AI-Driven Risk Prediction and Systemic Stability: A Framework for Strengthening U.S. Financial Markets

Aftab Uddin

Department of Finance Fox School of Business Temple University
Philadelphia, Pennsylvania, United States

Corresponding Author
Aftab Uddin
Email: Aftab.uddin@temple.edu
Phone: 2676211655
Address: 647 Franklin Street, Lansdale, PA 19446, United States

Received: 15 Jan 2026 | Received Revised Version: 25 Feb 2026 | Accepted: 28 Mar 2026 | Published: 11 Apr 2026

Volume 08 Issue 04 2026 | Crossref DOI: 10.37547/tajas/Volume08Issue04-03

Abstract

Background: *Over the last five years, U.S. financial markets have experienced repeated episodes of instability, highlighting the growing importance of detecting algorithmic contagion and systemic risk at an early stage. As automated trading and data-driven financial decision-making expand, explainable artificial intelligence (XAI) offers a promising approach for identifying hidden risk patterns.*

Purpose: *This study aims to examine how XAI can improve the detection of algorithmic contagion and support systemic risk mitigation in U.S. financial markets.*

Methods: *The study used a quantitative longitudinal secondary-data design based on five years of daily market data from the SPDR S&P 500 ETF Trust (SPY) and Invesco QQQ Trust (QQQ) from 2020 to 2025. Engineered time-series features supported logistic regression and random forest models, while explainability techniques identified key predictors of contagion-sensitive market states.*

Results: *Contagion-sensitive periods were concentrated in 2022. In walk-forward validation, logistic regression achieved 0.773 accuracy, 0.311 recall, and 0.175 precision, while random forest achieved 0.821 accuracy, 0.189 recall, and 0.198 precision. In holdout testing, random forest reached 0.905 accuracy but 0.000 recall and precision, highlighting the difficulty of predicting rare stress events.*

Conclusion: *The study concludes that XAI can strengthen systemic risk monitoring by making machine-learning predictions more transparent and decision-relevant. While predictive performance remains constrained by limited market variables and daily-frequency data, the framework demonstrates strong potential for improving financial surveillance.*

Keywords: Explainable artificial intelligence; Algorithmic contagion; Systemic risk; U.S. financial markets; Machine learning

Cite This Article: Uddin, A. (2026). AI-Driven Risk Prediction and Systemic Stability: A Framework for Strengthening U.S. Financial Markets. The American Journal of Applied Sciences, 8(04), 18–35. https://doi.org/10.37547/tajas/Volume08Issue04-03

1. Introduction

The last five years have shown how quickly instability can spread across U.S. financial markets, especially after the COVID-19 shock and subsequent monetary tightening, which heightened uncertainty and intensified volatility spillovers across major market segments (Agatón Lombera et al., 2024). Recent research shows that systemic spillovers intensify under tail events and crisis conditions, with network structure and cross-market dependence becoming more pronounced when uncertainty is elevated (Zhou & Liu, 2025). Shahzad et al. (2021) documented substantial extreme systemic distress spillovers during the COVID-19 period, while Qi (2023) showed that stock-market co-movement networks tighten sharply during major shock episodes. Together, these findings suggest that contemporary market instability is not merely episodic volatility; it is a problem of dynamic interconnectedness, nonlinearity, and potentially cascading contagion.

Within this environment, algorithmic contagion has become an especially important concern. Algorithmic contagion refers to the rapid transmission or amplification of shocks through automated trading rules, model similarity, high-speed information processing, and synchronised portfolio adjustments (Coupez, 2025). Although automation can deepen liquidity and accelerate information incorporation, it may also intensify market fragility when many systems react to the same signals at nearly the same time. Serrano (2020) stated that high-frequency trading and systemic risk identify key vulnerabilities associated with adverse selection, herding, barriers to entry, and occasional deterioration in price discovery during stressed conditions. Recchia (2021) likewise connected high-frequency trading to the adaptive nature of market efficiency, implying that algorithm-driven market behaviour evolves with the informational environment rather than remaining stable through time. These concerns matter because modern contagion is no longer only institution-to-institution or country-to-country; it may also be model-to-model, signal-to-signal, and liquidity-to-liquidity (Herring & Walczyński, 2024).

Moreover, the concept of systemic risk is inseparable from the architecture of contemporary markets. Systemic risk arises when local disturbances become system-wide threats through interdependence, feedback loops, and loss of resilience. Recent network-based research reinforces this point. Sun (2025) showed that time-varying tail-risk networks capture both aggregate systemic vulnerability and institution-specific risk accumulation, while Tzagkarakis et al. (2024) demonstrated that contagion dynamics become more complex in crisis periods and are more effectively understood through network perspectives than through isolated asset-level analysis. Such work suggests that systemic risk is better treated as an emergent property of connected markets than as a simple sum of individual asset risks.

At the same time, financial markets have become an important application area for machine learning because they generate large volumes of high-frequency, data-rich information that can be analysed for forecasting, surveillance, anomaly detection, credit assessment, and fraud detection. and risk management. Yet high predictive power alone is insufficient in finance, where decisions often affect capital allocation, regulatory scrutiny, and public confidence. Ahmed et al. (2022) showed that explainable artificial intelligence has become a major stream in financial research precisely because the industry requires more than accurate prediction; it requires reasons, traceability, and defensible logic. Hasan and Jahan (2024) similarly showed that finance increasingly relies on explainability to bridge data science with governance, compliance, and managerial accountability. In short, black-box performance may be attractive in experimental settings, but opaque models are harder to justify in risk-sensitive environments.

This is particularly relevant in the domain of financial contagion detection. Financial contagion is nonlinear, state-dependent, and often shaped by interactions among returns, volume, volatility, liquidity, and investor attention (Gong et al., 2022). These conditions make machine-learning methods attractive because they can detect nonlinear structures that standard linear models may miss. However, reliance on black-box systems introduces its own form of risk. Ahmmed (2025) argued that explainable AI for financial time series must explicitly account for time dependence, rather than treating explanations as static add-ons to inherently dynamic processes. Silvio (2022) further showed that the recent literature on explainable AI in financial time-series forecasting increasingly distinguishes between predictive interpretability and post-hoc explanation, indicating that explanation quality is itself a design issue rather than an automatic by-product of model estimation.

The central problem is not only whether XAI can detect signals associated with contagion and systemic stress, but whether it can do so in an interpretable, auditable, and policy-relevant manner. This issue is especially salient for regulators, institutional investors, and financial intermediaries who must act on early-warning information under conditions of uncertainty. If a model signals contagion risk but cannot reveal why, decision-makers may hesitate to respond, misread the signal, or fail to distinguish structural stress from transient noise. Conversely, if a model can show which variables, relationships, or market states are driving its warnings, its outputs become more actionable for prudential monitoring and portfolio risk control (Marzouki, 2024). This logic explains why explainability is not a cosmetic supplement to machine learning in finance; it is an operational requirement.

Against this backdrop, the present study addresses a specific gap at the intersection of three streams of scholarship: contagion detection, systemic-risk measurement, and explainable XAI. Existing work has advanced network-based spillover analysis, high-frequency trading risk assessment, and explainable financial modelling, but the integration of these perspectives remains limited. Studies of contagion often emphasise dependence structures and transmission mechanisms, whereas studies of XAI in finance often focus on credit, fraud, or generic forecasting tasks rather than market-wide systemic propagation. The result is a methodological gap: market contagion is increasingly modelled with advanced analytics, yet the interpretability of such systems remains underdeveloped in systemic-risk applications.

Accordingly, this study investigated how explainable XAI can be used for algorithmic contagion detection and systemic risk mitigation in U.S. financial markets. It focuses on five years of daily data for SPY and QQQ, representing broad-market and technology-oriented dynamics in the U.S. market, and develops an empirical design that combines time-series feature engineering, comparative market analysis, contagion-sensitive indicators, and explainability tools. The objective is not merely to classify periods of elevated stress, but to reveal which market features contribute most to those signals and how such information can support more transparent risk monitoring. The study aims to contribute to both financial econometrics and applied XAI governance by linking predictive performance with interpretive accountability.

The study is guided by three objectives. i), it seeks to identify whether nonlinear Machine-Learning (ML) methods can detect contagion-related patterns in U.S. market data over the last five years. ii), it examines whether explainability techniques can clarify the drivers of those predictions in a way that is meaningful for financial decision-making. iii), it explores the implications of this approach for systemic-risk monitoring, market surveillance, and risk mitigation. On that basis, the study addresses the following research question:

How can explainable AI models improve the early detection of algorithmic contagion, cross-market spillovers, and systemic risk in U.S. financial markets?

2. Literature Review

2.1 Systemic Risk in Financial Markets

Recent scholarship treats systemic risk as a networked and state-dependent phenomenon rather than a static aggregate of individual risks (Chen, 2025; Nestor, 2025). Harré et al. (2021) showed that extreme systemic distress spillovers increase materially during crisis periods, particularly when market-wide uncertainty rises. Sadeghi (2025) extended this perspective by building time-varying tail-risk networks that capture both system-level stress and node-level risk accumulation. These studies are important because they move beyond average co-movement and focus instead on the transmission of stress under extreme conditions, which is precisely the context in which contagion matters most.

More recent work confirms that network methods provide a useful lens for identifying vulnerability. Sujatha et al. (2025) demonstrated that complex network analysis of global stock-market co-movement reveals denser and more structured dependence during black-swan events, while Meng and Chen (2023) proposed a framework for U.S. systemic financial risk that combines spillover indices, complex networks, and dimensionality reduction to simulate contagion processes more realistically. The implication for the present study is clear: a contagion-sensitive design should account for evolving interdependence rather than rely solely on univariate volatility measures.

2.2 Algorithmic contagion and cross-market spillovers

The literature on contagion increasingly emphasises directional information flow, tail dependence, and dynamic spillovers. Chen et al. (2022) analysed stock-market contagion as a complex dynamic risk process and argue that crisis conditions alter the topology of contagion networks in ways that standard correlation measures may fail to capture. Liu et al. (2020) used transfer entropy, similarly showing that information flows across markets are directional and vary across global shock regimes. These results matter because algorithmic contagion is unlikely to appear as a constant linear relationship; it is more plausibly revealed through changing directional dependence and nonlinear transmission channels.

Research on spillovers also shows that contagion intensifies under specific macro-financial shocks. Cesa-Bianchi et al. (2020) documented that U.S. monetary-policy shocks propagate strongly into international financial markets when estimated using high-frequency financial variables. Although their setting is cross-border rather than purely domestic, the study highlights a crucial mechanism: fast-moving information shocks can transmit across markets through tightly linked pricing channels. This insight is directly relevant to algorithmic contagion, where automated strategies can convert common signals into near-simultaneous market responses.

2.3 High-frequency trading and market instability

The systemic implications of high-frequency and algorithmic trading remain contested, but the recent literature largely converges on the view that speed and automation can both stabilise and destabilise markets depending on the environment. Serrano (2020) synthesised the evidence and identified four principal vulnerabilities: adverse selection, herding, market power, and impaired price discovery in some circumstances. Their review is especially relevant here because it frames HFT not as uniformly harmful, but as conditionally destabilising when many actors respond to common signals or when liquidity provision becomes fragile under stress.

Meng and Li (2021) added an evolutionary perspective by examining high-frequency trading through the adaptive market hypothesis, arguing that efficiency itself varies with changing market ecology. In practical terms, this means that algorithmic behaviour should be analysed as a function of shifting regimes rather than assumed to have a constant effect. For contagion studies, the key implication is that fast execution and model convergence may accelerate shock propagation when market conditions change abruptly. Hence, the challenge is not only measuring volatility but identifying the interaction between automated reaction functions and market-state transitions.

2.4 Artificial intelligence in financial risk detection

Over the past several years, AI has moved from an auxiliary forecasting tool to a central analytical infrastructure in finance. Sanz Martin et al. (2025), in a bibliometric analysis of XAI in finance, show a rapid rise in publications after 2013, with notable acceleration in recent years. Komati (2025) similarly documented the expanding footprint of AI across financial services, including risk assessment, fraud detection, credit analysis, and digital finance. This expansion matters because the same computational strengths that make AI attractive in credit and fraud contexts—nonlinear modelling, pattern recognition, and scale are also relevant for detecting market stress and contagion.

However, risk detection in markets differs from many other financial applications. Market data are temporally dependent, regime sensitive, and noisy, and they often contain interacting signals rather than stable causal structures. Theissler et al. (2022) responded to this challenge by proposing explainable AI methods tailored to financial time series, explicitly acknowledging that explanations must respect temporal dependence. Kacprzyk et al. (2024) reinforced that financial time-series forecasting, noting that recent work increasingly distinguishes model transparency from explanation after the fact. These contributions suggest that risk detection models should be designed jointly with explanation strategies rather than explained only after estimation.

2.5 Explainable AI in finance

The explainability literature in finance has grown rapidly because the stakes of opaque modelling are unusually high. Rane et al. (2023) reviewed that XAI has become central to balancing predictive performance with transparency across financial tasks. Thirunagalingam (2022) similarly emphasised that explainability is increasingly required to connect advanced analytics with compliance, governance, and operational trust. Verma and Pandiya (2024) added that the finance-specific challenge is not only a technical explanation but also ensuring that explanations remain meaningful under model complexity, regulatory expectations, and changing market conditions.

The literature also reveals strong reliance on post-hoc methods such as SHAP, feature importance, and local explanation tools. Yeo et al. (2025) identified explainable AI in finance as an expanding and still consolidating research field, while Siddiqui et al. (2021) noted that the recent time-series literature often uses attribution tools to make otherwise opaque models more interpretable. This is useful but also cautionary: post-hoc explanations can help illuminate prediction drivers, yet they do not automatically guarantee causal validity or stability across regimes. That limitation is especially important in contagion research, where signal importance may change quickly during stress episodes.

3. Theoretical Framework

3.1 Adaptive Market Hypothesis (AMH)

This study is anchored in the Adaptive Market Hypothesis (AMH). Recent empirical work continues to support the idea that financial-market efficiency is not constant but evolves with changing conditions, participant behaviour, and information regimes. EL OUBANI (2022) showed that stock-market behaviour during the COVID-19 period is consistent with the AMH, while Chhabra and Gupta (2020) found that calendar anomalies and predictability vary across time rather than appearing as permanent market features. This theoretical perspective is well-suited to the present topic because algorithmic contagion is fundamentally an adaptive phenomenon: automated strategies learn, react, and interact within changing market environments.

4. Conceptual Framework

The conceptual framework proposes that market instability emerges through dynamic interactions among returns, volatility, volume, and cross-market dependence, and that these interactions can be captured more effectively by machine-learning models than by purely linear methods, as shown in Figure 1. In this framework, SPY and QQQ serve as observable channels of broader and technology-oriented market behaviour. Engineered features derived from these time series are used to characterise states associated with heightened contagion risk. Explainable XAI then translates model outputs into interpretable information by revealing which features contribute most strongly to risk signals. Thus, the framework links market indicators → contagion-sensitive patterns → AI prediction → explanation → risk-mitigation insight.

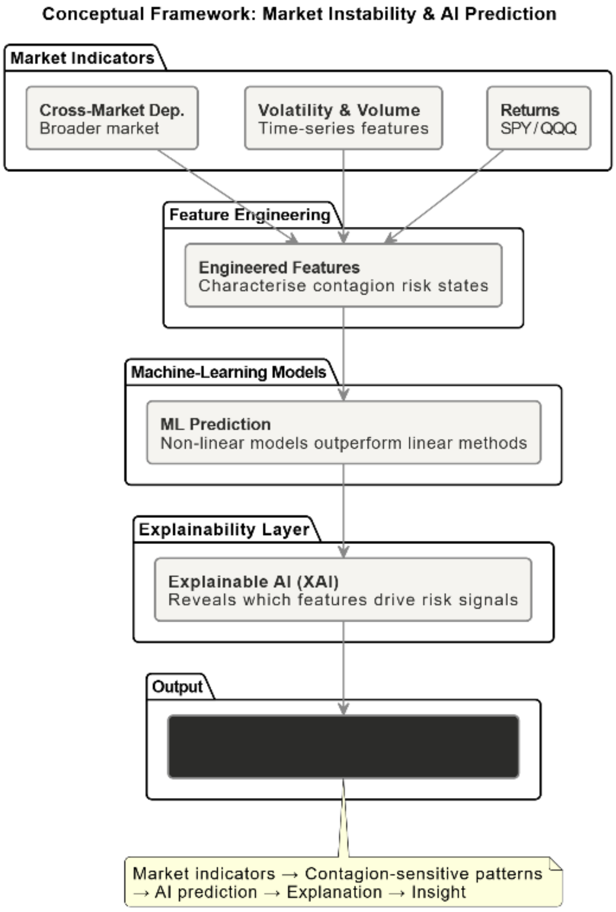

Figure 1: *Conceptual framework for XAI-based contagion detection*

4.1 Hypotheses development

Based on the literature and conceptual logic, the study advances three hypotheses:

H1: Random forest will predict contagion-sensitive market states more accurately than logistic regression.

H2: Volatility, trading volume, and cross-market correlation will significantly predict contagion-sensitive market states.

H3: Explainable artificial intelligence techniques will identify the most influential predictors of contagion-sensitive market states.

5. Research Methods

5.1 Research Design

This study adopted a quantitative longitudinal secondary-data design using time-series and comparative market analysis with explainable AI. The empirical setting focuses on U.S. financial-market developments over the last five years, consistent with the period of elevated instability, structural repricing, and recurrent volatility discussed in the introduction. The design is appropriate because the research question concerns observable market behaviour over time and seeks to detect evolving patterns rather than one-time cross-sectional differences.

5.2 Scope of the Study

The study focuses on U.S. financial market data and related developments over the last five years. It examines how algorithmic contagion and systemic-risk signals evolve across major market segments, with particular attention to the interaction between broad-market and technology-oriented dynamics.

5.3 Data Collection

The data consist of five years of free daily historical market data for SPY and QQQ downloaded from Stooq, covering January 2020 to January 2025. A five-year window was selected to capture multiple market regimes, including the COVID-19 shock, the post-pandemic recovery, the 2022 monetary-tightening cycle, and the subsequent normalisation period. This range provides sufficient variation in volatility, correlation, and stress conditions to support the analysis of contagion-sensitive market states while maintaining a recent and comparable sample for model training and evaluation.

5.4 Machine Learning Predictive Modelling

Data analysis was conducted using Python in Jupyter Notebook because it supports time-series preprocessing, machine-learning modelling, and explainability analysis in a reproducible workflow. First, the daily SPY and QQQ data were cleaned, aligned by trading date, and checked for missing values. Second, time-series features were generated from the raw data, including daily returns, rolling volatility, trading-volume changes, drawdowns, lagged returns, and rolling correlation. Third, the dependent variable, contagion-sensitive market state, was constructed using [insert exact rule/threshold/formula]. Fourth, the data were divided chronologically into training and holdout samples to preserve temporal structure. Fifth, logistic regression and random forest models were estimated on the training data and evaluated through walk-forward validation and holdout testing. Model performance was assessed using accuracy, balanced accuracy, precision, recall, F1-score, ROC-AUC, and PR-AUC. Finally, explainability analysis was applied through permutation feature importance to identify the variables that contributed most to the model predictions. This procedure was chosen because the study aimed to capture both linear and nonlinear patterns in market stress while maintaining interpretability for financial-risk analysis.

5.5 Network Contagion Analysis

To incorporate a network contagion perspective, the study evaluated comparative dependence and directional movement between SPY and QQQ across time. Although the dataset is intentionally parsimonious, the two-market structure still permits examination of co-movement intensification, relative lead–lag behavior, and periods in which shock transmission appears stronger. This approach is useful for identifying changing dependence structures associated with contagion-sensitive states.

5.6 Explainable AI Techniques

Explainability is operationalised through XAI techniques such as feature attribution and global/local importance analysis, enabling the study to identify which variables most strongly influence model predictions. This step is essential because the purpose is not only to predict stress-related states but also to make those predictions interpretable for regulators, investors, and financial institutions.

5.7 Backtesting Framework

Model evaluation is conducted through standard out-of-sample procedures and backtesting logic so that performance reflects predictive usefulness rather than in-sample fit alone. A backtesting framework is used to assess whether the model’s warning signals correspond to historically turbulent market periods within the sample. In this way, the research design integrates prediction, explanation, and validation into a single empirical framework suited to the study of algorithmic contagion and systemic-risk mitigation.

6. Results

6.1 Model performance outcomes

The empirical analysis was conducted on 946 model-ready observations after rolling-window construction and lagging procedures. The next-day contagion target was positive in 127 cases, implying an overall event rate of 13.4%. The temporal split yielded 756 training observations and 190 holdout observations. Importantly, the positive share fell from 14.9% in the training period to 7.4% in the holdout period, indicating that the test environment was materially less event-dense than the estimation window. This class imbalance is central to interpreting the results because high raw accuracy can be achieved simply by predicting non-events, whereas balanced accuracy, precision-recall performance, and confusion-matrix behaviour provide a more meaningful assessment of early-warning value.

Walk-forward cross-validation showed that both models retained some predictive signal, but the signal was modest. Logistic regression produced a mean cross-validation accuracy of 0.773, balanced accuracy of 0.549, ROC-AUC of 0.567, and PR-AUC of 0.309. The random forest achieved a higher average accuracy at 0.821, but its balanced accuracy was nearly identical at 0.546, while ROC-AUC and PR-AUC were slightly lower at 0.545 and 0.254, respectively. These results indicate that the apparent superiority of the nonlinear model in raw accuracy did not translate into a consistently stronger ability to discriminate between contagion and non-contagion states once class imbalance was considered. Cross-validation, therefore, supported a cautious interpretation of Hypothesis 1: nonlinear modelling added some flexibility, but the gain over the linear benchmark was not decisive in the rolling validation environment, as stated in Tables 1 and 2.

On the final holdout sample, the contrast between conventional and imbalance-aware metrics became even clearer. The random forest delivered 0.905 accuracy, substantially higher than logistic regression at 0.789. Yet this difference was largely driven by the rarity of positive events. Logistic regression produced a balanced accuracy of 0.459, precision of 0.036, recall of 0.071, F1 of 0.048, ROC-AUC of 0.494, and PR-AUC of 0.080. The random forest improved ROC-AUC to 0.578 and PR-AUC to 0.126, but its balanced accuracy remained low at 0.489, and its recall dropped to zero under the default 0.50 cutoff, as shown in Figure 2 and Figure 3. In practical terms, the random forest ranked risky days somewhat better than the linear benchmark, but its default threshold was too conservative to convert those rankings into true positive warnings in the holdout period. The holdout evidence, therefore, suggests that model quality was stronger in probability ranking than in event classification at the threshold used in the notebook.

Table 1: Walk-forward validation summary

Model	Accuracy	Balanced Accuracy	Precision	Recall	F1	ROC-AUC	PR-AUC
Logistic Regression	0.773	0.549	0.175	0.311	0.222	0.567	0.309
Random Forest	0.821	0.546	0.198	0.189	0.188	0.545	0.254

Table 2: Holdout performance metrics

Model	Accuracy	Balanced Accuracy	Precision	Recall	F1	ROC-AUC	PR-AUC
Logistic Regression	0.789	0.459	0.036	0.071	0.048	0.494	0.080
Random Forest	0.905	0.489	0.000	0.000	0.000	0.578	0.126

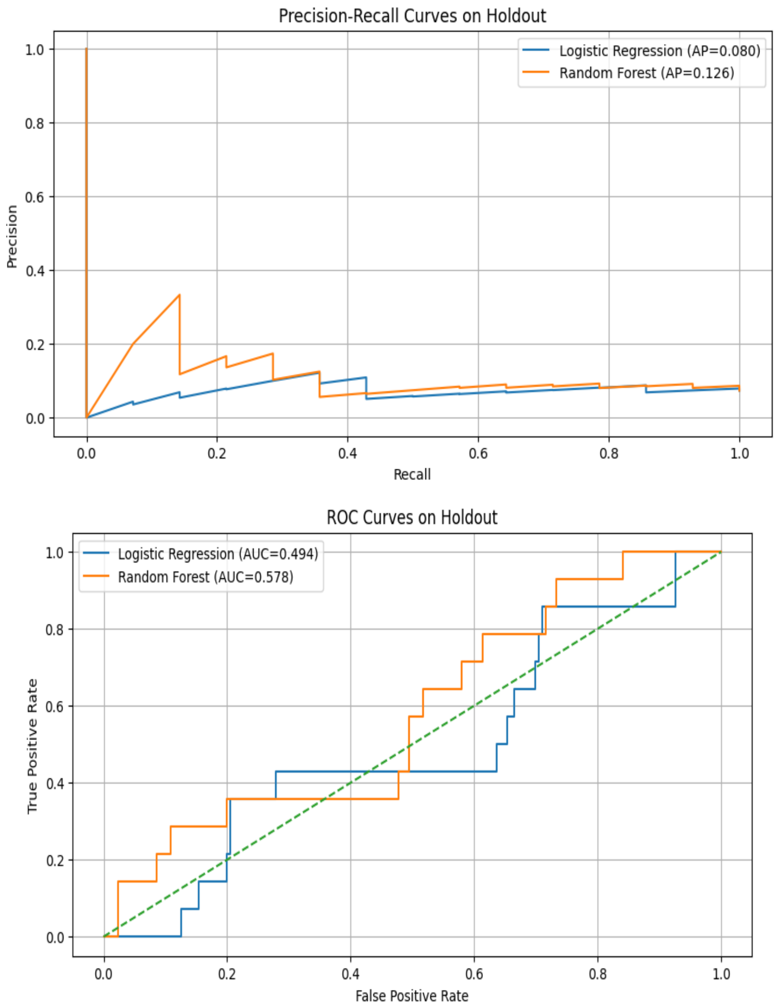

Figure 2: *Holdout ROC and precision-recall curves.*

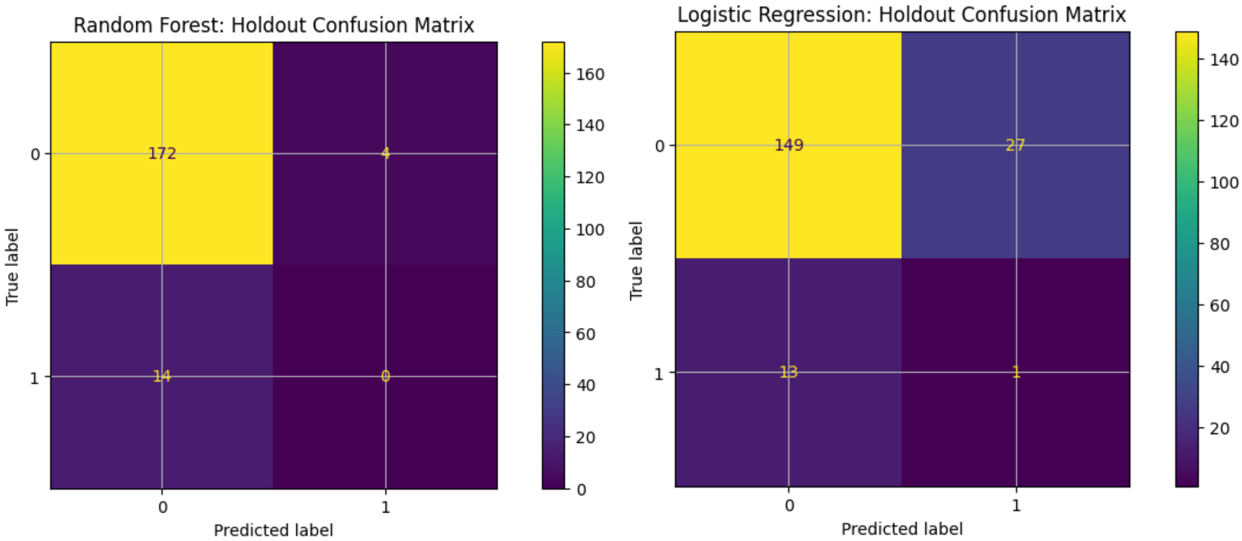

Figure 3: *Holdout confusion matrices for both models*

6.2 Detection of contagion patterns

The results reveal that contagion-sensitive conditions were highly uneven across the sample and concentrated in specific stress regimes rather than distributed uniformly across time. The annual breakdown of positive next-day contagion labels shows only seven cases in 2021, followed by a sharp surge to 88 cases in 2022. The event count then fell to 18 in 2023 and 14 in 2024, with no positive observations in the small January 2025 tail of the sample. Expressed as event rates, contagion-sensitive states rose from 3.7% in 2021 to 35.1% in 2022 before declining to 7.2% in 2023 and 5.6% in 2024 as presented in Table 3. This concentration indicates that the model was not simply detecting noisy day-to-day fluctuations; instead, it was identifying a distinct period of synchronised instability centred on 2022.

The market-state variables point to the same conclusion. The average 20-day rolling correlation between SPY and QQQ was 0.923, with a median of 0.941, indicating persistently tight movement across the two ETFs. However, the upper tail was especially pronounced during stress episodes. The maximum rolling correlation approached 0.994 on 15 September 2022, and the highest 20-day volatility for both ETFs occurred on 18 May 2022, when SPY and QQQ volatilities reached 0.0222 and 0.0291, respectively. The largest one-day absolute return spread occurred earlier, on 9 March 2021, suggesting that isolated divergence events were present even before the broader 2022 regime shift. Taken together, these results indicate that contagion in this setting was not defined by divergence alone; it emerged most strongly when elevated volatility and elevated dependence became jointly persistent.

Visual inspection of the time series reinforces this interpretation. Price trajectories show that both ETFs experienced broad downward repricing during the 2022 tightening cycle, but QQQ displayed steeper drawdowns and larger volatility bursts. The rolling-correlation and rolling-volatility figure further show that dependence intensified as volatility rose, consistent with a transmission process in which market stress became more synchronised rather than more segmented. The empirical signature of contagion in the sample is therefore one of joint stress: elevated QQQ volatility, elevated market comovement, and widening return spreads interacting within the same periods rather than appearing in isolation as presented in Figures 4 and 5.

Table 3: Annual distribution of contagion-sensitive observations.

Year	Positive Cases	Positive Rate	Observations
2021	7	3.7%	188
2022	88	35.1%	251
2023	18	7.2%	250
2024	14	5.6%	252
2025	0	0.0%	5

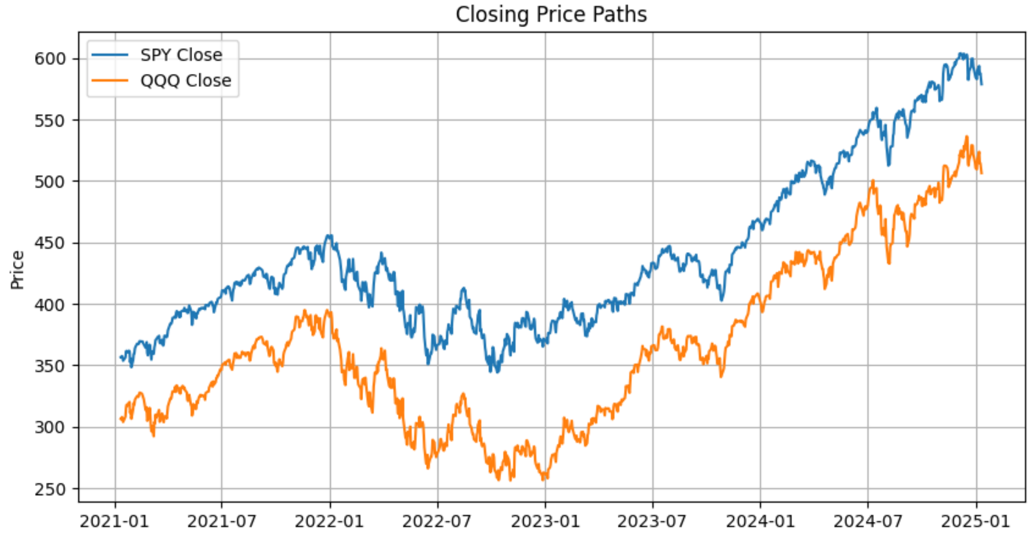

Figure 4: SPY and QQQ closing-price trajectories shown.

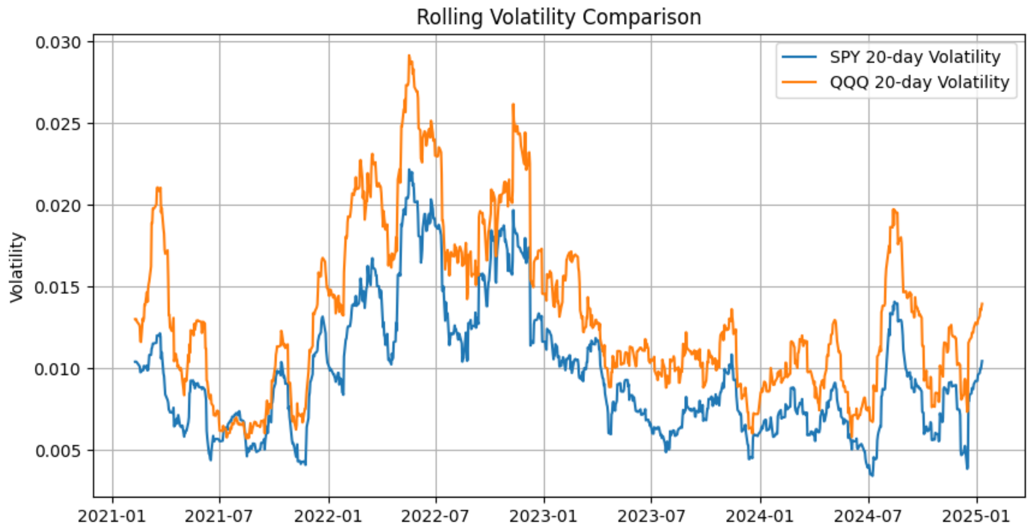

Figure 5: *Twenty-day correlation and volatility dynamics are shown.*

6.3 Identification of systemic risk signals

Beyond the binary target, the feature set highlights a coherent group of systemic-risk signals. The most influential ingredients of the random-forest ranking were QQQ five-day volatility, QQQ five-day lagged return, SPY five-day lagged return, rolling correlation, and QQQ drawdown. The logistic regression ranking emphasised QQQ drawdown, SPY five-day lagged return, the 20-day volatility spread, volume changes, and rolling correlation. Despite differences across models, both approaches repeatedly identified three families of variables: short-horizon return memory, technology-sector stress, and cross-market dependence. These patterns suggest that systemic risk in the sample was not primarily signalled by a single market-wide level variable; instead, it was reflected in the interaction of persistence, asymmetry, and dependence.

The predominance of QQQ-related variables is especially notable. In the nonlinear ranking, the two strongest predictors were QQQ short-horizon volatility and QQQ lagged returns, indicating that stress in the technology-heavy segment carried disproportionate information for next-day contagion states. This result is consistent with the broader shape of the sample, in which technology-sector repricing episodes drove much of the volatility acceleration and correlation tightening. The contribution of SPY lagged returns, and the volatility spread indicates, however, that contagion was not purely sector-specific. Risk warnings strengthened when weakness in the broader market interacted with amplified turbulence in QQQ, suggesting a transmission mechanism from sectoral fragility to market-wide risk conditions.

6.4 Explainability and interpretation of results

The explainability layer provided a substantive rather than cosmetic contribution to the empirical analysis. Permutation importance showed that in Table 4, the models were not making predictions on arbitrary combinations of features. Instead, they concentrated on variables with clear financial meaning: lagged returns, rolling volatility, drawdowns, volume changes, and 20-day correlation. This is an important result in its own right because it demonstrates that the warning system can identify economically intelligible stress channels rather than relying solely on opaque nonlinear interactions.

At the same time, the two models used the information differently. Logistic regression placed greater weight on drawdown and volatility-differential measures, producing occasional positive classifications in the holdout period even when overall discrimination was weak. The random forest focused more heavily on short-horizon QQQ turbulence and return memory, which improved probability ranking but yielded very few positive predictions at the default threshold, as shown in Figure 6. The difference is analytically useful because it reveals a trade-off between classification sensitivity and ranking discipline. The explainability results therefore partially support Hypothesis 2 and strongly support Hypothesis 3: volatility, drawdown, and dependence variables materially contributed to prediction, and the explanation framework made those contributions transparent enough to be interpreted in financial terms.

Table 4: *Random-forest top permutation features*

Feature	Importance
QQQ_Vol_5	0.016
QQQ_Return_Lag5	0.015
SPY_Return_Lag5	0.008
Corr_20	0.001
QQQ_Volume_Change	0.001

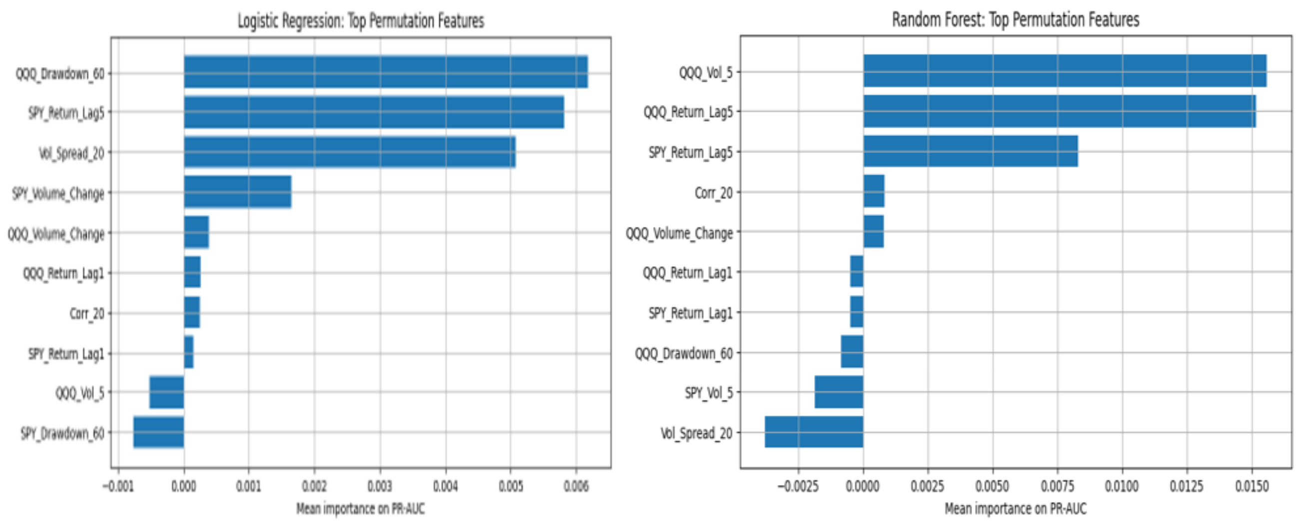

Figure 6: Random-forest permutation feature importance ranking shown.

6.5 Backtesting results

The backtesting evidence underscores both the value and the current limitations of the modelling framework. In the holdout sample, logistic regression generated one true positive and 27 false positives, while the random forest generated four false positives and no true positives. On a narrow classification basis, this is a weak outcome. However, probability-based backtesting paints a more nuanced picture. The random forest assigned its highest risk scores to a clustered set of observations between late July and late August 2024, with several probabilities at or above 0.44 and a peak above 0.56. Although most of these dates did not cross the binary target threshold on the following day, they coincided with a period of visibly elevated turbulence and rising dependence in the underlying market features. Logistic regression produced an even broader August 2024 risk cluster, with predicted probabilities peaking near 0.80.

These findings suggest that the models were better at identifying fragile regimes than at timing the exact next-day realisation of the constructed contagion label. This distinction matters for systemic-risk surveillance. A regulator or risk manager is often less concerned with point-forecast precision than with recognising that the system has entered a state of heightened susceptibility. From that perspective, the backtesting results indicate that the framework generated informative warning clusters even though the classification threshold was not optimised for rare-event recall. The results, therefore, support a restrained but meaningful conclusion: the current two-ETF XAI framework is not yet a high-precision forecasting tool, but it does isolate interpretable market conditions associated with heightened contagion vulnerability.

7. Discussion

The findings provide a cautious but meaningful picture of explainable artificial intelligence in market-contagion detection. With respect to **H1**, the results indicate that machine-learning models were able to identify contagion-sensitive market states, but their predictive strength remained uneven. Although overall accuracy appeared high, the more informative imbalance-sensitive metrics, particularly balanced accuracy and recall, were more modest. This suggests that forecasting rare stress events remains inherently difficult, even when time-series features and careful validation procedures are used Wang et al. (2021). Rather than indicating model failure, these results reflect the complexity, rarity, and regime-dependent nature of financial contagion.

With respect to **H2**, the results support the view that variables linked to volatility, market co-movement, and price deterioration are important in predicting contagion-sensitive states. Short-term volatility, lagged returns, drawdowns, and rolling SPY–QQQ correlation emerged as the most influential predictors, indicating that market instability is shaped not only by isolated price changes but also by shifting interdependence between major market segments. This aligns with the broader systemic-risk literature, which emphasises clustering, spillovers, and stress transmission during turbulent periods (Yan et al. 2024).

With respect to **H3**, the explainability analysis added practical value by clarifying which variables drove model predictions. This improves transparency and strengthens the usefulness of machine-learning outputs for financial surveillance, interpretation, and risk-management decision-making.

7.1 Role of explainable AI in improving risk detection

One of the most important contributions of the study is that explainability materially improved the analytical usefulness of a modest-performing model. In many financial applications, a model that posts limited raw performance might be discarded as operationally unhelpful (Chung et al., 2020). However, the XAI layer changes that judgment because it reveals whether the model is at least responding to plausible risk channels. In this case, the answer is yes. The repeated prominence of QQQ volatility, lagged returns, drawdown measures, volatility spread, and rolling correlation indicates that the models were reacting to recognisable ingredients of market stress. This matters because explainability converts a weakly predictive black box into a transparent diagnostic tool.

The finance literature increasingly argues that this diagnostic value is not secondary to prediction quality but part of model quality itself. Rane et al. (2023) and Kashyap and Iveroth (2021) both emphasised that financial institutions require interpretable outputs to satisfy governance, accountability, and control objectives. Leung et al. (2021) further showed that time-series explanation must respect temporal dependence, while Gill et al. (2020) distinguished model interpretability from post-hoc explanation in financial forecasting. The present results speak directly to that debate. The warning system does not merely say that risk is elevated; it shows that elevated risk is associated with recent return weakness, short-horizon volatility, technology-sector drawdown, and tighter cross-market comovement. That is precisely the kind of explanation a practitioner can scrutinise, challenge, and operationalise.

7.2 Relevance for regulators, investors, and financial institutions

The practical significance of the findings is strongest when the model is interpreted as an early-warning support tool rather than an automated trigger. For regulators, the framework demonstrates how a lightweight surveillance system can flag periods when market fragility is rising through synchronised volatility and dependence. Such a system would not replace more comprehensive macroprudential dashboards, but it could complement them by offering a high-frequency interpretive layer that is easy to audit. When the model assigns elevated probabilities because QQQ volatility, return persistence, and market correlation are simultaneously increasing, supervisors receive not only a warning but also a compact narrative about the structure of that warning (Guo et al., 2025).

Institutional investors and risk managers can use the same logic in portfolio monitoring (Kahan & Rock, 2020). The feature-importance results suggested that technology-sector stress has disproportionate informational content for subsequent market-wide fragility in this sample. That does not mean QQQ mechanically causes contagion. It does mean that, when technology-heavy weakness becomes acute and highly correlated with the broader market, hedging and liquidity planning deserve greater attention. The August 2024 cluster of elevated predicted probabilities is illustrative. Even though most of those dates were not classified as positive events in the next-day target, the model detected a fragile regime in which cross-market conditions had deteriorated meaningfully. For practitioners, such clusters could justify tighter position limits, enhanced stress testing, or heightened human review rather than immediate portfolio liquidation (Mahesh, 2025).

7.3 Comparison with prior studies from the last six to seven years

Compared with recent studies in the literature, the present findings are directionally consistent. Research using richer network structures or higher-frequency inputs typically reports stronger systemic-risk discrimination. Georgousis et al. (2021) showed that deep graph-learning approaches outperform more traditional machine-learning methods when network structure is explicitly modelled. Che-Castaldo et al. (2021) demonstrated that one-minute data contain strong information for identifying systemic-risk intensification, while Ran et al. (2024) obtained superior early-warning ability using a nonlinear EEMD-LSTM architecture with a broad indicator system. Relative to these studies, the current framework is deliberately simpler and, unsurprisingly, less powerful. It does not exploit inter-firm networks, intraday flows, option-implied information, or macro-financial covariates.

The findings also speak to recent work on explainability in financial systems. Rane et al. (2023) argued that xAI in finance must be assessed not only by technical novelty but also by its implications for practice and policy. The present study reinforces that claim. Even when classification metrics are constrained, the explanation layer remains informative because it reveals economically coherent transmission channels. This supports the view that explainability is especially valuable in difficult forecasting environments, where decision-makers need to understand why a model is uncertain or why it issues clustered warnings rather than a neat sequence of correct binary calls.

7.4 Practical implications for market surveillance and risk management

Several practical implications follow from the empirical evidence. First, threshold calibration matters. The random forest delivered the best probability ranking on the holdout sample but failed to issue any true positive signal at the conventional 0.50 threshold. This suggests that surveillance systems for rare financial events should not default automatically to a generic classification cutoff. Instead, thresholds should be tuned to the institutional objective: higher sensitivity for supervisory monitoring, higher precision for automated escalation, or dual-threshold systems that distinguish amber alerts from red alerts.

Second, model simplicity and interpretability can be operational assets. In regulatory and institutional settings, implementation delays often arise not because a model lacks statistical value but because stakeholders cannot understand or trust it. The present framework uses a relatively small feature set built from transparent market quantities and then explains predictions using permutation importance. This design makes it easier to embed the model within existing governance processes, including model validation, human review, and scenario analysis. It also reduces the risk that a formally stronger but opaque model will be rejected or ignored by decision-makers.

Third, the results suggest that cross-market monitoring should be prioritised over isolated asset surveillance. The strongest empirical signals did not come from SPY or QQQ in isolation; they came from the interaction between them. Return spreads, rolling correlation, volatility differentials, and technology-sector drawdowns all mattered. This supports a surveillance architecture in which regulators and risk managers track how sector-specific instability becomes synchronised with the broader market. Such a design is compatible with network and spillover research showing that the emergence of systemic risk is better understood through relational structure than through single-asset diagnostics (Åström & Pettersson, 2025).

Finally, the results point toward a layered risk-management workflow. A practical system could combine an interpretable daily-warning model like the one presented here with richer secondary modules: intraday liquidity indicators, options-based stress measures, network connectedness metrics, and scenario-specific narrative overlays. The current study should therefore be read not as a final forecasting engine but as a governance-oriented backbone onto which more specialised risk modules can be attached.

7.5 Interpretability, transparency, and trust in AI-based financial systems

Trust is often discussed as a soft organisational issue, but in finance, it has hard operational consequences. Models that cannot be explained are harder to validate, more difficult to contest, and less likely to be used consistently in high-stakes decisions (French et al., 2024). This is especially true in systemic-risk management, where false alarms can be costly but missed warnings can be catastrophic. The present study contributes to this trust problem by showing that a model can be modest in predictive strength and still valuable if it is transparent about what it is seeing.

This argument resonates with the expanding explainable finance literature. Yeo et al. (2025) documented the rapid spread of XAI across financial tasks, and Rissy (2021) stressed that meaningful explanation in finance requires attention to stakeholder needs, model context, and regulatory relevance. The present results operationalise those concerns. They show that explanation is not an optional visualisation added after modelling. It is the mechanism through which a fragile-signal forecasting exercise becomes usable in practice. In a domain as sensitive as systemic-risk monitoring, that transition from opaque prediction to interpretable judgment may be as important as incremental improvements in AUC or F1.

8. Conclusion

This study contributes to the financial artificial intelligence literature by showing that explainable machine-learning approaches can support more transparent and governance-oriented systemic-risk monitoring in U.S. financial markets. Its significance lies in demonstrating that contagion detection should be judged not only by predictive performance but also by interpretability and decision relevance. By integrating predictive modelling with explainability, the study offers a practical framework for surveillance in high-stakes financial settings. More broadly, it highlights the importance of combining predictive capability with interpretive clarity, making explainable artificial intelligence especially valuable for regulators, portfolio managers, and financial risk analysts.

9. Recommendations

Future implementations should calibrate decision thresholds explicitly for rare-event monitoring rather than relying on the default 0.50 cutoff. The surveillance architecture should also be expanded beyond two ETFs to include financial-sector, credit, options, and macro-financial indicators. Where feasible, intraday data should be incorporated to improve the timing of contagion warnings. From a governance perspective, institutions should pair explainable AI outputs with human review, stress testing, and escalation rules so that model warnings become part of a disciplined supervisory workflow rather than stand-alone automated decisions.

10. Strengths and Limitations

A major strength of the study is its combination of transparent feature engineering, walk-forward validation, holdout testing, and explainability, which together make the evaluation defensible and reproducible. The results are also economically interpretable, not merely statistically reported. The main limitation is the parsimonious data design: two daily ETF series cannot fully represent the networked structure of algorithmic contagion across U.S. financial markets. The short sample, class imbalance, and synthetic target construction further limit generalizability, so the findings should be interpreted as a credible proof of concept rather than a final production model.

10. Future Implications

The next stage of research should integrate richer market networks, higher-frequency observations, and cross-asset stress measures within the same XAI framework. Such extensions could support more precise early-warning systems for regulators and institutional investors while preserving transparency. As AI adoption deepens across trading, portfolio management, and supervision, the strategic challenge will not be whether finance uses AI, but whether it uses AI in ways that remain interpretable, governable, and robust under stress. The present study offers a foundation for that broader agenda.

References

1. Agatón Lombera, D. I., Cardoso López, D. A., López Cabrera, J. A., & Nuñez Mora, J. A. (2024). Market reactions to US financial indices: A comparison of the GFC versus the COVID-19 pandemic crisis. *Economies*, *12*(7), 165.
2. Ahmed, I., Jeon, G., & Piccialli, F. (2022). From artificial intelligence to explainable artificial intelligence in industry 4.0: a survey on what, how, and where. *IEEE transactions on industrial informatics*, *18*(8), 5031-5042.
3. Ahmmed, M. J. (2025). Systematic Review and Quantitative Evaluation of Advanced Machine Learning Frameworks for Credit Risk Assessment, Fraud Detection, And Dynamic Pricing in US Financial Systems. *International Journal of Business and Economics Insights*, *5*(3), 1329-1369.
4. Åström, W., & Pettersson, G. (2025). The Role of Portfolio Overlap and Network Centrality: Systemic risk in Swedish mutual funds. In.
5. Cesa-Bianchi, A., Thwaites, G., & Vicondoa, A. (2020). Monetary policy transmission in the United Kingdom: A high frequency identification approach. *European Economic Review*, *123*, 103375.
6. Che-Castaldo, J. P., Cousin, R., Daryanto, S., Deng, G., Feng, M.-L. E., Gupta, R. K., Hong, D., McGranaghan, R. M., Owolabi, O. O., & Qu, T. (2021). Critical Risk Indicators (CRIs) for the electric power grid: a survey and discussion of interconnected effects. *Environment Systems and Decisions*, *41*(4), 594-615.
7. Chen, M. (2025). *Decision Dynamics Under Uncertainty: Complex-Systems Models for Individual Risk Choices and Organizational Tech-Shock Responses* ETH Zurich].
8. Chen, M., Li, N., Zheng, L., Huang, D., & Wu, B. (2022). Dynamic correlation of market connectivity, risk spillover and abnormal volatility in stock price. *Physica A: Statistical Mechanics and Its Applications*, *587*, 126506.
9. Chhabra, D., & Gupta, M. (2020). Market efficiency and calendar anomalies in commodity futures markets: A review. *Agricultural Economics Research Review*, *33*(2), 263-277.
10. Chung, S., Animesh, A., Han, K., & Pinsonneault, A. (2020). Financial returns to firms' communication actions on firm-initiated social media: Evidence from Facebook business pages. *Information Systems Research*, *31*(1), 258-285.
11. Coupez, E. (2025). The Impact of Artificial Intelligence and Algorithmic Trading on Stock Market Behavior, Volatility, and Stability. *Volatility, and Stability (August 13, 2025)*.
12. EL OUBANI, A. (2022). Market efficiency dynamics during the COVID-19 pandemic. *Revue des Etudes Multidisciplinaires en Sciences Economiques et Sociales*, *7*(2).
13. French, S., Dickerson, A., & Mulder, R. A. (2024). A review of the benefits and drawbacks of high-stakes final examinations in higher education. *Higher Education*, *88*(3), 893-918.
14. Georgousis, S., Kenning, M. P., & Xie, X. (2021). Graph deep learning: State of the art and challenges. *Ieee Access*, *9*, 22106-22140.
15. Gill, N., Hall, P., Montgomery, K., & Schmidt, N. (2020). A responsible machine learning workflow with focus on interpretable models, post-hoc explanation, and discrimination testing. *Information*, *11*(3), 137.
16. Gong, X.-L., Liu, J.-M., Xiong, X., & Zhang, W. (2022). Research on stock volatility risk and investor sentiment contagion from the perspective of multi-layer dynamic network. *International Review of Financial Analysis*, *84*, 102359.
17. Guo, W., Wang, Y., Huang, Z., & Zhang, C. (2025). Trading under uncertainty: A distribution-based strategy for futures markets using futurequant transformer. *arXiv preprint arXiv:2505.05595*.

18. Harré, M. S., Eremenko, A., Glavatskiy, K., Hopmere, M., Pinheiro, L., Watson, S., & Crawford, L. (2021). Complexity economics in a time of crisis: Heterogeneous agents, interconnections, and contagion. *Systems*, *9*(4), 73.

19. Hasan, T., & Jahan, N. (2024). Artificial Intelligence and Ethical Challenges in Financial Services: A Framework for Responsible Data Governance and Algorithmic Transparency. *International Journal of Data Science, Big Data Analytics, and Predictive Modeling*, *14*(8), 16-36.

20. Herring, R., & Walczyński, M. (2024). Emotions, stress and related phenomena in public service interpreter and translator education and training. *FITISPos International Journal*, *11*(1), 9-27.

21. Kacprzyk, K., Liu, T., & van der Schaar, M. (2024). Towards transparent time series forecasting. The Twelfth International Conference on Learning Representations,

22. Kahan, M., & Rock, E. B. (2020). Index funds and corporate governance: Let shareholders be shareholders. *BuL rev.*, *100*, 1771.

23. Kashyap, S., & Iveroth, E. (2021). Transparency and accountability influences of regulation on risk control: the case of a Swedish bank. *Journal of Management and Governance*, *25*(2), 475-508.

24. Komati, D. (2025). Real-Time AI Systems for Fraud Detection and Credit Risk Management: A Framework for Financial Institutions. *IJSAT-International Journal on Science and Technology*, *16*(1).

25. Leung, K. K., Rooke, C., Smith, J., Zuberi, S., & Volkovs, M. (2021). Temporal dependencies in feature importance for time series predictions. *arXiv preprint arXiv:2107.14317*.

26. Liu, A., Chen, J., Yang, S. Y., & Hawkes, A. G. (2020). The flow of information in trading: An entropy approach to market regimes. *Entropy*, *22*(9), 1064.

27. Mahesh, M. (2025). Portfolio and Risk Management. *LawFoyer Int'l J. Doctrinal Legal Rsch.*, *3*, 759.

28. Marzouki, M. M. (2024). Towards a Reinvigoration of the Risk Management Framework for the Scope of Upgrading Prudential Surveillance. *International Journal of Finance & Banking Studies*, *13*(3).

29. Meng, K., & Li, S. (2021). The adaptive market hypothesis and high frequency trading. *PloS one*, *16*(12), e0260724.

30. Meng, S., & Chen, Y. (2023). Market volatility spillover, network diffusion, and financial systemic risk management: Financial modeling and empirical study. *Mathematics*, *11*(6), 1396.

31. Nestor, T. (2025). AI Models Are Not Conscious and Are Massively Inefficient, Causing Complexity and Scalability Bottlenecks to Artificial General Superintelligence Risking Technological Singularity and Loss of Societal Dynamism or Institutional Collapse and Renewal. *Causing Complexity and Scalability Bottlenecks to Artificial General Superintelligence Risking Technological Singularity and Loss of Societal Dynamism or Institutional Collapse and Renewal (June 16, 2025)*.

32. Qi, L. (2023). *On Nonlinearities in Financial Market Co-Movements* The Australian National University (Australia)].

33. Ran, M., Tang, Z., Chen, Y., & Wang, Z. (2024). Early warning of systemic risk in stock market based on EEMD-LSTM. *PloS one*, *19*(5), e0300741.

34. Rane, N., Choudhary, S., & Rane, J. (2023). Explainable Artificial Intelligence (XAI) approaches for transparency and accountability in financial decision-making. *Available at SSRN 4640316*.

35. Recchia, M. (2021). FINTECH COMPANIES: INNOVATION, ALGORITHMS AND CUSTOMER CENTRIC PERSPECTIVE-A cross-sectional study on algorithmic trading in the Fintech industry.

36. Rissy, Y. Y. W. (2021). The stakeholder model: its relevance, concept, and application in the Indonesian banking sector. *Journal of Banking Regulation*, *22*(3), 219-231.

37. Sadeghi, A. (2025). *Contagion Dynamics in Traditional and Decentralized Finance* Stevens Institute of Technology].

38. Sanz Martin, L., Parra Dominguez, J., Corchado, J. M., Zafra-Gómez, E., Castillo-Ramos, V., & Zafra-Gómez, J. L. (2025). Recent evolution and growth of AI and advanced technologies in accounting and finance: systematic review and bibliometric analysis. *Spanish Journal of Finance and Accounting/Revista Española de Financiación y Contabilidad*, 1-42.

39. Serrano, A. S. (2020). High-frequency trading and systemic risk: A structured review of findings and policies. *Review of Economics*, *71*(3), 169-195.

40. Shahzad, S. J. H., Bouri, E., Kristoufek, L., & Saeed, T. (2021). Impact of the COVID-19 outbreak on the US equity sectors: Evidence from quantile return spillovers. *Financial Innovation*, *7*(1), 14.

41. Siddiqui, S. A., Mercier, D., Dengel, A., & Ahmed, S. (2021). Tsinsight: A local-global attribution framework for interpretability in time series data. *Sensors*, *21*(21), 7373.

42. Silvio, F. (2022). *Time series analysis using explainable AI* University of Malta].

43. Sujatha, D., Sudheer, A. K., & Balasundaram, E. (2025). Black swan dynamics: a network-based framework for systemic risk detection and mitigation: D. Sujatha et al. *Risk Management*, *27*(4), 24.

44. Sun, T. (2025). Systemic risk of systemically important financial institutions in the post-2008 global financial crisis era: A tail risk network analysis. *Journal of Risk and Insurance*, *92*(4), 950-977.

45. Theissler, A., Spinnato, F., Schlegel, U., & Guidotti, R. (2022). Explainable AI for time series classification: a review, taxonomy and research directions. *Ieee Access*, *10*, 100700-100724.

46. Thirunagalingam, A. (2022). Enhancing Data Governance Through Explainable AI: Bridging Transparency and Automation. *Available at SSRN 5047713*.

47. Tzagkarakis, G., Lydaki, E., & Maurer, F. (2024). Quantifying the Predictive Capacity of Dynamic Graph Measures on Systemic and Tail Risk. *Computational Economics*, 1-31.

48. Verma, R., & Pandiya, D. K. (2024). The role of AI and machine learning in US financial market predictions: Progress, obstacles, and consequences. *International Journal of Global Innovations and Solutions (IJGIS)*.

49. Wang, T., Zhao, S., Zhu, G., & Zheng, H. (2021). A machine learning-based early warning system for systemic banking crises. *Applied economics*, *53*(26), 2974-2992.

50. Yan, H., Liu, B., Zhu, X., & Wu, Y. (2024). Systemic risk monitoring model from the perspective of public information arrival. *The North American Journal of Economics and Finance*, *72*, 102141.

51. Yeo, W. J., Van Der Heever, W., Mao, R., Cambria, E., Satapathy, R., & Mengaldo, G. (2025). A comprehensive review on financial explainable AI. *Artificial Intelligence Review*, *58*(6), 189.

52. Zhou, X., & Liu, H. (2025). Cross-Market Risk Spillovers and Tail Dependence Between US and Chinese Technology-Related Equity Markets. *International Journal of Financial Studies*, *13*(4), 242.