial manipulation using FGSM and PGD.

Baseline models were compared using no-preprocessing, SMOTE-only, and non-robust configurations to highlight HMLF advantages (Table 5, Figure 3).

Table 5. Real-time performance indicators under simulation.

Metric	Value
Avg. Prediction Latency	150 ms
Drift Response Time	<1 day
False Positive Rate	Low /Stable
Alert Volume Stability	High

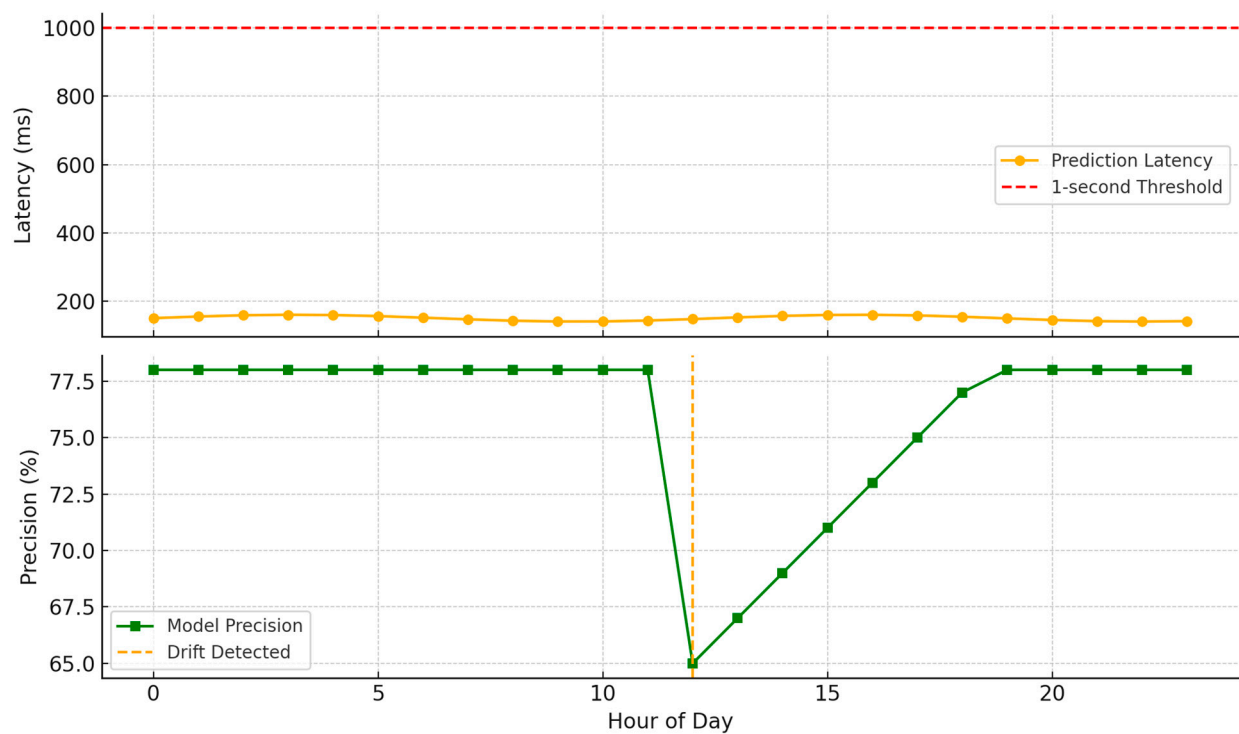


Figure 3. Prediction latency and drift recovery timeline.

4.4. Adversarial Robustness

In adversarial settings, where attackers attempt to subtly manipulate inputs to evade detection, the HMLF demonstrated notable resilience. This capability is especially critical for financial institutions in the GCC, where adversarial threats are increasing but formal defense protocols remain underdeveloped.

The HMLF integrates two major robustness mechanisms:

1. Adversarial Training: The model is trained using both clean and perturbed data (e.g., via FGSM, PGD), helping it to learn decision boundaries that are less sensitive to small, malicious changes;
2. FraudGAN: A generative adversarial network simulates sophisticated fraud patterns that may not yet exist in the labeled data, enriching the training set with diverse and hard-to-detect examples.

These techniques combined significantly improved the model’s robust accuracy—defined as the accuracy on adversarial examples—from 45% (baseline) to 85%. At the same time, the attack success rate (ASR) dropped from 35% to 5%, meaning most attempted attacks failed to bypass detection. Furthermore, perturbation sensitivity—the degree to

which small changes affect predictions—was greatly reduced, indicating a more stable and reliable model [67] (Table 6, Figure 4).

Table 6. Adversarial robustness comparison: Baseline vs. HMLF.

Metric	Baseline (%)	HMLF (%)
Robust Accuracy	45	85
Attack Success Rate (ASR)	35	5
Perturbation Sensitivity	High	Low

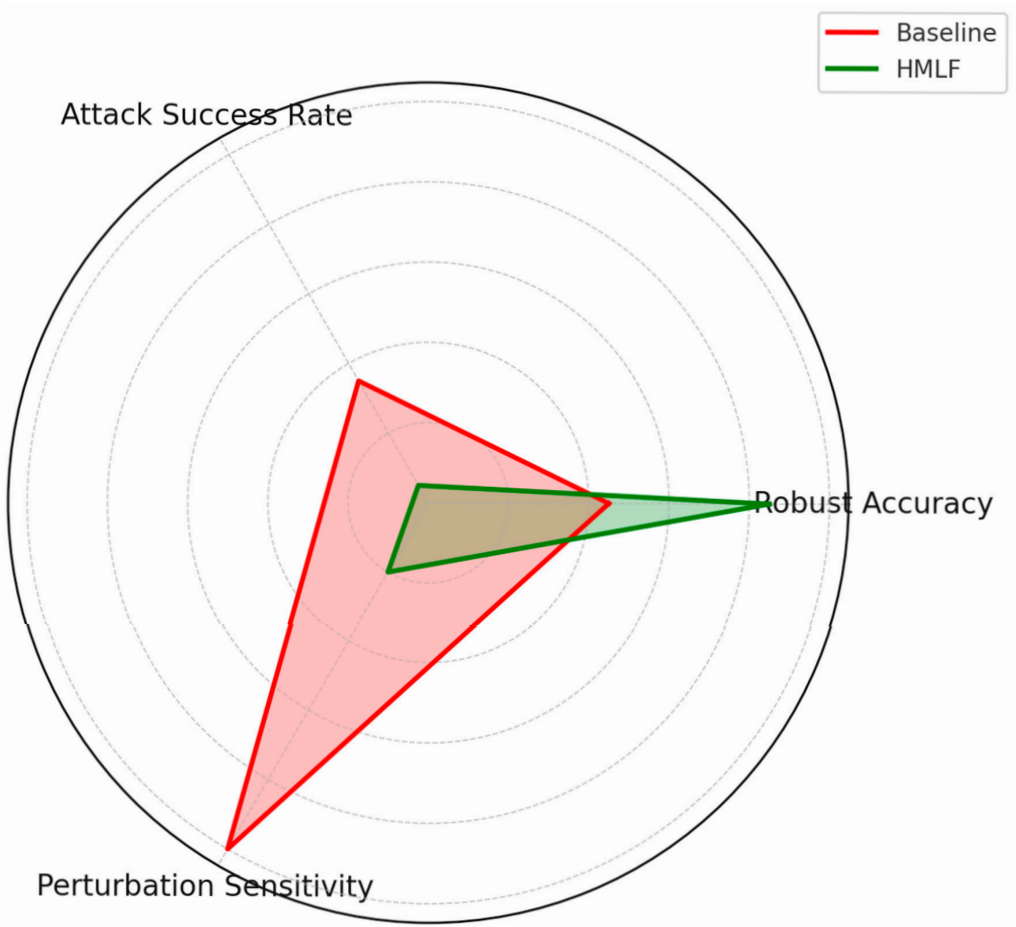


Figure 4. Adversarial Robustness Radar Chart.

4.5. Adaptive Learning and Drift Management

Fraud-detection models deployed in financial institutions must operate in rapidly changing environments. Regulatory updates, seasonal shopping patterns, and evolving attacker strategies can all introduce concept drift—a shift in the underlying data distribution that degrades model performance over time.

The HMLF responds to this challenge through adaptive-learning mechanisms that integrate incremental retraining with real-time drift detection [68,69].

Two drift-detection techniques, ADWIN (adaptive windowing) and DDM (drift-detection method) were implemented in the system in order to constantly track the prediction errors and to identify sudden or gradual changes in the patterns of fraud. Upon the detection of a drift event, the system will initiate purposeful retraining based on new data seen to ensure minimal downtime before restoring performance.

Simulation experiments prove that the HMLF has high levels of recall and precision even with extreme drift. In all three scenarios—seasonal shift, sudden fraud pattern changes, and simulated regulatory effects—temporary declines in recall were seen but were recovered to pre-drift levels in one day. This speaks well to the framework’s success in maintaining minimal disruption in operations while retaining detection quality (Table 7, Figure 5).

Table 7. Drift scenario simulation outcomes and recovery metrics.

Scenario	Recall (Before/After)	Precision (Before/After)	Recovery Time
Seasonal Drift	70% → 84%	68% → 80%	<24 h
Sudden Fraud Pattern Change	65% → 83%	66% → 78%	<1 day
Regulatory Shift (Simulated)	60% → 81%	63% → 77%	<18 h

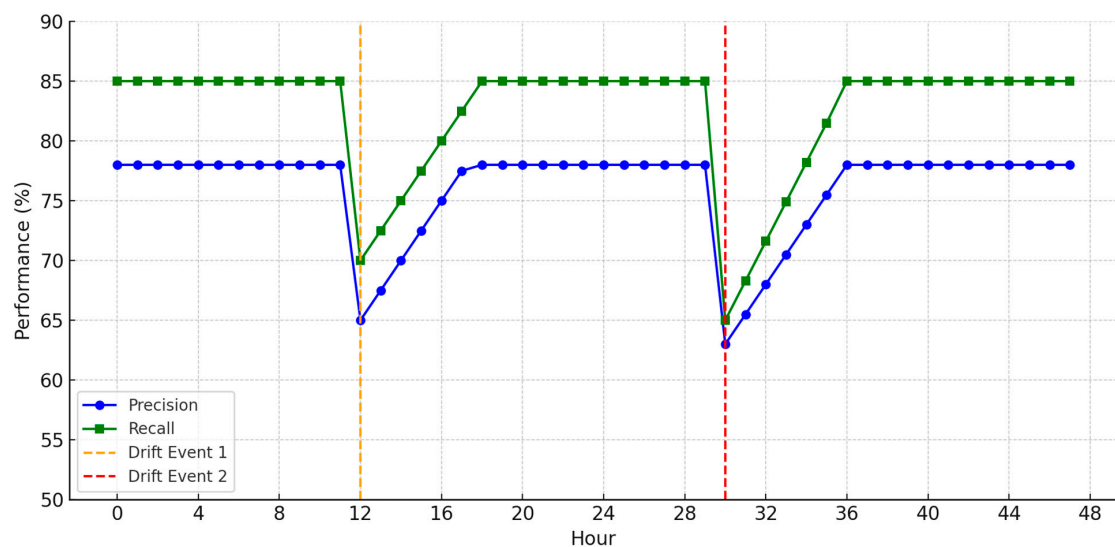


Figure 5. Drift Detection and Model Recovery Timeline.

5. Discussion

This research illustrates the viability and implications of using a modular, interpretable, and fault-tolerant fraud-detection system in the real-world context of Gulf Cooperation Council (GCC) financial institutions. The findings are theoretically and practically significant and are informed by the previous literature review.

First, the hybrid imbalance management approach with SMOTEBoost and cost-sensitive learning also improved recall substantially with little loss of precision. These findings support earlier evidence that single-process methodologies typically under-perform in financial fraud scenarios where fraud instances are few and it is expensive to misclassify them [1,51]. The successful implementation of human-in-the-loop (HITL) supervision also aligns with research promoting expert verification in high-stakes fields to guarantee equity, explainability, and accuracy [70,71].

Second, the adversarial robustness modules of the framework, such as adversarial training as well as the use of GAN-created fraud scenarios, reduced the success of attacks from 35% to 5%. This verifies that adversarial simulations such as FraudGAN offer critical protection in environments with minimal pre-existing adversarial protection infrastruc-

ture [2,14]. Notably, these functions worked well in institutions with limited computational power, supporting the argument that strong defense does not need excessively complex structure if modularized.

Third, the adaptive-learning approach using DDM and ADWIN facilitated recovery from concept drift within 24 hours, in line with research in the field pointing to the risk of using stationary models in high-frequency transaction environments [3]. In the context of the GCC—one where changing fraudulent methods tend to align with regulatory delays—this ability is not only useful but also critical.

Fourth, embedded HITL reviews and explainability attributes (SHAP, LIME) guaranteed compliance with auditing and trust demands. Although post hoc explainability methods have been faulted for instability and computational demand [28], integration with domain review countered the trust deficit and facilitated coherent outputs with updates. This approach specifically meets calls from regulators for transparent reasoning in automation decision making [4].

Fifth, modular architecture solved integration problems pervasive in disjointed financial ecosystems. Data silos, legacy infrastructure, and schema discrepancies are well documented in the GCC [30,32]. Through the deployment of centralized warehousing and real-time pipelines, the HMLF closed the gap between operational necessity and technical viability.

Ultimately, the research also recognizes the complexity–practical deployment trade-off. Although HMLF leverages several high-end components, each component (FraudGAN and SMOTEBoost are examples) can be turned on and off or replaced. This makes it easier for low-resource banks or small banks to start with lightweight or interpretable models and gradually increase as time goes on. This can overcome an issue identified in the existing research: the “AI ambition gap” in which institutions’ capacities are outpaced by what resources are available [23].

In summary, the HMLF presents an alignment of cutting-edge machine learning with the infrastructure, governance, and risk profiles of regulated financial environments. Its high-stakes real-time performance in the GCC context adds to its real-world viability and presents a model for deployable AI at-scale within comparable jurisdictions in an explainable and compliant way.

6. Implications

Aiming to operationalize AI-based fraud-detection systems in challenging, resource-constrained financial contexts, such those prevalent across the Gulf Cooperation Council (GCC), this study provides specific implications for practitioners, legislators, and system developers. The outcomes stress the requirement of modularity, explainability, adaptability, and human supervision in ensuring effective and dependable implementation.

6.1. For Practitioners

This study shows that, in fraud-detection performance, operational readiness and human supervision are equally important as the underlying algorithmic structure. Accordingly, practitioners should implement the following recommendations:

- Implement human-in-the-loop (HITL) approaches at key junctures in the fraud-detection process—label verification, anomaly analysis, and post-decision auditing—to establish system confidence and to lower false positives;
- Implement robust evaluation tools with model drift alerts, user-friendly dashboards with fraud probability ratings, flagged transactions, and model confidence indicators. Beyond sophisticated black-box models, these software technologies tell fraud investigators what the findings signify;

- Start with explainable models or simple methods such as oversampling in organizations with minimal data science strength of staff, then move to more complicated methods such as adaptive drift control or adversary testing as resources and skills expand;
- Provide audit-ready outputs in support of internal reviews and regulatory compliance in the form of detailed case annotations and decision-making traceable logs.

6.2. For Policymakers and Regulators

This study underlines the need for well-defined, technologically neutral governance concepts able to guide prudent AI implementation in the financial sector. The authorities should implement the following recommendations:

- Mandate explainability protocols above and beyond technical documentation, e.g., have models generate structured, human-readable justifications of high-risk decisions;
- Implement baseline standards for metadata logging, drift monitoring, retraining cycles, and post-deployment auditability. These allow for ensuring AI systems' conformity as datasets and behaviors change;
- Encourage the adoption of HITL and modular-governance frameworks by providing regulatory sandboxes, incentives, or compliance credits for institutions that adopt transparent and auditable AI;
- Encourage the regional harmonization of regulation, particularly in the GCC, where cross-border banking and differing levels of AI maturity create alignment issues.

6.3. For Technology Developers and System Architects

Technical deployment success depends also on integration compatibility, adaptability, and governance compatibility in addition to performance. Developers should inspire to enact the following recommendations:

- Implement modular API-first architectures in which each component, such as preprocessing, adversarial defense, and explainability, can be separately triggered, turned off, or swapped out as needed by the institution;
- Support legacy infrastructure through the creation of compatibility layers or lightweight versions of high-level modules to be utilized in low-compute environments;
- Implement schema validation, feature lineage tracking, and audit-ready metadata logging in the system pipeline for compliance in an automated manner;
- Offer configurable switches between interpretable and high-performance models to allow phased rollout and experimentation without sacrificing stability and traceability.

7. Conclusions

This study demonstrates that the hybrid machine-learning framework (HMLF) offers a robust, adaptable, and regulation-ready approach to financial fraud detection within the Gulf Cooperation Council (GCC). By combining SMOTEBoost with cost-sensitive learning, the framework effectively addresses class imbalance. Adversarial training and FraudGAN simulations significantly reduce susceptibility to manipulation, while adaptive-learning mechanisms (DDM and ADWIN) enable rapid response to concept drift. Explainability tools (SHAP, LIME) and embedded human-in-the-loop (HITL) processes further enhance stakeholder trust, accountability, and auditability. Additionally, the framework's two-tier data integration model resolves long-standing fragmentation issues common to financial institutions in the region.

The results validated all seven hypotheses put forward in the research (H1–H7), therefore verifying the success of the modular architecture of the framework in balancing performance, interpretability, and governance over limited contexts.

From these results, several actionable recommendations emerge:

- Adopt hybrid imbalance strategies to improve rare-event detection without sacrificing precision;
- Integrate adversarial defense mechanisms early in model design, particularly in high-risk or cross-border systems;
- Implement adaptive-learning pipelines to address dynamic fraud patterns in near real-time;
- Prioritize explainability and HITL oversight as built-in components, not as post hoc features;
- Invest in data integration architecture that supports traceability, schema validation, and cross-departmental collaboration;
- Start with light-weight modules, such as decision trees or sampling-based techniques, in resource-limited organizations, then scale toward more sophisticated components as capacity increases;
- Support continuous monitoring, internal review, and regulatory audits by means of adjustable HITL checkpoints and automated audit logging.

A key strength of this study lies in its real-world grounding and emphasis on deployment feasibility within complex regulatory environments. However, its focus on GCC institutions also presents one limitation: while the findings are highly relevant to regions with similar infrastructure and regulatory pressures, they may require contextual adaptation for high-resource or unregulated environments. This specificity presents an opportunity rather than a constraint, however, providing a tested, modular blueprint that can be tailored to fit diverse financial or compliance-driven sectors.

Future studies should explore generalizing HMLF principles to other fields such as healthcare, insurance, and the public sector, where risk regulation and dispersed infrastructure pose analogous problems.

More research should be conducted into the following areas:

- Briefed versions of HMLF for application in low-resource settings;
- More sophisticated fraud types including collusive, internal, and synthetic identity fraud;
- HITL process automation and adaptive threshold roles in real-time environments.

Author Contributions: Conceptualization, K.I.A.-D. and I.A.A.-A.; methodology, K.I.A.-D.; software, K.I.A.-D.; validation, K.I.A.-D. and I.A.A.-A.; formal analysis, K.I.A.-D.; investigation, K.I.A.-D.; resources, K.I.A.-D.; data curation, K.I.A.-D.; writing—original draft preparation, K.I.A.-D.; writing—review and editing, K.I.A.-D. and I.A.A.-A.; visualization, K.I.A.-D.; supervision, I.A.A.-A.; project administration, K.I.A.-D.; funding acquisition, I.A.A.-A. All authors have read and agreed to the published version of the manuscript.

Funding: The APC was funded by Middle East University, Amman, Jordan.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (Ethics Committee) of Middle East University (MEU Research Committee) (protocol code MEU-IRB-2024-1220 and date of approval 20-12-2024).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on reasonable request from the corresponding author. The data are not publicly available due to privacy and ethical restrictions imposed by the Institutional Review Board of Middle East University.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Mujahid, M.; Kina, E.; Rustam, F.; Villar, M.G.; Alvarado, E.S.; De La Torre Diez, I.; Ashraf, I. Data Oversampling and Imbalanced Datasets: An Investigation of Performance for Machine Learning and Feature Engineering. *J. Big Data* **2024**, *11*, 87. [\[CrossRef\]](#)
- Taheri, R.; Shojafar, M.; Arabikhan, F.; Gegov, A. Unveiling Vulnerabilities in Deep Learning-Based Malware Detection: Differential Privacy Driven Adversarial Attacks. *Comput. Secur.* **2024**, *146*, 104035. [\[CrossRef\]](#)
- Arora, S.; Rani, R.; Saxena, N. A Systematic Review on Detection and Adaptation of Concept Drift in Streaming Data Using Machine Learning Techniques. *WIREs Data Min. Knowl. Discov.* **2024**, *14*, e1536. [\[CrossRef\]](#)
- Salih, A.M.; Raisi-Estabragh, Z.; Galazzo, I.B.; Radeva, P.; Petersen, S.E.; Lekadir, K.; Menegaz, G. A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME. *Adv. Intell. Syst.* **2025**, *7*, 2400304. [\[CrossRef\]](#)
- Tayebi, M.; El Kafhali, S. Generative Modeling for Imbalanced Credit Card Fraud Transaction Detection. *J. Cybersecur. Priv.* **2025**, *5*, 9. [\[CrossRef\]](#)
- Mehmood, H.; Khalid, A.; Kostakos, P.; Gilman, E.; Pirttikangas, S. A Novel Edge Architecture and Solution for Detecting Concept Drift in Smart Environments. *Future Gener. Comput. Syst.* **2024**, *150*, 127–143. [\[CrossRef\]](#)
- Vimbi, V.; Shaffi, N.; Mahmud, M. Interpreting Artificial Intelligence Models: A Systematic Review on the Application of LIME and SHAP in Alzheimer's Disease Detection. *Brain Inform.* **2024**, *11*, 10. [\[CrossRef\]](#)
- Mostofi, F.; Tokdemir, O.B.; Toğan, V. Generating Synthetic Data with Variational Autoencoder to Address Class Imbalance of Graph Attention Network Prediction Model for Construction Management. *Adv. Eng. Inform.* **2024**, *62*, 102606. [\[CrossRef\]](#)
- Zhang, Q.; Zhu, Y.; Yang, M.; Jin, G.; Zhu, Y.; Chen, Q. Cross-to-Merge Training with Class Balance Strategy for Learning with Noisy Labels. *Expert. Syst. Appl.* **2024**, *249*, 123846. [\[CrossRef\]](#)
- Chang, S.; Wang, L.; Shi, M.; Zhang, J.; Yang, L.; Cui, L. Extended Attention Signal Transformer with Adaptive Class Imbalance Loss for Long-Tailed Intelligent Fault Diagnosis of Rotating Machinery. *Adv. Eng. Inform.* **2024**, *60*, 102436. [\[CrossRef\]](#)
- Fandio, A.; Awe, O.O. Comparative Study of Supervised Machine Learning Algorithms for Predicting Oversampled Imbalanced Medical Data. In *Practical Statistical Learning and Data Science Methods*; Awe, O.O., A. Vance, E., Eds.; STEAM-H: Science, Technology, Engineering, Agriculture, Mathematics & Health; Springer Nature: Cham, Switzerland, 2025; pp. 667–696. ISBN 978-3-031-72214-1.
- Altalhan, M.; Algarni, A.; Alouane, M.T.-H. Imbalanced Data Problem in Machine Learning: A Review. *IEEE Access* **2025**, *13*, 13686–13699. [\[CrossRef\]](#)
- Alam, N.; Rahman, M.A.; Islam, M.R.; Hossain, M.J. Ensemble Adversarial Training-Based Robust Model for Multi-Horizon Dynamic Line Rating Forecasting against Adversarial Attacks. *Electr. Power Syst. Res.* **2025**, *241*, 111289. [\[CrossRef\]](#)
- Morshed, A.; Ramadan, A.; Maali, B.; Khrais, L.T.; Baker, A.A.R. Transforming accounting practices: The impact and challenges of business intelligence integration in invoice processing. *J. Infrastruct. Policy Dev.* **2024**, *8*, 4241. [\[CrossRef\]](#)
- Malik, J.; Muthalagu, R.; Pawar, P.M. A Systematic Review of Adversarial Machine Learning Attacks, Defensive Controls and Technologies. *IEEE Access* **2024**, *12*, 99382–99421. [\[CrossRef\]](#)
- Nirmal, S.; Patil, P.; Shinde, S. Adversarial Measurements for Convolutional Neural Network-Based Energy Theft Detection Model in Smart Grid. *E-Prime-Adv. Electr. Eng. Electron. Energy* **2025**, *11*, 100909. [\[CrossRef\]](#)
- Gopal, O.N. UAE Signs Two Agreements with the International Monetary Fund to Support the “Resilience and Sustainability Trust” and the “Poverty Reduction and Growth Trust” Funds; United Arab Emirates Ministry of Finance: Abu Dhabi, United Arab Emirates, 2024.
- Noor, K.; Imoize, A.L.; Li, C.-T.; Weng, C.-Y. A Review of Machine Learning and Transfer Learning Strategies for Intrusion Detection Systems in 5G and Beyond. *Mathematics* **2025**, *13*, 1088. [\[CrossRef\]](#)
- Aguiar, G.J.; Cano, A. A Comprehensive Analysis of Concept Drift Locality in Data Streams. *Knowl. Based Syst.* **2024**, *289*, 111535. [\[CrossRef\]](#)
- Morshed, A.; Khrais, L.T. Cybersecurity in Digital Accounting Systems: Challenges and Solutions in the Arab Gulf Region. *J. Risk Financ. Manag.* **2025**, *18*, 41. [\[CrossRef\]](#)
- Li, Y.; Xiong, S.; Li, Q.; Chen, Z. A TG-AGD Anomaly Image Detection Model Based on Residual Bottleneck Attention and Time Series Prediction. *Appl. Soft Comput.* **2025**, *173*, 112746. [\[CrossRef\]](#)
- Modiha, P. Digital Transformation in the Finance and Banking Sector. In *Digital Transformation in South Africa*; Moloi, T., Ed.; Professional Practice in Governance and Public Organizations; Springer Nature: Cham, Switzerland, 2024; pp. 95–117. ISBN 978-3-031-52402-8.
- Melnyk, V. Transforming the Nature of Trust between Banks and Young Clients: From Traditional to Digital Banking. *Qual. Res. Financ. Mark.* **2024**, *16*, 618–635. [\[CrossRef\]](#)
- Wang, Y.; Wang, M.; Pan, Y.; Chen, J. Joint Loan Risk Prediction Based on Deep Learning-optimized Stacking Model. *Eng. Rep.* **2024**, *6*, e12748. [\[CrossRef\]](#)
- 6.2.3 Fraud Investigation and Reporting. CBUAE Rulebook. Available online: https://rulebook.centralbank.ae/en/rulebook/62-3-fraud-investigation-and-reporting?utm_source=chatgpt.com (accessed on 4 April 2025).

26. Emami, Y.; Almeida, L.; Li, K.; Ni, W.; Han, Z. Human-In-The-Loop Machine Learning for Safe and Ethical Autonomous Vehicles: Principles, Challenges, and Opportunities. *arXiv* **2024**, arXiv:2408.12548.
27. Badhon, B.; Chakraborty, R.K.; Anavatti, S.G.; Vanhoucke, M. A Multi-Module Explainable Artificial Intelligence Framework for Project Risk Management: Enhancing Transparency in Decision-Making. *Eng. Appl. Artif. Intell.* **2025**, *148*, 110427. [[CrossRef](#)]
28. Retzlaff, C.O.; Angerschmid, A.; Saranti, A.; Schneeberger, D.; Roettger, R.; Mueller, H.; Holzinger, A. Post-Hoc vs Ante-Hoc Explanations: xAI Design Guidelines for Data Scientists. *Cogn. Syst. Res.* **2024**, *86*, 101243. [[CrossRef](#)]
29. Abekoon, T.; Sajindra, H.; Rathnayake, N.; Ekanayake, I.U.; Jayakody, A.; Rathnayake, U. A Novel Application with Explainable Machine Learning (SHAP and LIME) to Predict Soil N, P, and K Nutrient Content in Cabbage Cultivation. *Smart Agric. Technol.* **2025**, *11*, 100879. [[CrossRef](#)]
30. AlBenJasim, S.; Takruri, H.; Al-Zaidi, R.; Dargahi, T. Development of Cybersecurity Framework for FinTech Innovations: Bahrain as a Case Study. *Int. Cybersecur. Law. Rev.* **2024**, *5*, 501–532. [[CrossRef](#)]
31. Kumar, S.; Datta, S.; Singh, V.; Datta, D.; Singh, S.K.; Sharma, R. Applications, Challenges, and Future Directions of Human-in-the-Loop Learning. *IEEE Access* **2024**, *12*, 75735–75760. [[CrossRef](#)]
32. Khan, H.U.; Malik, M.Z.; Alomari, M.K.B.; Khan, S.; Al-Maadid, A.A.S.A.; Hassan, M.K.; Khan, K. Transforming the Capabilities of Artificial Intelligence in GCC Financial Sector: A Systematic Literature Review. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 8725767. [[CrossRef](#)]
33. Salhab, H.; Zoubi, M.; Khrais, L.T.; Estaitia, H.; Harb, L.; Al Huniti, A.; Morshed, A. AI-Driven Sustainable Marketing in Gulf Cooperation Council Retail: Advancing SDGs Through Smart Channels. *Adm. Sci.* **2025**, *15*, 20. [[CrossRef](#)]
34. Ismail, A.; Szali, F.H.; Jawaddi, S.N.A.; Mutalib, S. Stream ETL Framework for Twitter-Based Sentiment Analysis: Leveraging Big Data Technologies. *Expert. Syst. Appl.* **2025**, *261*, 125523. [[CrossRef](#)]
35. Sun, H.; Liu, Y.; Al-Tahmeesschi, A.; Nag, A.; Soleimanpour-Moghadam, M.; Canberk, B.; Arslan, H.; Ahmadi, H. Advancing 6G: Survey for Explainable AI on Communications and Network Slicing. *IEEE Open J. Commun. Soc.* **2025**, *6*, 1372–1412. [[CrossRef](#)]
36. Qin, C.; Chen, X.; Wang, C.; Wu, P.; Chen, X.; Cheng, Y.; Zhao, J.; Xiao, M.; Dong, X.; Long, Q.; et al. SciHorizon: Benchmarking AI-for-Science Readiness from Scientific Data to Large Language Models. *arXiv* **2025**, arXiv:2503.13503.
37. Gyau, E.B.; Appiah, M.; Gyamfi, B.A.; Achie, T.; Naeem, M.A. Transforming Banking: Examining the Role of AI Technology Innovation in Boosting Banks Financial Performance. *Int. Rev. Financ. Anal.* **2024**, *96*, 103700. [[CrossRef](#)]
38. Alshammari, R. Building a Unified Legislative Framework for the Development and Registration for the Emerging Technologies in the GCC. In *Innovation and Development of Knowledge Societies*; Routledge: London, UK, 2025; pp. 36–52.
39. Verma, A.; Allen, T. A Sociotechnical Readiness Level Framework for the Development of Advanced Nuclear Technologies. *Nucl. Technol.* **2024**, *210*, 1722–1739. [[CrossRef](#)]
40. Kothandapani, H.P. AI-Driven Regulatory Compliance: Transforming Financial Oversight through Large Language Models and Automation. *Emerg. Sci. Res.* **2025**, *3*, 12–24.
41. Claessens, S.; Cong, L.W.; Moshirian, F.; Park, C.-Y. Opportunities and Challenges Associated with the Development of FinTech and Central Bank Digital Currency. *J. Financ. Stab.* **2024**, *73*, 101280. [[CrossRef](#)]
42. Al-Debei, M.M.; Hujran, O.; Al-Adwan, A.S. Net Valence Analysis of Iris Recognition Technology-Based FinTech. *Financ. Innov.* **2024**, *10*, 59. [[CrossRef](#)]
43. Hafez, I.Y.; Hafez, A.Y.; Saleh, A.; Abd El-Mageed, A.A.; Abohany, A.A. A Systematic Review of AI-Enhanced Techniques in Credit Card Fraud Detection. *J. Big Data* **2025**, *12*, 6. [[CrossRef](#)]
44. Young, B.; Anderson, D.T.; Keller, J.; Petry, F.; Michael, C.J. SPARC: A Human-in-the-Loop Framework for Learning and Explaining Spatial Concepts. *Information* **2025**, *16*, 252. [[CrossRef](#)]
45. Abdallah, A.; Ahmad, A.; Said, B. Balancing Privacy and Usability: A Design Science Research Approach for Cookie Consent Mechanisms. *J. Open Innov. Technol. Mark. Complex.* **2025**, *11*, 100520. [[CrossRef](#)]
46. Nandan Prasad, A. Data Quality and Preprocessing. In *Introduction to Data Governance for Machine Learning Systems*; Apress: Berkeley, CA, USA, 2024; pp. 109–223. ISBN 979-8-8688-1022-0.
47. Butt, A.Q.; Shangguan, D.; Ding, Y.; Banerjee, A.; Mukhtar, M.A.; Taj, K. Evaluation of Environmental Impact Assessment and Mitigation Strategies for Gulpur Hydropower Project, Kotli, Pakistan. *Discov. Appl. Sci.* **2024**, *6*, 137. [[CrossRef](#)]
48. Shan, J.; Che, Z.; Liu, F. Accurate Rotor Temperature Prediction of Permanent Magnet Synchronous Motor in Electric Vehicles Using a Hybrid RIME-XGBoost Model. *Appl. Sci.* **2025**, *15*, 3688. [[CrossRef](#)]
49. Marra, M.; Nielsen, B.B. Research Methodology: Best Practices for Rigorous, Credible, and Impactful Research: Herman Aguinis. *J. Int. Bus. Stud.* **2025**, 1–3. [[CrossRef](#)]
50. Sahlaoui, H.; Alaoui, E.A.A.; Agoujil, S.; Nayyar, A. An Empirical Assessment of Smote Variants Techniques and Interpretation Methods in Improving the Accuracy and the Interpretability of Student Performance Models. *Educ. Inf. Technol.* **2024**, *29*, 5447–5483. [[CrossRef](#)]
51. Salmi, M.; Atif, D.; Oliva, D.; Abraham, A.; Ventura, S. Handling Imbalanced Medical Datasets: Review of a Decade of Research. *Artif. Intell. Rev.* **2024**, *57*, 273. [[CrossRef](#)]

52. Hou, Z.; Tang, J.; Li, Y.; Fu, S.; Tian, Y. MVQS: Robust Multi-View Instance-Level Cost-Sensitive Learning Method for Imbalanced Data Classification. *Inf. Sci.* **2024**, *675*, 120467. [[CrossRef](#)]
53. Ünal, S.; Günay, O.; Akkurt, I.; Gunoglu, K.; Tekin, H.O. A Comparative Study on Breast Cancer Classification with Stratified Shuffle Split and K-Fold Cross Validation via Ensembled Machine Learning. *J. Radiat. Res. Appl. Sci.* **2024**, *17*, 101080. [[CrossRef](#)]
54. Eleftheriadis, C.; Symeonidis, A.; Katsaros, P. Adversarial Robustness Improvement for Deep Neural Networks. *Mach. Vis. Appl.* **2024**, *35*, 35. [[CrossRef](#)]
55. Shan, J.; Xie, Y.; Wei, L. A GAN-Based Hybrid Sampling Method for Transaction Fraud Detection. In Proceedings of the 2024 International Conference on Networking, Sensing and Control (ICNSC), Marseille, France, 25–27 October 2023; IEEE: New York, NY, USA, 2024; pp. 1–7.
56. Yumlembam, R.; Issac, B.; Jacob, S.M.; Yang, L. Comprehensive Botnet Detection by Mitigating Adversarial Attacks, Navigating the Subtleties of Perturbation Distances and Fortifying Predictions with Conformal Layers. *Inf. Fusion.* **2024**, *111*, 102529. [[CrossRef](#)]
57. Goncalves, V.P.M.; Silva, L.P.; Nunes, F.L.S.; Ferreira, J.E.; Araújo, L.V. Concept Drift Adaptation in Video Surveillance: A Systematic Review. *Multimed. Tools Appl.* **2024**, *83*, 9997–10037. [[CrossRef](#)]
58. Makhmudov, F.; Kilichev, D.; Giyosov, U.; Akhmedov, F. Online Machine Learning for Intrusion Detection in Electric Vehicle Charging Systems. *Mathematics* **2025**, *13*, 712. [[CrossRef](#)]
59. Kotsiopoulos, T.; Papakostas, G.; Vafeiadis, T.; Dimitriadis, V.; Nizamis, A.; Bolzoni, A.; Bellinati, D.; Ioannidis, D.; Votis, K.; Tzovaras, D. Revolutionizing Defect Recognition in Hard Metal Industry through AI Explainability, Human-in-the-Loop Approaches and Cognitive Mechanisms. *Expert. Syst. Appl.* **2024**, *255*, 124839. [[CrossRef](#)]
60. Yau, K.L.A.; Saleem, Y.; Chong, Y.W.; Fan, X.; Eyu, J.M.; Chieng, D. The augmented intelligence perspective on human-in-the-loop reinforcement learning: Review, concept designs, and future directions. *IEEE Trans. Hum.-Mach. Syst.* **2024**, *54*, 1023–1038. [[CrossRef](#)]
61. Dworzanczyk, A.R.; Viqueira-Moreira, M.; Langhorn, J.D.; Libeau, M.A.; Brehm, C.; Parziale, N.J. On Aerobreakup in the Stagnation Region of High-Mach-Number Flow over a Bluff Body. *J. Fluid. Mech.* **2025**, *1002*, A1. [[CrossRef](#)]
62. Alshaleel, M.K. The Extraterritoriality of the Gdpr and Its Effect on Gcc Businesses. *Glob. J. Comp. Law.* **2024**, *13*, 201–226. [[CrossRef](#)]
63. Al-Dmour, A.; Al-Dmour, R.; Al-Dmour, H.; Al-Adwan, A. Blockchain Applications and Commercial Bank Performance: The Mediating Role of AIS Quality. *J. Open Innov. Technol. Mark. Complex.* **2024**, *10*, 100302. [[CrossRef](#)]
64. Miller, K.M.; Lukic, K.; Skiera, B. The Impact of the General Data Protection Regulation (GDPR) on Online Tracking. *Int. J. Res. Mark.* **2025**, 1–97. [[CrossRef](#)]
65. Salehi, A.R.; Khedmati, M. A Cluster-Based SMOTE Both-Sampling (CSBBoost) Ensemble Algorithm for Classifying Imbalanced Data. *Sci. Rep.* **2024**, *14*, 5152. [[CrossRef](#)]
66. Agbedanu, P.R.; Yang, S.J.; Musabe, R.; Gatere, I.; Rwigema, J. A Scalable Approach to Internet of Things and Industrial Internet of Things Security: Evaluating Adaptive Self-Adjusting Memory K-Nearest Neighbor for Zero-Day Attack Detection. *Sensors* **2025**, *25*, 216. [[CrossRef](#)]
67. Zhang, Y.; Bryan, J.; Richards, G.; Wang, H. Development of Surrogate-Optimization Models for a Novel Transcritical Power Cycle Integrated with a Small Modular Reactor. *Energy AI* **2024**, *15*, 100311. [[CrossRef](#)]
68. Taqa, S.B.A. The Mediating Role of Remote Communication on the Relationship Between Electronic Human Resource Management Practices and Organizational Performance in Iraqi Commercial Banks. *Middle East J. Commun. Sci.* **2025**, *5*, 1–52.
69. Ahmed, A.K.; Nahar, H.M.; Manajrah, M.M.N. Effect of Social Media on Shaping the Agenda of the Communicator in the Jordanian TV Channels. *Middle East J. Commun. Sci.* **2023**, *3*, 1–40. [[CrossRef](#)]
70. Oregat, A. The Degree of Satisfaction of Facebook Users About Its Features, Usage Motives and Achieved Gratifications: An Applied Study on Students of the Faculty of Mass Communication at the Middle East University. *Middle East J. Commun. Sci.* **2021**, *1*, 1–35. [[CrossRef](#)]
71. Al-Muntasir, M. The Phenomenon of Information Flow from Traditional and New Media About the Corona Pandemic from the Perspective of Newly Graduated Media Professionals in Yemen. *Middle East J. Commun. Sci.* **2022**, *2*, 1–25. Available online: <https://digitalcommons.aaru.edu.jo/mejcs/vol2/iss2/1> (accessed on 3 March 2025).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.