Journal of Mathematical Finance and Risk Management
ISSN: 3080-9983 (Print), ISSN: 3080-9991 (Online) | Vol. 2, No. 1, 2026

Explainable Machine Learning for Detecting Irregular Patterns in Corporate Bond Markets

Francesca Romano [1], Lorenzo Bianchi [2, *], Elena Conti [1], Davide Marchetti [2]

[1] Department of Computer Science, University of Milan, Milan, Italy

[2] Department of Computer Science, University of Pisa, Pisa, Italy

Abstract: Corporate bond markets periodically display irregular behavior, ranging from abrupt spread repricing to breakdowns in the usual relations between rating segments and Treasury yields, while the indicators used for monitoring are mostly univa riate and say little about why a market state is unusual. This paper develops an explainable machine-learning framework that detects such irregular patterns and attributes each alarm to its economic drivers. Using daily U.S. data from 1999 to 2024 on ICE BofA high-yield and BB option-adjusted spreads, Moody's seasoned Aaa and Baa yield spreads, and Treasury yields, fifteen trailing-window features are organized into five interpretable families. An Isolation Forest is re-estimated annually in a walk-forward design and compared with six baseline detectors, including a spread z-score rule. Alarms are explained with exact TreeSHAP attributions aggregated to feature families and summarized by a depth-three surrogate tree. Against 14 documented stress episodes over 2004–2024, the detector reaches an AUC-PR of 0.546 and an alarm precision of 0.715 at a false-alarm rate of 1.7%, and it attains the highest mean AUC (0.935) on four types of injected irregularities. The explanations separate episodes by mechanism: the 2013 taper tantrum is dominated by rates–credit linkage features, the COVID-19 shock by spread volatility and momentum, and the 2008 Bear Stearns episode by cross-segment dislocation. The surrogate rules reproduce the alarms of the detector with 98.1% accuracy.

Keywords: Corporate bonds, credit spreads, anomaly detection, isolation forest, explainable artificial intelligence, Shapley values.

1. Introduction

The U.S. corporate bond market is a primary source of external finance for non-financial firms, and its pricing conditions carry information about default risk, investor risk appetite and future economic activity [1–5]. Credit spreads also contain a sizable non-default component related to liquidity and to the risk-bearing capacity of intermediaries [6–8]. Because these components move at different speeds, stress does not always appear as a smooth widening of spreads. It can take the form of an abrupt repricing within a few days, an unusual divergence between rating segments, a burst of volatility, or a temporary breakdown of the usual co-movement between credit spreads and Treasury yields.

Recent episodes illustrate this variety. During the global financial crisis, funding and market liquidity evaporated and the illiquidity component of corporate bond spreads increased sharply [9, 10]. In March 2020, investment-grade bonds traded at deep discounts relative to comparable credit default swaps and dealers withdrew from market making until the Federal Reserve announced its corporate credit facilities [11–13], while outflows from bond funds amplified the sell-off [14, 15]. The 2013 taper tantrum and the 2022 tightening cycle, by contrast, originated in the rates market rather than in credit. A monitoring tool that only tracks spread levels therefore misses irregularities that do not appear as level shifts, and it gives no indication of the nature of the irregularities it does flag.

Supervisors and risk managers rely on composite stress indices to summarize market conditions [16, 17], and a dedicated corporate bond market distress index has recently been proposed [18]. Machine learning has meanwhile become a standard tool in empirical finance, including bond return prediction [19–22] and financial crisis prediction [23]. Unsupervised anomaly detection is a natural fit for irregularity monitoring because stress episodes are rare and their forms change over time [24, 25]. Its adoption in financial supervision and risk management depends on whether alarms can be explained in economically meaningful terms [26, 27].

This paper develops an explainable framework for detecting irregular patterns in the U.S. corporate bond market by combining interpretable features, a walk-forward Isolation Forest [28] and exact Shapley-value attributions [29–31]. The contributions are threefold. First, fifteen trailing-window features grouped into five families (level, dynamics, volatility, cross-segment structure and rates–credit linkage) are built from public daily data, so that every alarm can be attributed to drivers with a clear market interpretation. Second, the detector is evaluated against 14 documented stress episodes between 2004 and 2024 and against four types of injected irregularities, alongside six baseline detectors. Third, the attributions are aggregated into family-level episode profiles and summarized by surrogate rules, which reveal how credit-driven, rates-driven and segment-specific episodes differ. Fig. 1 summarizes the framework.

Section II reviews related work, Sections III and IV present the data and methodology, Section V describes the experimental design, Section VI reports the results, and Sections VII and VIII discuss and conclude.

2. Related Work

2.1. Corporate Bond Spreads and Market Stress

Structural models link credit spreads to leverage and asset volatility [1], yet empirical decompositions show that expected default explains only part of the level and of the changes of spreads [2, 3]. A large literature documents the role of liquidity [6–8], 10] and of trading costs, which declined

after the dissemination of TRACE transaction data [32, 33]. Intermediation capacity is a key channel in stress periods: post-crisis regulation reduced dealers' willingness to absorb customer selling [34], and investor flows can turn bond funds into a source of fragility [14, 15]. Studies of individual events, including the General Motors and Ford downgrades of 2005 [35], the euro-area debt crisis [36], the 2010 flash crash [37] and the 2023 regional bank failures [38], document the market dynamics behind the episode calendar used in this study. Regime-switching models [39] and stress indices [16–18] are the classical tools for identifying stressed market states.

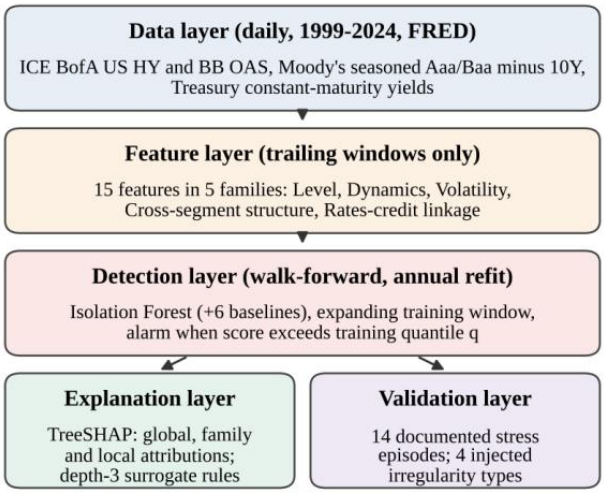


Figure 1. Overview of the explainable framework for detecting irregular patterns in corporate bond markets.

2.2. Anomaly Detection in Financial Time Series

Anomaly detection methods are commonly grouped into density-, distance-, boundary-, reconstruction- and isolation-based approaches [24, 25, 40]. The Local Outlier Factor (LOF) compares local densities [41], the one-class support vector machine estimates the support of normal data [42], robust Mahalanobis distances build on the minimum covariance determinant [43], and principal component or autoencoder reconstructions flag points that the normal subspace cannot represent [44, 45]. Isolation Forest separates observations with random axis-parallel splits and scores them by their average path length [28, 46]; it requires little tuning, scales linearly with the sample size and performs competitively in benchmark studies [47–49]. Open-source toolkits have made these detectors widely accessible [50, 51]. In finance, anomaly detection has been applied mainly to fraud and market manipulation [52–54], whereas the machine-learning literature on crises has largely adopted a supervised early-warning formulation [23]. Unsupervised and explainable detection of irregular states of the corporate bond market as a whole has received comparatively little attention.

2.3. Explainable Machine Learning

Post-hoc explanation methods include local surrogate models such as LIME [55] and additive feature attributions based on the Shapley value [29, 30, 56], for which TreeSHAP provides exact solutions for tree ensembles in polynomial time [31]. Surveys organize the rapidly growing set of techniques [27, 57], and methodological work clarifies how attributions should be read when features are correlated or when explanations are used as importance measures [26, 58]. Global surrogate models that mimic a black box with a compact decision tree date back to [59]. Explainable anomaly detection adapts these ideas to unsupervised scores, including depth-based feature importance for Isolation Forest [60], Shapley values of reconstruction errors [61, 62] and broader taxonomies [63]. In finance, Shapley-based explanations have been used for credit risk [64] and crisis prediction [23]. The present study applies exact TreeSHAP to an Isolation Forest that monitors corporate bond conditions and aggregates the attributions into economically defined feature families.

3. Data and Feature Construction

3.1. Data

The analysis uses daily series from the FRED database of the Federal Reserve Bank of St. Louis, listed in Table I. ICE BofA option-adjusted spreads (OAS) cover the U.S. high-yield (HY) index and its BB-rated segment, Moody's seasoned Aaa and Baa corporate yields are measured relative to the 10-year Treasury yield, and Treasury constant-maturity yields describe the risk-free curve. After aligning trading days, the sample contains 6,313 observations from January 4, 1999 to May 30, 2024. The HY OAS ranges from 2.41% to 21.82%, with its maximum on December 15, 2008, and the Baa–10Y spread peaks at 6.16% on December 4, 2008 (Fig. 2).

Table I. Daily Series Used in the Study (Jan. 1999–May 2024)

Series	FRED ID	Mean (%)	Range (%)
ICE BofA US HY OAS	BAMLH0A0HYM2	5.49	2.41–21.82
ICE BofA BB OAS	BAMLH0A1HYBB	3.65	1.65–14.68
Moody's Baa – 10Y	BAA10Y	2.51	1.40–6.16
Moody's Aaa – 10Y	AAA10Y	1.50	0.63–3.20
10-year Treasury	DGS10	3.36	0.52–6.79
1-year Treasury	DGS1	2.09	0.04–6.44

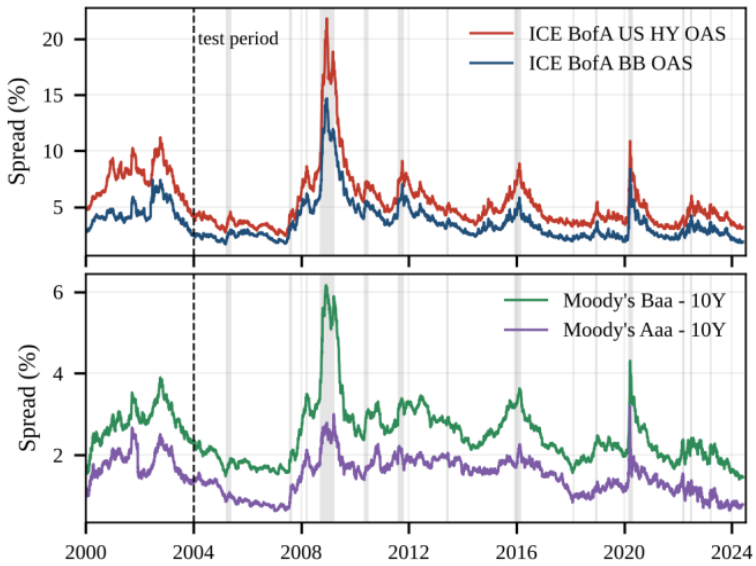


Figure 2. Daily corporate bond spreads, 2000–2024. Shaded areas mark the documented stress episodes of Table III, and the dashed line marks the start of the walk-forward test period.

3.2. Feature Families

Irregularity is defined relative to recent market conditions, so every feature is computed from trailing windows that end on the day being scored. For a spread or yield v_t measured in basis points, level features are rolling z-scores

$$z_t = (v_t - \mu_{t,W}) / \sigma_{t,W} \quad (1)$$

Where $\mu_{t,W}$ and $\sigma_{t,W}$ are the mean and standard deviation over the previous W = 252 trading days. Dynamics features are five-day log changes of spreads, and volatility features are 21-day standard deviations of daily changes together with the ratio of short- to long-run HY volatility. Cross-segment features capture the HY–BB quality spread, the Baa–Aaa default premium and the 63-day correlation between HY and Baa spread changes. Rates-linkage features capture five-day Treasury shocks, the 63-day correlation between HY spread changes and 10-year yield changes, and the slope of the yield curve. Table II lists the fifteen features. Complete feature vectors are available from January 7, 2000, giving 6,061 daily observations.

Table II. Feature Families and Definitions

Family	Symbol	Definition
Level	z_{HY}	Rolling z-score of the HY OAS (W = 252)
	z_{Baa}	Rolling z-score of the Baa–10Y spread
	z_{Aaa}	Rolling z-score of the Aaa–10Y spread
Dynamics	r_{HY}	Five-day log change of the HY OAS (%)
	r_{BB}	Five-day log change of the BB OAS (%)
	r_{Baa}	Five-day log change of the Baa–10Y spread (%)
Volatility	σ_{HY}	21-day s.d. of daily HY OAS log changes
	σ_{Baa}	21-day s.d. of daily Baa–10Y changes (bp)
	v_{HY}	Ratio of 21-day to 252-day HY volatility
Cross-segment	z_{QS}	z-score of the HY–BB quality spread
	z_{DP}	z-score of the Baa–Aaa default premium
	$\rho_{HY,Baa}$	63-day correlation of daily HY and Baa–10Y changes
Rates linkage	$\Delta_5 y_{10}$	Five-day change of the 10Y yield (bp)
	$\rho_{HY,10Y}$	63-day correlation of daily HY OAS and 10Y yield changes
	z_{slope}	z-score of the 10Y–1Y slope

3.3. Documented Stress Episodes

To evaluate alarms on real data, Table III compiles 14 stress episodes in the 2004–2024 test period from the literature and from public event chronologies. Each window starts at the triggering event and covers the main repricing phase. Together, the windows contain 588 of the 5,075 test days (11.6%). The calendar spans credit events, liquidity shocks and rate shocks, which allows the explanation profiles of different episode types to be compared.

Table III. Documented Stress Episodes in the Test Period

No.	Episode	Window
1	GM/Ford downgrades [35]	16 Mar – 31 May 2005
2	Subprime turmoil onset [9]	16 Jul – 31 Aug 2007
3	Bear Stearns rescue	3 Mar – 31 Mar 2008
4	Lehman failure / GFC peak [10]	15 Sep 2008 – 31 Mar 2009
5	Euro crisis I / flash crash [36, 37]	3 May – 30 Jun 2010
6	U.S. downgrade / euro crisis II	1 Aug – 31 Oct 2011
7	Taper tantrum	22 May – 28 Jun 2013
8	Oil collapse / HY energy stress	1 Dec 2015 – 29 Feb 2016
9	Volatility shock	2 Feb – 28 Feb 2018
10	Q4-2018 risk-off	3 Dec – 31 Dec 2018
11	COVID-19 shock [11–13]	24 Feb – 30 Apr 2020
12	Russia–Ukraine invasion	24 Feb – 31 Mar 2022
13	Mid-2022 tightening	9 Jun – 15 Jul 2022
14	Regional bank failures [38]	9 Mar – 31 Mar 2023

4. Methodology

4.1. Walk-Forward Protocol

Let $x_t \in \mathbb{R}^{15}$ denote the feature vector on day t. For each test year $y \in \{2004, \ldots, 2024\}$, every detector is trained on all observations before January 1 of year y and applied to the trading days of year y. This expanding-window protocol reproduces real-time monitoring: no future information enters a model, and each annual refit incorporates the most recent episodes into the reference distribution.

4.2. Isolation Forest

Isolation Forest grows T random trees on subsamples of size ψ, choosing at each node a feature and a split value uniformly at random. Irregular observations are isolated after fewer splits, and the anomaly score is

$$s(x) = 2^{-E[h(x)]/c(\psi)} \quad (2)$$

Where $h(x)$ is the path length of x in a tree and $c(\psi) = 2H(\psi - 1) - 2(\psi - 1)/\psi$ is the average path length of an unsuccessful search in a binary search tree, with $H(\cdot)$ the harmonic number [28], [46]. We use $T = 300$ and $\psi = 256$. Because split values are drawn uniformly over the range of each feature, the score is invariant to affine rescaling of individual features, so the detector operates directly on interpretable raw feature values.

4.3. Adaptive Alarm Threshold

Score scales differ across detectors and annual refits. Each test score is therefore normalized with the training-score distribution of the same model,

$$\tilde{s}_t = (s_t - m_y) / (\tau_y - m_y) \quad (3)$$

Where m_y is the median and τ_y the q-quantile of the training scores for year y. An alarm is raised when $\tilde{s}_t > 1$. The default $q = 0.975$ implies an alarm rate of about 2.5% under conditions resembling the training sample; Section VI-C reports the full trade-off.

4.4. Shapley Attributions

TreeSHAP decomposes the ensemble output for a single day into additive feature contributions [30, 31],

$$E[h(x_t)] = \varphi_0 + \Sigma_{j=1}^{15} \varphi_j(x_t) \quad (4)$$

Where φ_0 is the expected path length over the training data. Since shorter paths indicate irregularity, we define the anomaly contribution $\varphi^{A}_j = -\varphi_j$, so that positive values push a day toward an alarm. For all 21 trained forests, the attributions reproduce the path lengths implied by the model scores with a maximum absolute error of 2.1×10^{-14}. For economically interpretable summaries, positive contributions are aggregated over the features of family F_k and over a set of days D,

$$S_k(D) = \Sigma_{t \in D} \Sigma_{j \in Fk} [\varphi^{A}_j(x_t)]^+ / \Sigma_{t \in D} \Sigma_j [\varphi^{A}_j(x_t)]^+ \quad (5)$$

Where $[\cdot]^+ = \max(\cdot, 0)$. For an episode, D contains its alarm days; for an episode without alarms, the five highest-scoring days in its window are used.

4.5. Surrogate Rules and Baselines

As a global summary, a decision tree of depth three is fitted to the alarm indicator of the Isolation Forest over the test period [59]. Its fidelity, measured by accuracy and F1-score with respect to the detector's own alarms, shows how well a few readable rules reproduce the decisions of the detector. Six baselines are trained under the same protocol: LOF with $k = 50$ neighbors [41], a one-class SVM (OCSVM) with an RBF kernel and $\nu = 0.05$ [42], the PCA reconstruction error retaining 90% of the variance (PCA-Rec) [44], a robust Mahalanobis distance based on the minimum covariance determinant (R-Maha) [43], an autoencoder with a 12–4–12 tanh architecture (AE) [45], and a practitioner rule (Z-HY) that uses the HY OAS z-score alone. The learned baselines operate on features robustly scaled with training-window statistics, and all detectors use the threshold in (3).

5. Experimental Setup

5.1. Evaluation on Documented Episodes

Days inside the windows of Table III are labeled positive. Threshold-free performance is measured by the area under the ROC curve (AUC-ROC) and the average precision (AUC-PR) of the normalized scores. At the alarm threshold, the hit rate is the share of episodes with at least one alarm in their window, precision is the share of alarms that fall inside a window, the false-alarm rate (FAR) is the share of non-episode days with an alarm, and the detection delay is the number of trading days from the start of an episode to its first alarm.

5.2. Injected Irregularities

Because documented episodes are dominated by spread

widening, four types of irregularity are injected into the raw series on calm test days, defined as days at least 30 trading days away from any episode window. A spread jump raises all four credit spreads by 10% in log terms for five days. A segment dislocation raises the HY OAS by 12% while the BB and investment-grade spreads stay unchanged. A volatility burst adds zero-mean noise with four times the unconditional daily standard deviation to the HY OAS and the Baa–10Y spread for ten days. A rates–credit co-move raises the 10-year yield by 30 bp and the HY and BB spreads by 6% for five days, reversing their usual negative co-movement. For each type, 30 non-overlapping events are drawn with each of three random seeds, features are recomputed on the modified series, and the trained models score the test days. The 21 trading days after each event are excluded from evaluation because trailing windows carry the injected values forward.

5.3. Implementation

The experiments are implemented in Python with scikit-learn [51] and the SHAP library [31]. The complete walk-forward run with seven detectors and 21 annual refits takes about one minute on a single CPU core.

6. Results

6.1. Detection of Documented Episodes

Table IV compares the detectors at q = 0.975. The Isolation Forest (IF) attains the highest AUC-PR (0.546) and the highest alarm precision (0.715) at a false-alarm rate of 1.7%, and it detects 9 of the 14 episodes with a median delay of four trading days. The Z-HY rule reaches a slightly higher AUC-ROC (0.857), as expected for episodes defined largely by spread widening, but it detects fewer episodes (8) with a longer median delay (6.5 days). The two detectors also differ in which episodes they capture: IF flags the taper tantrum and the 2023 bank failures, which the spread rule misses, whereas the spread rule alone detects the slow-building oil-driven episode of 2015–2016. OCSVM and LOF reach a wide coverage only by flagging 22% and 17% of all test days, respectively.

Table IV. Detection of Documented Episodes (q = 0.975)

Detector	AUC-ROC	AUC-PR	Hits	Prec.	FAR	Delay
IF	0.845	0.546	9	0.715	0.017	4.0
LOF	0.724	0.293	10	0.245	0.148	1.5
OCSVM	0.748	0.463	13	0.283	0.180	1.0
PCA-Rec	0.680	0.338	9	0.435	0.044	4.0
R-Maha	0.772	0.424	7	0.512	0.028	2.0
AE	0.762	0.372	9	0.430	0.039	5.0
Z-HY	0.857	0.525	8	0.709	0.012	6.5

Hits: episodes with at least one alarm (out of 14); Delay: median trading days to the first alarm. Best values in bold.

Fig. 3 shows the normalized scores. IF produces concentrated alarm clusters at the onset of the subprime turmoil, after the Lehman failure, during the COVID-19 shock and during the 2022 sell-offs, with peak scores of about 2.5 in July 2007, October 2008 and March 2020. Of its 267 alarms, 84.3% fall within 21 trading days of an episode window. The remaining alarms coincide with market events outside the reference calendar, including the credit turmoil of November–December 2007, the global equity sell-off of January 2008, the Treasury flash rally of October 2014, the oil-driven HY sell-off of December 2014, the market swings of June 2020 and the vaccine announcement of November 2020.

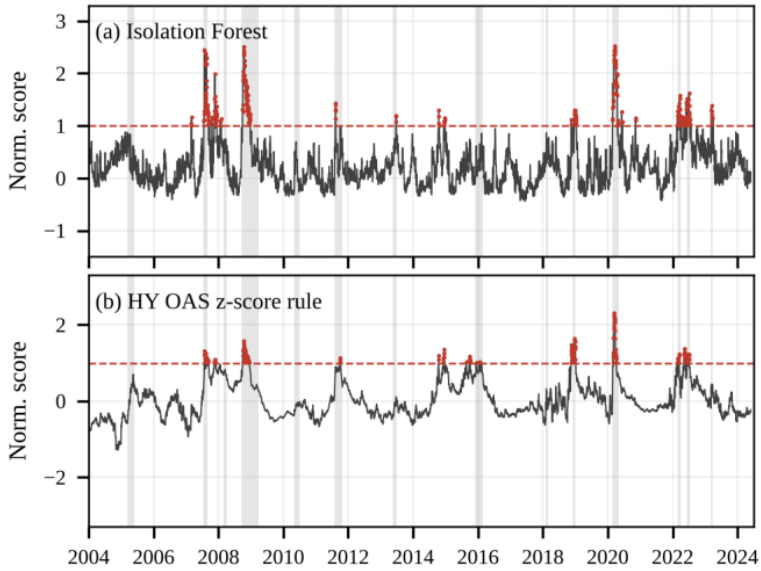


Figure 3. Normalized anomaly scores of (a) the Isolation Forest and (b) the HY OAS z-score rule. Red dots mark alarms ($\tilde{s}t > 1$), and shaded areas mark documented episodes.

6.2. Injected Irregularities

Table V. AUC-ROC on Injected Irregularities (Mean of Three Seeds)

Detector	Jump	Disloc.	Vol. burst	Co-move	Mean
IF	0.972	0.945	0.923	0.899	0.935
LOF	0.764	0.855	0.855	0.705	0.795
OCSVM	0.864	0.912	0.866	0.733	0.844
PCA-Rec	0.820	0.985	0.928	0.735	0.867
R-Maha	0.960	0.993	0.951	0.744	0.912
AE	0.802	0.963	0.934	0.755	0.863
Z-HY	0.717	0.746	0.535	0.648	0.661

Table V reports AUC-ROC on the injected irregularities. IF achieves the highest mean AUC (0.935) and the best results for spread jumps (0.972) and rates–credit co-moves (0.899). The robust Mahalanobis detector is the strongest on segment dislocations (0.993) and volatility bursts (0.951), which appear as departures from the covariance structure of normal days, and its mean AUC-PR (0.611) is close to that of IF (0.608). The co-move type is the most difficult one for all distance- and reconstruction-based detectors, whose AUC values lie between 0.705 and 0.755. These detectors aggregate deviations over all fifteen dimensions, so a moderate shift confined to a few rates-linkage features is diluted, whereas axis-parallel isolation reacts to unusual values of individual features. The Z-HY rule performs poorly on all four types (mean 0.661) and is close to random for volatility bursts (0.535), confirming that level monitoring misses irregularities that leave the spread level largely unchanged. Across seeds, the standard deviation of AUC-ROC is at most 0.033 for every detector and type.

6.3. Threshold Trade-Off and Sensitivity

Fig. 4 traces alarm precision against the episode hit rate as q varies from 0.90 to 0.995. The IF curve lies above the curves of the Z-HY rule and the robust Mahalanobis detector over almost the entire range. At q = 0.95, IF detects all 14 episodes with a precision of 0.480 and a FAR of 6.3%; at q = 0.985, its precision rises to 0.914 with a FAR of 0.3% while 8 episodes are still detected. The Z-HY rule never detects more than 9 episodes in the examined range, because episodes without a pronounced level shift stay below its threshold. Table VI shows that threshold-free performance is stable across rolling-window lengths and forest configurations, with AUC-ROC between 0.838 and 0.864 and AUC-PR between 0.524 and 0.557. A 504-day window raises the hit rate to 11 episodes at a similar FAR.

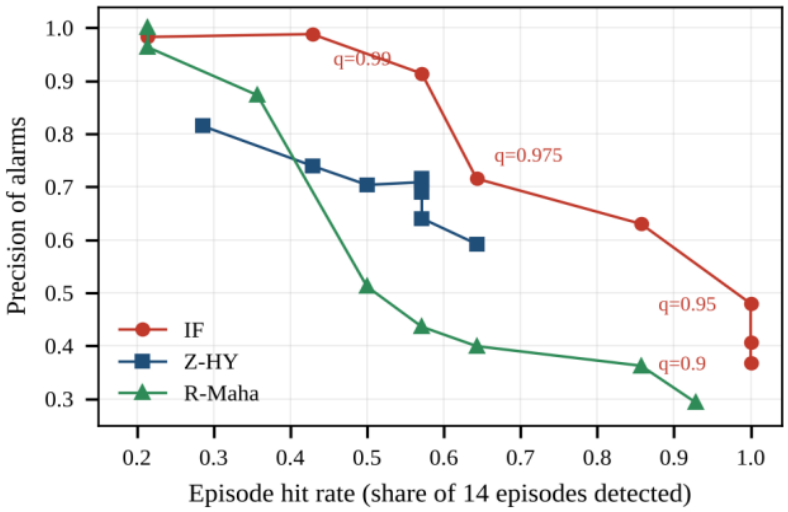


Figure 4. Alarm precision versus episode hit rate as the alarm quantile q varies from 0.90 to 0.995.

Table VI. Sensitivity of the Isolation Forest

Setting	AUC-ROC	AUC-PR	Hits	Prec.	FAR
q = 0.95	0.845	0.549	14	0.480	0.063
q = 0.975 (default)	0.845	0.546	9	0.715	0.017
q = 0.99	0.847	0.552	6	0.988	0.000
W = 126	0.864	0.541	9	0.712	0.014
W = 504	0.849	0.524	11	0.681	0.018
T = 100, ψ = 256	0.849	0.545	9	0.724	0.016
T = 500, ψ = 256	0.851	0.553	9	0.710	0.018
T = 300, ψ = 128	0.851	0.557	10	0.728	0.016
T = 300, ψ = 512	0.838	0.538	10	0.700	0.018

W: rolling window (trading days); T: number of trees; ψ: subsample size.

6.4. Global Explanations

Fig. 5 contrasts two views of global importance. Averaged over all test days, the largest mean absolute attribution belongs to the rates–credit correlation ρHY,10Y, followed by HY OAS volatility and the HY–Baa correlation; these features shape the score landscape in normal times. On alarm days the ranking changes: HY OAS volatility contributes most, with an average of about 0.7 path-length units, followed by the HY OAS z-score, the volatility of the Baa–10Y spread, the five-day HY log change and the HY–BB quality spread. At the family level, alarms are driven by volatility (28.7% of

positive attribution), dynamics (22.6%) and level features (19.2%), while cross-segment (15.2%) and rates-linkage features (14.3%) account for the remainder. For five forests trained with different random seeds on pre-2020 data, the global importance rankings on the 2020–2024 test days have a mean Spearman correlation of 0.917, so the explanations are robust to the randomness of the ensemble.

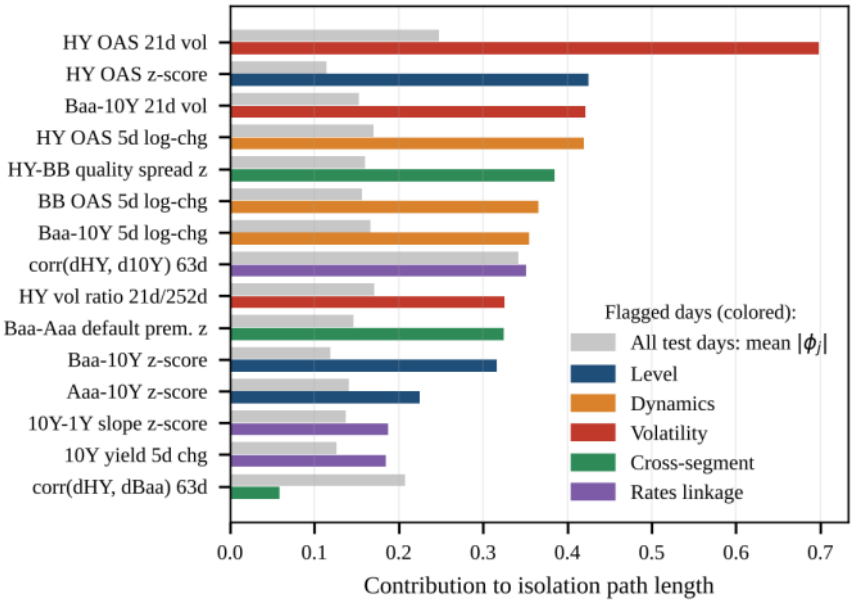


Figure 5. Global attributions. Gray bars show the mean absolute attribution over all test days; colored bars show the mean positive anomaly contribution on alarm days, colored by feature family.

6.5. Episode-Level Explanations

Fig. 6 decomposes the positive attribution of each episode into the five families and reveals distinct mechanisms. The taper tantrum is the clearest rates-driven episode, with 48% of its attribution coming from rates-linkage features, led by $\rho_{HY,10Y}$. The mid-2022 tightening episode also has a large rates-linkage share (29%), whereas the Russia–Ukraine episode is dominated by volatility (39%) with a sizable cross-segment component (19%) through the default premium. The COVID-19 shock combines volatility (35%) and dynamics (26%), reflecting the speed of the repricing. Abrupt and broad sell-offs such as the U.S. downgrade in August 2011 and the 2023 bank failures load mainly on dynamics (51% and 48%). The Bear Stearns episode, which did not trigger an alarm at $q = 0.975$, has the most cross-segment-driven profile (36%), with the HY–BB quality spread as the leading feature, consistent with stress concentrated in lower-rated credit. The Lehman/GFC episode has a balanced profile, with every family contributing between 12% and 28%, as expected for a systemic episode that affected all dimensions of the market.

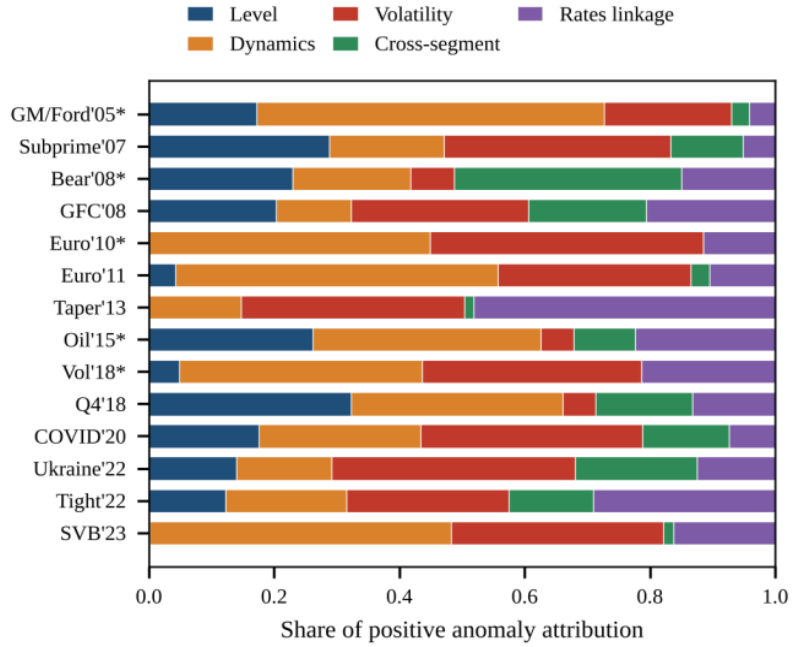


Figure 6. Family composition of positive anomaly attribution by episode. Episodes marked with an asterisk produced no alarm at $q = 0.975$ and are profiled on their five highest-scoring days.

6.6. Local Explanations and Surrogate Rules

Fig. 7 shows local explanations for two alarm days. On June 25, 2013, the alarm is driven by an elevated HY volatility ratio (2.03), a weakening of the usual negative correlation between HY spread changes and Treasury yield changes to

−0.36, against a pre-2013 median of −0.72, a 40 bp rise of the 10-year yield within five days, and a steep yield curve with a slope z-score of 3.66. The HY spread itself had widened only moderately, by 7.4% in five days, so a level-based rule does not single out this day, while the explanation identifies it as a rates-driven irregularity. On March 18, 2020, the leading contributions come instead from an extreme HY OAS z-score (6.93), a 31% five-day widening of the HY OAS, a 24% widening of the Baa–10Y spread and high volatility in both segments, the signature of a broad and fast credit repricing.

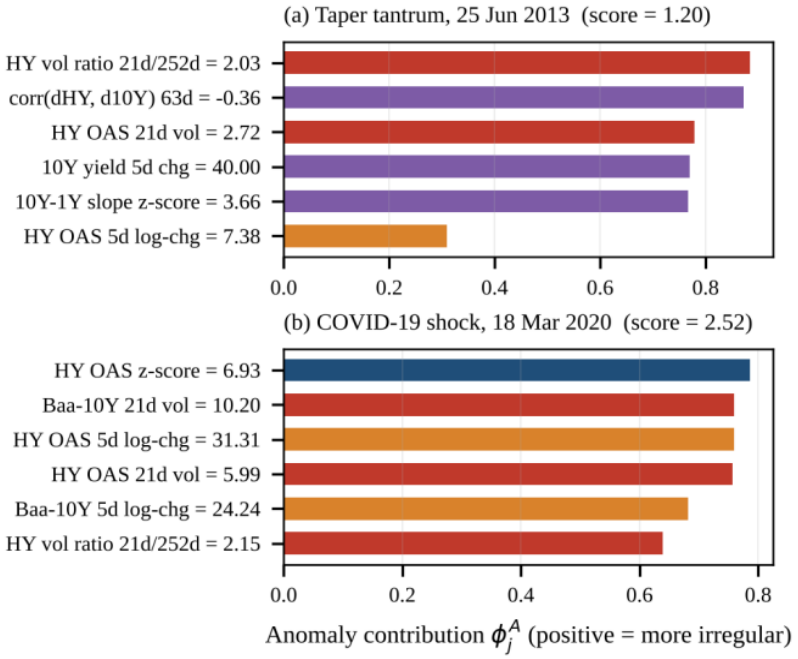


Figure 7. Local explanations for two alarm days. Labels report feature values, and bars show the anomaly contributions φAj, colored by feature family.

The depth-three surrogate tree reproduces the IF alarms with an accuracy of 98.1% and an F1-score of 0.80, and its rules are economically transparent. When the HY OAS z-score exceeds 2.72, a day is flagged if the 21-day volatility of the Baa–10Y spread exceeds 3.92 bp or if the yield-curve slope is close to or above its one-year average. Below that level, a day is flagged if the HY OAS widens by more than 12.6% in five days or if the BB OAS tightens by more than 16.2% in five days. The last rule shows that the detector also treats abrupt compressions as irregular; 21 alarms satisfy it, including April 9–16, 2020, when the Federal Reserve expanded its corporate credit facilities, and November 9–10, 2020, when the first COVID-19 vaccine trial results were announced.

7. Discussion

The results carry three implications for monitoring corporate bond markets. First, spread rules and multivariate detectors are complementary. Spread levels remain the most direct signal of slow-building credit deterioration, as the 2015–2016 energy episode illustrates, whereas the multivariate detector also captures irregularities in volatility, segment structure and the rates–credit relation, as the injection experiments confirm. Second, family-level attributions turn an anomaly score into a diagnosis. Distinguishing a rates-driven episode such as the taper tantrum from a liquidity-driven episode such as March 2020 matters for the policy response, which in the first case concerns duration risk and fund flows [14] and in the second dealer balance-sheet capacity and market-functioning facilities [12, 34].

Third, the walk-forward protocol, the training-quantile threshold and the depth-three surrogate rules make the detector auditable, in line with the requirements for models used in high-stakes financial decisions [26, 64]. Because Shapley attributions describe how a model uses its inputs [58], aggregating them over economically coherent families, as in (5), yields stable and interpretable diagnoses. The framework relies only on public daily index data and extends directly to transaction-level TRACE data, to other currency areas and to reconstruction-based detectors with model-specific attributions [61, 62].

8. Conclusion

This paper presented an explainable machine-learning framework for detecting irregular patterns in the U.S. corporate bond market. Fifteen interpretable features, a walk-forward Isolation Forest and exact TreeSHAP attributions deliver alarms together with their economic drivers. Against 14 documented stress episodes between 2004 and 2024, the detector attains an AUC-PR of 0.546 and a precision of 0.715 at a false-alarm rate of 1.7%, and it detects all 14 episodes at a false-alarm rate of 6.3% when the threshold is relaxed. On injected irregularities it achieves the highest mean AUC (0.935) among seven detectors, including irregularity types that a spread-level rule does not detect. The family-level explanations distinguish rates-driven, volatility-driven, momentum-driven and segment-driven episodes, and depth-three surrogate rules reproduce the alarms with 98.1% accuracy. Future work will extend the framework to bond-level transaction data, to sector- and issuer-level irregularities and to real-time integration with market-functioning indicators.

References

[1] Zou, X., Wei, K., & Lu, Z. (2024). AEGIS: Calibrated anomaly-risk-driven checkpointing and proactive migration for

fault-tolerant high-concurrency stream processing. Innovation and Technology Studies.

[2] Shen, A., & Huang, S. (2022). A systems-aware framework for intelligent large-scale recommendation models. Journal of Computer Science and Electrical Engineering, 4(1), 36–44.

[3] Zhou, F., & Liu, W. (2026). Physics-anchored residual learning for ship fuel consumption and carbon dioxide emission prediction. Energy and AI, 100945.

[4] Hu, X., Guo, L., Wang, J., & Liu, Y. (2025). Computational fluid dynamics and machine learning integration for evaluating solar thermal collector efficiency-based parameter analysis. Scientific Reports, 15, 24528.

[5] Ping, W., & Fan, H. (2024). Disclosure quality and valuation risk in Level 3 investments: A disclosure-implied risk framework for U.S. business development companies. Social Science and Management, 1(1), 78–87.

[6] Chen, J., & Fan, H. (2025). Beyond automation in tax compliance through artificial intelligence and professional judgment. Frontiers in Business and Finance, 2(2), 399–418.

[7] Liu, Y., Ren, S., Wang, X., & Zhou, M. (2024). Temporal logical attention network for log-based anomaly detection in distributed systems. Sensors, 24(24), 7949.

[8] Wang, J., Liu, J., Ge, Y., & Jiang, B. (2023). Risk-aware ensemble learning for credit default prediction under economic uncertainty. World Journal of Economics and Business Research, 1(1), 18–32.

[9] Wang, Y., & Zhang, C. (2023). Aeris: Adaptive executor-scoped resource management for privacy-preserving data processing in Apache Spark. Eurasia Journal of Science and Technology, 5(2), 25–33.

[10] Zhang, F., & Yang, J. S. (2025). Learning driven decision intelligence for autonomous driving through multimodal understanding world modeling and policy optimization. Frontiers in Artificial Intelligence Research, 2(3), 616–634.

[11] Chen, J., Liang, Y., Liu, J., & Zhou, M. (2026). Temporal transformer with conditional tabular GAN for credit card fraud detection: A sequential deep learning approach. Mathematics, 14(7), 1183.

[12] Li, P., Ren, S., Zhang, Q., Wang, X., & Liu, Y. (2024). Think4SCND: Reinforcement learning with thinking model for dynamic supply chain network design. IEEE Access, 12, 195974–195985.

[13] Jiao, Y., & Liang, Y. (2024). An integrated process-mining and explainable machine-learning framework for bottleneck and internal-control risk detection in the multi-entity financial close. Social Science and Management, 1(3), 96–107.

[14] Chen, X. (2023). Algorithmic bias and discrimination in AI-driven lending: A Rashomon-based audit of less discriminatory alternatives and adverse-action accountability. World Journal of Sociology and Law.

[15] Chen, J., Zhang, X., & Han, X. (2023). Explainable identification of tax-related financial reporting anomalies in U.S. public companies: Cross-fitted tax-consistency residuals and Owen-value Shapley propagation. World Journal of Economics and Business Research, 1(1), 33–45.

[16] Xing, S., & Wang, Y. (2025). Proactive data placement in heterogeneous storage systems via predictive multi-objective reinforcement learning. IEEE Access, 13, 117986–117998.

[17] Lu, Z., Zou, X., & Wei, K. (2022). Machine learning-based detection of BGP routing anomalies for enhanced network security and resilience: A hybrid multi-scale temporal learning approach. Journal of Computer Science and Electrical Engineering, 4(1), 62–71.

[18] Xing, S., Wang, Y., & Liu, W. (2025). Multi-dimensional anomaly detection and fault localization in microservice architectures: A dual-channel deep learning approach with causal inference for intelligent sensing. Sensors, 25(11), 3396.

[19] Li, P., Bian, M., & Lin, Y. (2026). Carbon-aware offline reinforcement learning for joint inventory replenishment and transportation scheduling in multi-echelon supply chains. Highlights in Business, Economics and Management, 69, 251–262.

[20] Xing, S., & Wang, Y. (2022). MCG: Marginal–conditional graph learning for joint anomaly detection and root cause localization in distributed cloud systems. Journal of Computer Science and Electrical Engineering, 4(1), 81–88.

[21] Fan, H., Jiao, Y., Wang, M., Chen, J., & Yang, S. (2026). Continual graph learning under strategic adversarial drift in dynamic networks. Scientific Reports.

[22] Yao, W. (2024). PrivTok-XR: Privacy-preserving real-time speech processing for accessible extended reality systems. Innovation and Technology Studies.

[23] Chen, J., Cui, Y., Zhang, X., Yang, J., & Zhou, M. (2024). Temporal convolutional network for carbon tax projection: A data-driven approach. Applied Sciences, 14(20), 9213.

[24] He, X., & Zhou, L. (2024). Adaheal: An adaptive framework for fault detection and automated recovery in large-scale cloud-native microservices. AI and Data Science Journal, 1(1), 132–141.

[25] Wang, B., Wang, Z., Zhao, W., & Liu, Y. (2025). Network fabric simulation and validation for data center routing convergence under large-scale failure scenarios. Computer Science Bulletin, 8(1), 310–326.

[26] Xu, Q., & Zhu, R. (2024). CoSTAR: Explainable and cost-sensitive machine learning for tax-aware corporate fraud risk detection. Innovation and Technology Studies.

[27] Lin, Y., Wang, Y., & Bian, M. (2024). XOR-EW: An explainable and cost-aware artificial intelligence framework for operational risk early warning in financial institutions. AI and Data Science Journal, 1(1), 94–102.

[28] Liang, Y., & Jiao, Y. (2024). Real-time financial fraud detection using a feature-light temporal dynamic graph neural network with contrastive learning. Journal of Trends in Finance and Economics, 1(1), 76–84.

[29] Qin, Y., Rhee, M., Teng, D., Zhang, C., & Zi, B. (2026). Neuro-symbolic constraint verification for LLM-driven internet finance transaction execution. Scientific Reports.

[30] Li, P., Feng, E., Wang, G., & Mo, T. (2024). Hybrid deep reinforcement learning and integer optimization for resource scheduling in heterogeneous AI computing clusters. Innovation and Technology Studies, 1(1), 48–56.

[31] Zhang, F., & Yang, J. (2022). Accuracy-aware model compression for real-time visual perception under resource constraints. Journal of Computer Science and Electrical Engineering, 4(1), 45–52.

[32] Wei, K., Lu, Z., & Zou, X. (2024). A scalable causal machine learning framework for data-driven decision optimization. AI and Data Science Journal, 1(1), 122–131.

[33] Rhee, M., Zou, J., & Qin, Y. (2024). CLARITY: A closed loop framework for duplicate aware failure prediction and budgeted automated remediation in cloud infrastructure. AI and Data Science Journal, 1(1), 76–85.

[34] Zhang, F., & Yang, J. (2024). Robustness-aware model compression for multimodal perception under challenging conditions. AI and Data Science Journal, 1(1), 103–111.

[35] Xu, Q. (2023). Explainable machine learning for tax reporting anomaly detection and compliance risk assessment. World Journal of Economics and Business Research, 1(1), 56–66.

[36] Zou, J., Qin, Y., & Rhee, M. (2024). Beyond model scale: A task-aware reliability benchmark for large language models in clinical decision support. Innovation and Technology Studies, 1(1), 39–47.

[37] Ren, S., Jin, J., Niu, G., & Liu, Y. (2025). ARCS: Adaptive reinforcement learning framework for automated cybersecurity incident response strategy optimization. Applied Sciences, 15(2), 951.

[38] Xing, S., Wang, Y., & Liu, W. (2025). Self-adapting CPU scheduling for mixed database workloads via hierarchical deep reinforcement learning. Symmetry, 17(7), 1109.

[39] Sun, T., Yang, J., Li, J., Chen, J., Liu, M., Fan, L., & Wang, X. (2024). Enhancing auto insurance risk evaluation with transformer and SHAP. IEEE Access, 12, 116546–116557.

[40] Chen, J., Han, X., Zhang, X., & Yang, Y. (2023). Exact concept-level explanations for early detection of financial misstatements. Eurasia Journal of Science and Technology, 5(2), 34–48.

[41] Liu, Y., Guo, L., Hu, X., & Zhou, M. (2025). Sensor-integrated inverse design of sustainable food packaging materials via generative adversarial networks. Sensors, 25(11), 3320.

[42] Qin, Y., Zou, J., & Rhee, M. (2022). An AI-driven intelligent transaction execution framework for internet finance platforms: Cost-sensitive real-time integrity interdiction. Eurasia Journal of Science and Technology, 4(2), 36–44.

[43] Wang, J., Tan, Y., Jiang, B., Wu, B., & Liu, W. (2025). Dynamic marketing uplift modeling: A symmetry-preserving framework integrating causal forests with deep reinforcement learning for personalized intervention strategies. Symmetry, 17(4), 610.

[44] Zhou, F. (2026). Moment-constrained side-force allocation for wind-assisted merchant ships: Hydrodynamic performance of retractable anti-leeway fins and flap rudders. Innovation and Technology Studies, 3(1), 27–41.

[45] Ping, W., & Fan, H. (2023). Beyond isolated ratios: Accrual–revenue–cash flow indicators and interaction-aware learning for detecting financial reporting irregularities in U.S. public companies. World Journal of Economics and Business Research, 1(1), 46–55.

[46] Zhou, L., & He, X. (2022). Dependency-aware root cause analysis and automated fault recovery for cloud-native microservices. Journal of Computer Science and Electrical Engineering, 4(1), 72–80.

[47] Lin, Y., Yue, X., Bian, M., Zhu, R., & Teng, D. (2026). Subgraph-level temporal graph transformer with laundering-motif prototypes for anti-money laundering on blockchains. IEEE Access.

[48] Fan, H., & Chen, J. (2025). From ledgers to large language models in contemporary accounting practice. American Journal of Big Data, 6(5).

[49] Zhu, R., & Yue, X. (2024). An explainable machine learning framework for predicting dividend sustainability and financial risk of US REITs under interest-rate shocks. Journal of Trends in Finance and Economics, 1(2), 48–57.

[50] Zhang, F., & Wu, B. (2025). Large language models as general purpose intelligence systems for reasoning, planning and decision making. American Journal of Artificial Intelligence and Neural Networks, 6(4), 45–72.

[51] Wang, G., Li, P., Feng, E., & Mo, T. (2024). Uncertainty-aware LLM-as-a-judge evaluation for reliable search and recommendation systems: Calibrated confidence and impact-aware selective human review. AI and Data Science Journal, 1(1), 112–121.

[52] Chen, X. (2026). Can people challenge an AI decision? Measuring the contestability chain in law and corporate disclosure. AI and Ethics, 6(6), 642.

[53] Chen, X. (2024). Who is accountable for AI decisions? Rethinking legal liability and ethical responsibility through counterfactual harm apportionment. Innovation and Technology Studies.

[54] Wang, J., Liu, J., Zheng, W., & Ge, Y. (2025). Temporal heterogeneous graph contrastive learning for fraud detection in credit card transactions. IEEE Access.

[55] Wang, J., Ge, Y., Liu, J., & Jiang, B. (2024). Early warning of liquidity stress in U.S. banks with seasonally aligned frailty hazards and delayed-feedback conformal alarms. Journal of Trends in Finance and Economics, 1(1), 85–100.

[56] Bian, M., Wang, Y., & Lin, Y. (2024). Explainable supplier risk prediction for critical telecommunications infrastructure procurement: An empirical XGBoost–SHAP framework. Social Science and Management, 1(1), 64–77.

[57] Wang, Y., Bian, M., & Lin, Y. (2022). An explainable temporal graph neural network for disruption prediction and risk propagation in multi-tier supply chains. Journal of Computer Science and Electrical Engineering, 4(1), 17–25.

[58] Mo, T., Zhang, C., Zou, J., Guo, Z., & Rhee, M. (2026). Self-evolving AI agents with dual memory for automated software testing and bug localization. IEEE Access.

[59] Mo, T., Wang, G., Feng, E., & Li, P. (2023). Execution-feedback-guided context selection for large language model-based software testing and debugging. World Journal of Information and Knowledge Management, 1(1), 55–64.

[60] Wang, G., Li, P., Feng, E., Shen, A., & Rhee, M. (2026). A dual-process memory-augmented large language model agent for explainable sequential recommendation. IEEE Access.

[61] Ji, E., Wang, Y., Xing, S., & Jin, J. (2025). Hierarchical reinforcement learning for energy-efficient API traffic optimization in large-scale advertising systems. IEEE Access.

[62] Yao, W. (2022). Raptor: Risk-adaptive security and privacy-preserving transaction processing for large-scale mobile payment systems. Journal of Computer Science and Electrical Engineering, 4(1), 89–97.

[63] Wang, Y., & Xing, S. (2025). AI-driven CPU resource management in cloud operating systems. Journal of Computer and Communications, 13(6), 135–149.

[64] Zhou, F. (2026). Spectral co-design of ship rudder geometry and heading autopilot for energy efficiency and manoeuvring safety. Transactions on Environment, Energy and Earth Sciences, 7, 158–181.