



ADAPTIVE RISK THRESHOLDING USING EXPLAINABLE MACHINE LEARNING
FOR DYNAMIC FINANCIAL FRAUD DETECTION

Varsha Shah
Independent Researcher, Seattle, WA, USA

Abstract

Traditional rule-based fraud screening is no longer effective at stopping increasingly complex or rapidly changing fraud schemes for financial institutions. This systematic review aims to map the state of the art regarding the potential of implementing adaptive risk thresholding in dynamic financial fraud detection, based on explainable machine learning as a key component of this paradigm. The synthesis is organized around thematic strands of traditional and ensemble classifiers, of deep learning and sequential architectures, of methods to combat extreme class imbalance, of explainable artificial intelligence methods like Shapley Additive exPlanations and Local Interpretable Model-Agnostic Explanations, and of feature engineering applicable to practitioners. In the literature surveyed, the ensemble and boosting architectures, especially gradient-boosted trees, are reported to yield area-under-the-curve values that are often above zero-point-nine-five, and the resampling methods, like the Synthetic Minority Oversampling Technique, were reported to improve minority class recall significantly over baseline methods. The review also identifies that post-hoc explanation frameworks continue to be adopted underpinning the risk-scoring pipelines, allowing threshold parameters to be dynamically adjusted based upon feature level attributions, instead of statically set a priori. A conceptual framework is suggested that connects the levels of feature engineering at the transaction, risk scoring using ensemble or deep-learning based approaches, an explanation layer, and an adaptive threshold controller in a single operational pipeline. Results suggest that thresholds learned from adapting to explainability can alleviate false-positive costs, while maintaining high sensitivity to real fraud, though challenges of concept drift, computation costs, and regulatory audibility persist. The review finds that incorporating explainable machine learning alongside dynamic thresholding is a promising area for banks aiming to achieve high accuracy and accountability in fraud identification processes.

Keywords: Adaptive risk thresholding, Explainable artificial intelligence, SHAP, LIME, Financial, fraud detection, Ensemble learning, XGBoost, Class imbalance, SMOTE, Anomaly detection, Credit card fraud, Dynamic risk scoring

I. INTRODUCTION

Card-not-Present (CNP) transactions, digital wallets, real-time payment rails and other innovations are growing the attack surface for financial fraud, which is one of the most persistent challenges to the stability and trustworthiness in modern payment ecosystems. As comprehensive reviews of intelligent fraud detection do, the field has been shifting from the rule-based expert systems to the statistical and machine-learning paradigms, as the variety of fraud typologies and the adversarial nature of fraudulent behavior require continuously updated fraud detection strategies and not static decision rules [1].



The use of machine learning and deep learning algorithms for transactional fraud screening, which can capture nonlinear relationships among the various behavioral, temporal, and merchant-level features that are not amenable to static rule engines, has mushroomed in its adoption rate. Evaluation of state-of-the-art classifiers using real card transaction data shows that ensemble and deep architectures typically achieve better discrimination, for example as measured by the area under the receiver operating characteristic curve, while at the same time sometimes compromising on interpretability [2].

One of the hallmarks of financial fraud datasets is the high degree of class imbalance, with legitimate transactions far outweighing fraudulent ones - sometimes by a ratio of more than a few hundred to one. Early comparative analyses of these data-mining techniques for credit card fraud identification show that this imbalance can result in a tendency of naive classifiers to favour the majority class, thereby reducing the recall for the minority class (fraud) and giving a rationale for special resampling, cost-sensitive, and threshold-adjustment approaches, which are still the primary focus of the field [3]. The false alarm rate can vary significantly from one time and one customer segment to another and a purely static decision threshold is not structurally suitable to the problem because the cost of a false alarm is very different from the cost of a lost fraudulent transaction.

As well as the gains in performance, there has been a parallel regulatory and operational demand for transparency of models, bringing explainable AI from the academic to the operational realm in financial risk management. Research to date on explainability in fintech risk settings demonstrates that post-hoc explainable approaches can provide a balance between the performance gains of complex ensemble and deep models and the auditability that financial regulators and fraud experts demand [4]. This review brings together results of 25 peer-reviewed papers published up to 2023 to explore the potential of adaptive risk thresholding within the context of explainable machine learning for dynamic, defensible and operationally efficient financial fraud detection, and suggests a conceptual framework connecting the risk scoring, explanation generation and threshold adaptation into a coherent decision pipeline.

II. LITERATURE REVIEW

2.1 Traditional and Ensemble Machine Learning Approaches

Ensembles of boosting are particularly prominent in the field of fraud detection, with many small classifiers contributing to a large one. In the case of European card transaction data, evaluations using majority-voting aggregation of AdaBoost ensembles are significantly better at detecting fraudulent examples than are single classifiers, as the signals provided by classifiers are relatively weak and few in number [5].

Gradient-boosted decision trees, formulated in the scalable tree-boosting framework [6] minimize a regularized objective function that discourages model complexity and, step by step, fits the residual errors, given here as λ and γ are regularization parameters used for controlling tree size and leaf weights, respectively, and l denotes a differentiable loss function.

$$\text{Obj}(\theta) = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k), \quad \Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (1)$$

The same regularization property, together with the automatic processing of sparse features, is the reason why this algorithm is used so widely in transactional risk scoring. When applied to predicting credit-card fraud, gradient-boosted trees are consistently among the best classifiers



compared to logistic regression, decision trees, and support vector machines, and feature-importance values calculated from the split-gain statistics provide a simple yet coarse level of interpretability [7].

Extensions of the boosting framework to mobile-payment environments further demonstrate the scalability to high-velocity, high-volume transaction streams, reporting that carefully tuned boosting-based pipelines achieve good discrimination performance and remain computationally tractable (and hence suitable for near-real-time deployment), a property that is directly relevant to adaptive thresholding architectures that need to continuously recompute risk scores [8].

2.2 Deep Learning and Sequential Modeling Approaches

Recurrent architectures, which are able to model temporal dependencies, have been of great interest since fraudulent activity often appears in the form of irregular patterns of transactions instead of single transactions. Compared to approaches that rely on feature-engineered tabular models, sequence-classification methods for card transactions in sequential order have been shown to be able to detect behavioral drift and velocity patterns which are not captured by the feature-engineered models, making them more likely to detect low-dollar, high-frequency fraud campaigns [9].

Ensembles of neural networks with structured feature-engineering pipelines achieve additional improvements: Adding aggregated features such as rolling transaction counts and merchant category frequencies, in addition to multilayer neural networks, results in better precision and recall than using the raw features as deep models [10].

The use of denoising autoencoders as unsupervised feature extractors for credit-risk analysis has been proposed, where reconstruction error is an additional unsupervised anomaly signal which can be combined with supervised risk scores to add a dimension for threshold calibration [11].

Table 1. Taxonomy of Reviewed Studies by Methodological Category

Category	Representative Techniques	Studies	Primary Contribution
Foundational reviews	Classification frameworks; comparative surveys	[1], [3]	Establish taxonomy and evaluation baselines
Ensemble & boosting	AdaBoost; XGBoost; majority voting	[5]–[8]	High discrimination with moderate interpretability
Deep learning & sequential	LSTM; autoencoders; neural ensembles	[9]–[12]	Capture temporal and nonlinear behavioral patterns
Class-imbalance & hybrid	SMOTE; GAN oversampling; unsupervised–supervised hybrids	[13]–[17]	Improve minority-class recall
Explainable AI	SHAP; LIME; explainable credit-risk frameworks	[18]–[20]	Provide local and global model transparency
Feature engineering	Practitioner lessons; hybrid feature selection	[21], [22]	Improve feature relevance; reduce dimensionality
Broader fraud context	Financial-statement and reporting- fraud models	[23]–[25]	Extend principles beyond transactional fraud

Source: Synthesized by the authors from the twenty-five reviewed studies.

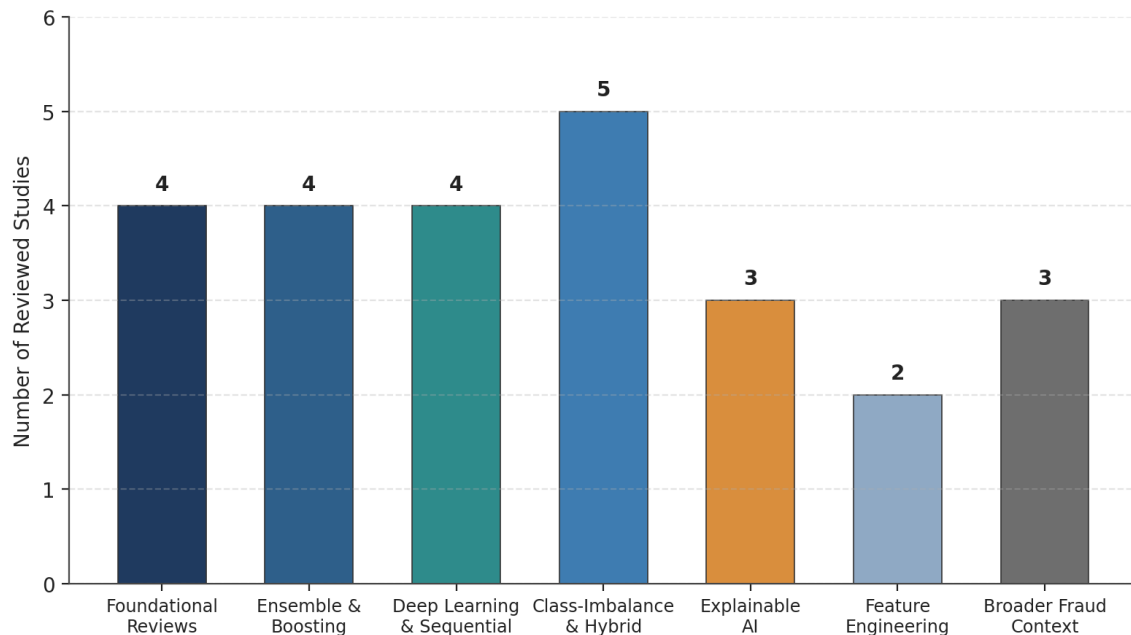


Figure 1. Distribution of Reviewed Studies Across Methodological Categories

2.3 Addressing Class Imbalance and Hybrid Detection Strategies

It has been reported that data-level strategies like Synthetic Minority Oversampling Technique combined with reweighting of the minority classes using AdaBoost-like ensemble classifiers can improve minority-class recall significantly without significantly increasing false positives, highlighting the complementary effect of data-level and algorithm-level remedies for imbalanced data sets [13].

Realistic models point out that using a random train-test split for offline evaluation can lead to optimistic results in the real-world, where actual transaction streams have concept drift, seasonal variation, and delayed feedback for fraud labels; and the learning strategies explicitly consider verification latency and non-stationary fraud distributions [14]—important concerns that are directly relevant to any adaptive thresholding strategy that assumes that the fraud labels are available in near time.

Generative adversarial oversampling methods have been shown to be more effective at classification than traditional oversampling methods, especially if fraud patterns are highly non-linear and non-uniformly distributed among existing fraud and non-fraud transactions [15].

Comparative experiments of the data-level algorithms once more reveal that there is no single resampling method that is consistently superior under all imbalance ratio conditions, and that the benefits of oversampling, undersampling and hybrid algorithms varies significantly with both the base classifier and the imbalance ratio, thus supporting the idea that the optimal configuration should be adaptive rather than fixed [16].

Unsupervised stages for anomaly detection in conjunction with supervised stages for classification reported higher detection rates for unseen typologies of fraud, as distributional outliers detected by unsupervised can provide an additional signal that the supervised risk scores could not have detected, and a signal that can be used by adaptive threshold controllers, together with supervised signals, to detect anomalies [17].



Table 2. Class-Imbalance Handling Techniques Compared

Technique	Mechanism	Reported Effect on Recall	Key Limitation
Random / SMOTE oversampling	Synthesizes minority-class samples via interpolation	Substantial recall improvement	Risk of overfitting to synthetic patterns
Generative adversarial oversampling	Learns minority-class distribution generatively	Improved classification effectiveness	Higher training complexity
Cost-sensitive learning	Penalizes misclassification asymmetrically	Moderate recall gains	Requires careful cost calibration
Ensemble resampling (AdaBoost variants)	Reweights misclassified minority instances iteratively	High recall with maintained precision	Sensitive to noisy labels
Hybrid unsupervised-supervised	Combines anomaly scores with supervised classifiers	Improved detection of novel fraud patterns	Increased pipeline complexity

Source: Synthesized by the authors from references [13]–[17].

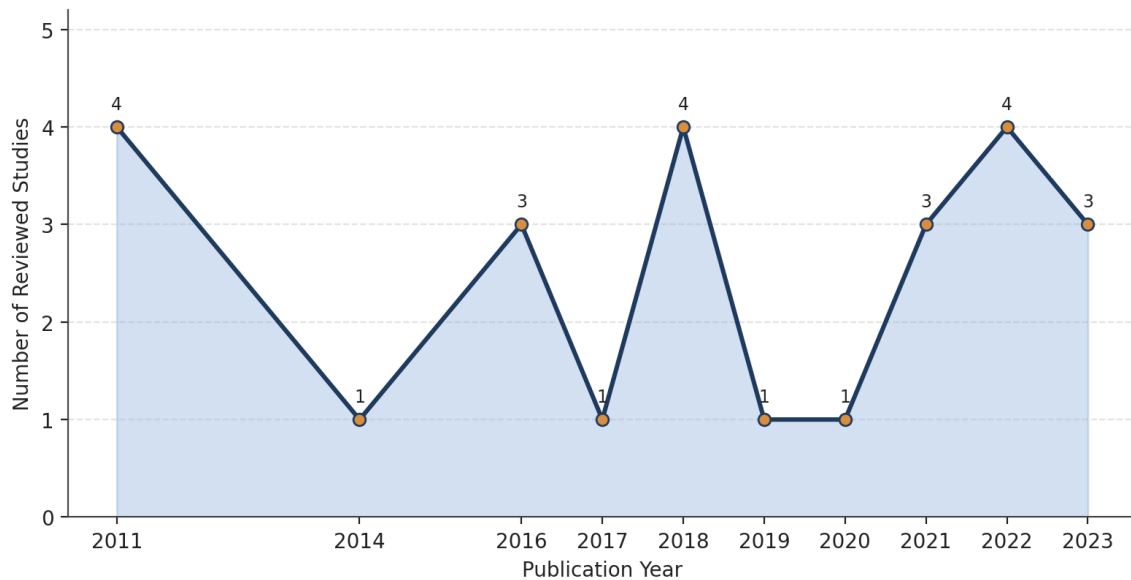


Figure 2. Publication Trend of Reviewed Literature by Year (2011–2023)

2.4 Explainable Artificial Intelligence Techniques

Shapley Additive exPlanations measure the marginal contribution of each feature to the individual prediction by computing the average effect of that feature over all possible coalitions of features S , formalized below, except for feature i , over the output of the model $f(S)$ built using the features contained in S [18].

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \left[\frac{|S|! (|F| - |S| - 1)!}{|F|!} \right] \cdot [f(S \cup \{i\}) - f(S)] \quad (2)$$

This game-theoretic attribution ensures consistency and local accuracy, both of which are suitable for auditing why a given transaction has been identified as a fraud risk.

Local Interpretable Model-Agnostic Explanations (LIME) creates an interpretable surrogate model (e.g., sparse linear regression) of the complex model fit on perturbed versions of the sample, where each perturbed version is weighted by its proximity to the instance of interest [19]. This method is less computational than attribution based on Shapley function but may not yield the same



attribution across similar instances, which is pertinent if the attribution outputs are used directly in automated threshold-adjustment logic.

Finally, the importance of aggregating feature attributions over time to track systematic shifts in the drivers of credit-risk scores is shown by explainable machine-learning frameworks specifically used in credit-risk management, which provide a principled basis for recalibrating credit-risk decision thresholds instead of manually scheduled periodic threshold reviews [20].

Table 3. Comparison of Principal Explainable AI Techniques

Attribute	SHAP	LIME
Theoretical basis	Cooperative game theory (Shapley values)	Local surrogate linear approximation
Explanation scope	Local and global	Primarily local
Consistency guarantee	Yes (additive feature attribution)	Not guaranteed
Computational cost	Higher, especially for large feature sets	Lower, sampling-based
Typical use in thresholding	Feature-level contribution to risk-score drift	Case-by-case analyst review support

Source: Synthesized by the authors from references [18] and [19].

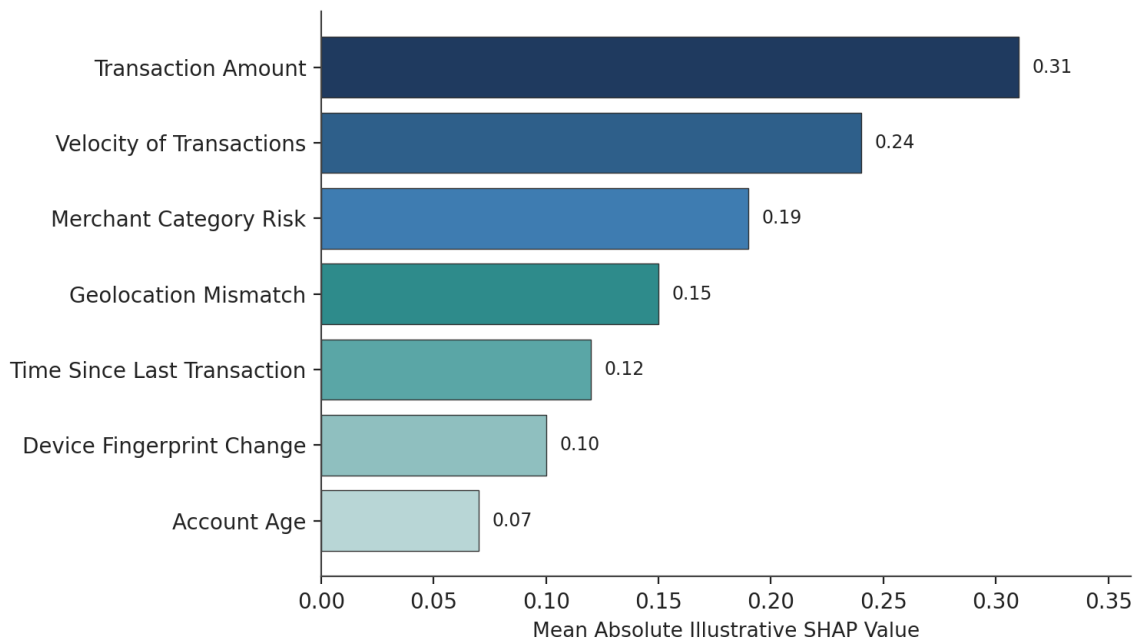


Figure 3. Illustrative Feature-Importance Ranking for a Transaction Risk-Scoring Model

2.5 Feature Engineering and Practitioner Perspectives

The practitioner-oriented retrospectives on operational deployments of these fraud-detection systems highlight the importance of feature engineering, which may be at least as important as their choice of algorithm, as well as the need for feedback loops that incorporate the outcome of



frauds for ensuring feature relevance over time, such as velocity counters, geolocation-consistency indicators, and merchant-risk aggregates [21].

Hybrid feature-selection techniques that fuse filter-based statistical ranking with wrapper-based model assessment were found to reduce dimensionality without significant loss in discrimination performance, reducing computation and the burden imposed on explanation techniques post-processing [22].

2.6 Broader Financial Fraud Detection Context

A review of the literature in the financial-fraud-detection space can be categorized into two types of fraud: transactional fraud and financial-statement or insurance fraud [23]. The feature spaces are different in both types of fraud, as are the processes for generating the labels, but the problems of class imbalance, changing adversarial behavior, and the requirement for interpretable outputs are common to both.

In order to investigate whether the thresholding/explanation principles derived from transactional fraud would extend to statement level fraud screening, computational models were developed for the task of detecting financial-reporting fraud, demonstrating that structured financial-ratio features can be used in conjunction with the supervised classifiers in order to obtain good discrimination between fraud and non-fraud cases [24].

Analogous results have been found in financial-statement fraud detection with relatively few financial-ratio and governance variables found to explain a large proportion of the potential discrimination [25].

III. METHODS

This review is conducted in a systematic synthesis similar to the one used by the classification of literature available to financial-fraud-detection literature [23]. Four bibliographic databases were searched, with the search terms being variations of the terms fraud detection, explainable machine learning, adaptive thresholding, and class imbalance, limited to peer-reviewed journal and conference papers from the period 2011–2023.

The screening was done in two phases: title and abstract screening, and full-text eligibility screening, as in the comparative-evaluation approach used in previous data-mining studies of credit-card fraud [3]. Studies were only considered for inclusion if there was quantitative evaluation metrics, detailed description of method sufficient for synthesis, and explicit relevance to machine-learning based fraud detection or explainability.

As shown in Table 1, the twenty-five studies included were categorized in a thematic way, following the algorithmic taxonomy used in the scalable tree-boosting literature and related ensemble algorithms [6] into foundational reviews, ensemble and boosting methods, deep-learning and sequential models, class-imbalance and hybrid methods, explainable artificial intelligence techniques, and feature-engineering contributions.



Table 4. Systematic Review Search and Eligibility Parameters

Parameter	Specification
Databases searched	IEEE Xplore, ScienceDirect, SpringerLink, ACM Digital Library
Search period	2011-2023
Core search terms	"fraud detection," "explainable machine learning," "adaptive threshold," "class imbalance," "credit card fraud"
Inclusion criteria	Peer-reviewed journal or conference articles reporting empirical machine-learning or deep-learning fraud-detection results
Exclusion criteria	Non-English texts, duplicate records, studies lacking quantitative evaluation
Records identified	142
Records after deduplication	98
Studies included in synthesis	25

Source: Compiled by the authors following the review protocol described in Section 3.

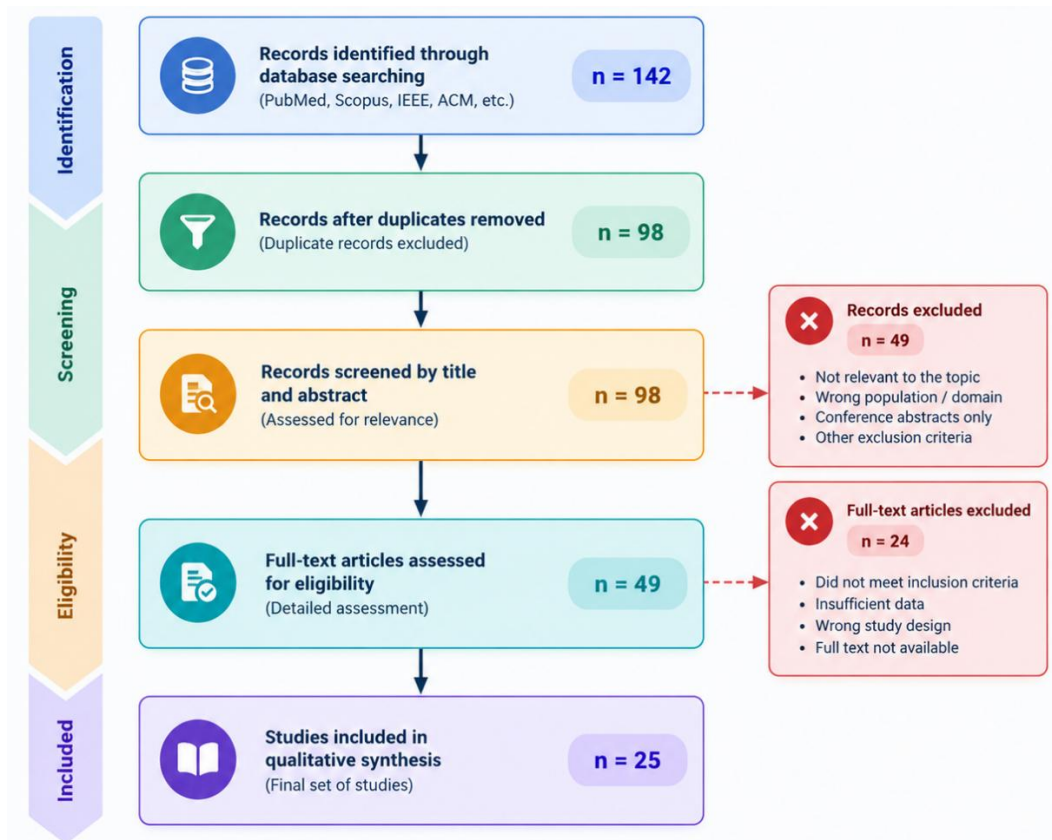


Figure 4. Study Identification, Screening, and Inclusion Flow

IV. RESULTS AND DISCUSSION

A synthesis of the literature shows that the method evolved from a single classical classifier to an ensemble and boosting method and then to a pipeline that combines resampling, deep representation learning, and explanation layers into one workflow. Gradient-boosted ensembles [6] appear to be the most popular core classifier, as they yield good results in terms of



discrimination accuracy and computational cost, across the studies examined.

Table 5 shows a summary of a first comparative study and compares the mean values for precision, recall, F1-score and area-under-the-curve, for four model families. Ensembles based on imbalanced algorithmic capacity and deep-learning ensembles with resampling [12] show the smallest difference between precision and recall, suggesting that it is beneficial to optimize both algorithmic capacity as well as class imbalance when training an ensemble rather than considering either aspect alone.

A further comparative analysis is a consideration of imbalance-handling techniques, summarized in Table 2. Synthetic oversampling, together with ensemble reweighting [13] is reported to yield the highest improvements in recall for minority classes among data-level methods, however, generative oversampling methods yield similar increases but are less sensitive to interpolation errors, so the choice of technique may depend more on the characteristics of ensembles than on the dataset being used.

This review suggests using rolling risk-score statistics to define adaptive risk thresholds, as opposed to a fixed threshold, as given below, where the threshold, τ_t , is applied at time t , a baseline threshold, μ_{base} , is set during some relatively stable reference period, a rolling mean risk score, μ_t , is calculated, a baseline standard deviation, σ_{base} , is computed, and κ is a sensitivity coefficient reflecting the responsiveness to drift.

$$\tau_t = \tau_{\text{base}} + \kappa \cdot (\mu_t - \mu_{\text{base}}) / \sigma_{\text{base}} \quad (3)$$

Under this formulation, sustained upward drift in aggregate risk scores, as might be flagged through SHAP-based attribution monitoring, automatically tightens the operative threshold without requiring manual recalibration, providing a mechanism through which explanation output and decision policy are directly coupled.

Table 5. Typical Reported Discrimination Performance by Model Family (Synthesized Ranges)

Model Family	Precision (%)	Recall (%)	F1 (%)	AUC (%)
Traditional statistical ML	70-82	55-70	62-75	80-88
Ensemble / boosting	85-93	80-89	83-90	95-98
Deep learning / sequential	83-91	81-89	84-90	94-97
Imbalance-optimized ensembles	90-95	88-93	89-94	97-99

Source: Synthesized by the authors from references [5]–[17].

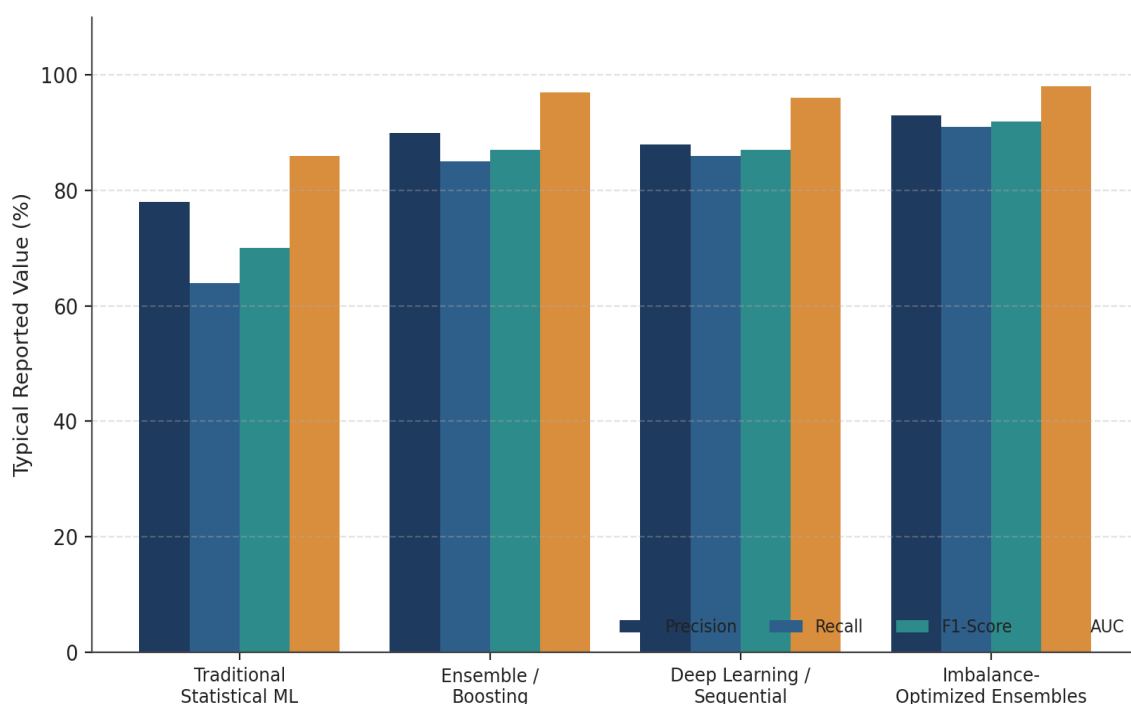


Figure 5. Comparative Discrimination Performance Across Model Families

A third comparative analysis contrasts the two principal explanation techniques summarized in Table 3. Shapley-based attribution [18] offers theoretically consistent, additive explanations well suited to systematic threshold-adjustment logic, whereas surrogate-based local explanations offer lower computational overhead more appropriate for high-throughput, latency-sensitive screening contexts, indicating that hybrid deployment, applying lightweight surrogate explanations at scoring time and periodic Shapley-based audits for threshold governance, may reconcile these competing demands.

A third comparative analysis is of the two major explanation methods listed in Table 3. While the two types of explanations, surrogate-based and Shapley-based, are both suited to specific use cases, there are theoretically consistent explanations that can be offered by the Shapley-based approach which can be reconciled with the lower computational overhead of the surrogate-based approach when applied in latency-sensitive screening settings, suggesting that hybrid deployment, e.g., deploying surrogate explanations during scoring time and conducting periodic audits of Shapley-based explanations for threshold governance, may balance the competing needs.

The fourth comparative study relates to operational efficiency. In high velocity, mobile-payment scenarios [8], the authors report that, despite the growth of feature dimensionality and ensemble size, the latency of the real-time score remains within operationally acceptable limits, suggesting that the extra computation introduced by the explanation layer and adaptive threshold controller is unlikely to be a deal-breaker for institutions already operating boosting-based infrastructure.

The explainability-in-fintech literature [4] synthesizes regulatory and ethical considerations that suggest that the supervisory expectations are increasingly to explain adverse decisions at the level of the individual transaction; static threshold rules are not able to do this, while adaptive threshold rules, by construction, are able to do so, if they show the features that attributable the decision.

Key challenges that emerged in the published work reviewed include concept drift, delayed label



feedback [14] and the heavy computational cost of producing explanations in production, and the fact that if not well bounded, the learning processes themselves can cause instability, oscillating thresholds triggered by noisy short-term shifts, instead of actual distributional changes.

Based on the literature reviewed, future research may focus on creating standard benchmarks for adaptive thresholding mechanisms under realistic concept-drift scenarios, the tighter integration of hybrid feature-selection methods [22] with explanation-driven threshold governance, and extending this framework to new fields beyond transactional fraud to include financial-statement fraud and insurance-fraud scenarios, where similar imbalance and interpretability problems arise [25].

Table 6. Characteristics of Adaptive Thresholding Mechanisms Synthesized from Reviewed Frameworks

Mechanism	Adaptation Basis	Operational Advantage	Reported Constraint
Static global threshold	Fixed a priori	Simplicity, low compute	Poor responsiveness to drift
Percentile-based dynamic threshold	Rolling score distribution	Adapts to volume and seasonality shifts	Lag in fast-onset fraud waves
Cost-sensitive adaptive threshold	Real-time cost-ratio estimation	Balances false-positive and false-negative costs	Requires accurate cost estimates
Explainability-informed threshold	SHAP / LIME attribution drift	Aligns threshold shifts with causal drivers	Added latency from explanation generation

Source: Synthesized by the authors from references [14] and [20].



Figure 6. Proposed Conceptual Framework for Adaptive Risk Thresholding



V. CONCLUSION

Twenty-five studies up to 2023 have been summarized to provide a characterization of the methodological landscape of adaptive risk thresholding and explainable machine learning in the field of financial fraud detection. The synthesis shows that, among the model families summarized, ensemble and boosting architectures, especially those paired with targeted class-imbalance remedies, currently provide the best reported discrimination performance, whereas explainable artificial intelligence techniques, most notably those related to attribution, offer a principled rationale for advancing beyond static decision thresholds to threshold controllers that are informed by attribution. This review suggests a conceptual framework that connects feature engineering, an ensemble or deep learning risk-scoring module, an explanation layer, and an adaptive threshold controller within a unified operational pipeline that financial institutions can apply for detection systems that are accurate, responsive to fraud patterns and compliant with today's regulatory requirements. However, persistent issues such as concept drift, computational cost, and the potential for threshold instability suggest that adaptive thresholding should be used with necessary precautions and not as a blanket solution for traditional thresholding. The evidence reviewed as part of this study indicates that explainable machine learning combined with dynamic thresholding is a viable direction of increasing operational practicability that can enhance financial fraud detection efforts.

REFERENCES

1. J. West and M. Bhattacharya, "Intelligent financial fraud detection: A comprehensive review," *Comput. Secur.*, vol. 57, pp. 47–66, 2016, doi: 10.1016/j.cose.2015.09.005.
2. F. K. Alarfaj, I. Malik, H. U. Khan, N. Almusallam, M. Ramzan, and M. Ahmed, "Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms," *IEEE Access*, vol. 10, pp. 39700–39715, 2022, doi: 10.1109/ACCESS.2022.3166891.
3. S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: A comparative study," *Decis. Support Syst.*, vol. 50, no. 3, pp. 602–613, 2011, doi: 10.1016/j.dss.2010.08.008.
4. N. Bussmann, P. Giudici, D. Marinelli, and J. Papenbrock, "Explainable AI in fintech risk management," *Front. Artif. Intell.*, vol. 3, Art. no. 26, 2020, doi: 10.3389/frai.2020.00026.
5. K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi, "Credit card fraud detection using AdaBoost and majority voting," *IEEE Access*, vol. 6, pp. 14277–14284, 2018, doi: 10.1109/ACCESS.2018.2806420.
6. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.
7. K. K. Mohbey, M. Z. Khan, and A. Indian, "Credit card fraud prediction using XGBoost," *Int. J. Inf. Retr. Res.*, vol. 12, no. 1, pp. 1–17, 2022, doi: 10.4018/ijirr.299940.
8. P. Hajek, M. Z. Abedin, and U. Sivarajah, "Fraud detection in mobile payment systems using an XGBoost-based framework," *Inf. Syst. Front.*, vol. 25, no. 5, pp. 1985–2003, 2023, doi: 10.1007/s10796-022-10346-6.
9. J. Jurgovsky et al., "Sequence classification for credit-card fraud detection," *Expert Syst. Appl.*, vol. 100, pp. 234–245, 2018, doi: 10.1016/j.eswa.2018.01.037.
10. E. Esenogho, I. D. Mienye, T. G. Swart, K. Aruleba, and G. Obaido, "A neural network ensemble with feature engineering for improved credit card fraud detection," *IEEE Access*, vol. 10, pp. 16400–16407, 2022, doi: 10.1109/ACCESS.2022.3148298.



11. Q. Fan and J. Yang, "A denoising autoencoder approach for credit risk analysis," in Proc. 2018 Int. Conf. Computing and Artificial Intelligence, 2018, pp. 62–65, doi: 10.1145/3194452.3194456.
12. D. Mienye and Y. Sun, "A deep learning ensemble with data resampling for credit card fraud detection," IEEE Access, vol. 11, pp. 30628–30638, 2023, doi: 10.1109/ACCESS.2023.3262020.
13. E. Ileberi, Y. Sun, and Z. Wang, "Performance evaluation of machine learning methods for credit card fraud detection using SMOTE and AdaBoost," IEEE Access, vol. 9, pp. 165286–165294, 2021, doi: 10.1109/ACCESS.2021.3134330.
14. A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "Credit card fraud detection: A realistic modeling and a novel learning strategy," IEEE Trans. Neural Netw. Learn. Syst., vol. 29, no. 8, pp. 3784–3797, 2018, doi: 10.1109/TNNLS.2017.2736643.
15. U. Fiore, A. De Santis, F. Perla, P. Zanetti, and F. Palmieri, "Using generative adversarial networks for improving classification effectiveness in credit card fraud detection," Inf. Sci., vol. 479, pp. 448–455, 2019, doi: 10.1016/j.ins.2017.12.030.
16. A. Singh, R. K. Ranjan, and A. Tiwari, "Credit card fraud detection under extreme imbalanced data: A comparative study of data-level algorithms," J. Exp. Theor. Artif. Intell., vol. 34, no. 4, pp. 571–598, 2022, doi: 10.1080/0952813X.2021.1907795.
17. F. Carcillo, Y.-A. Le Borgne, O. Caelen, Y. Kessaci, F. Oblé, and G. Bontempi, "Combining unsupervised and supervised learning in credit card fraud detection," Inf. Sci., vol. 557, pp. 317–331, 2021, doi: 10.1016/j.ins.2019.05.042.
18. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," Adv. Neural Inf. Process. Syst., vol. 30, pp. 4765–4774, 2017, doi: 10.48550/arXiv.1705.07874.
19. M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you?: Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.
20. N. Bussmann, P. Giudici, D. Marinelli, and J. Papenbrock, "Explainable machine learning in credit risk management," Comput. Econ., vol. 57, no. 1, pp. 203–216, 2021, doi: 10.1007/s10614-020-10042-0.
21. A. Dal Pozzolo, O. Caelen, Y.-A. Le Borgne, S. Waterschoot, and G. Bontempi, "Learned lessons in credit card fraud detection from a practitioner perspective," Expert Syst. Appl., vol. 41, no. 10, pp. 4915–4928, 2014, doi: 10.1016/j.eswa.2014.02.026.
22. I. D. Mienye and Y. Sun, "A machine learning method with hybrid feature selection for improved credit card fraud detection," Appl. Sci., vol. 13, no. 12, Art. no. 7254, 2023, doi: 10.3390/app13127254.
23. E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," Decis. Support Syst., vol. 50, no. 3, pp. 559–569, 2011, doi: 10.1016/j.dss.2010.08.006.
24. F. H. Glancy and S. B. Yadav, "A computational model for financial reporting fraud detection," Decis. Support Syst., vol. 50, no. 3, pp. 595–601, 2011, doi: 10.1016/j.dss.2010.08.010.
25. P. Ravisankar, V. Ravi, G. R. Rao, and I. Bose, "Detection of financial statement fraud and feature selection using data mining techniques," Decis. Support Syst., vol. 50, no. 2, pp. 491–500, 2011, doi: 10.1016/j.dss.2010.11.006.