

Market Regime-Aware Explainability for AI-Driven Financial Decisions: A Counterfactual and Uncertainty-Based Framework

Hadi Hasan Hadi

Damghan University, Computer Science (Soft computing and artificial intelligence), Iran. Email: hadihasn2000@gmail.com

ARTICLE INFO

Article history:

Received: 18 /03/2026

Revised form: 20 /06/2026

Accepted : 22 /06/2026

Available online: 30 /09/2026

Keywords:

market regime detection

explainable artificial intelligence
counterfactual

explanations uncertainty
quantification

financial decision-making; financial
time series; risk-aware AI

algorithmic trading; interpretable
machine learning; trustworthy AI.

ABSTRACT

AI models now play a key role in financial decisions, including buy, sell, and hold recommendations, risk monitoring, market direction prediction, and portfolio rebalancing. In financial domains, though, high predictive precision isn't enough because financial decision-makers want to know why a model made a decision, whether the reasons behind the decision are sound given current market conditions, and what type of uncertainty is present in the model's output. Current explainable AI (XAI) methods in finance often provide static feature-attribution explanations. In contrast, financial markets are dynamic and non-stationary, exhibiting latent regimes such as bullish, bearish, sideways, and high-volatility phases. We present RACU-XAI, a framework for generating market regime-aware frameworks that combines market regime detection, Artificial Intelligence-related financial decision modeling, regime-conditioned explanations, counterfactual explanation generation, and uncertainty quantification. The framework is validated on daily market data for the S&P 500, NASDAQ-100, and Bitcoin/USD, using engineered price, technical, risk, and regime features. We proposed a three-class classification problem over a five-trading-day horizon for the main decision task. The findings indicate that RACU-XAI achieves the best predictive performance on S&P 500 data, with an accuracy of 0.724, an F1 score of 0.712, and an AUC of 0.769. Its performance improves financial performance after transaction costs, and the cumulative return is 61.4%, the Sharpe ratio is 1.31, and the maximum drawdown is -12.6%. Regime-aware SHAP analysis shows that the explanatory drivers vary widely across market conditions, and counterfactual explanations indicate the minimum changes required to shift the decision from Sell to Hold or from Buy to Hold. Finally, uncertainty estimates are strongly associated with error rates, suggesting that uncertainty could serve as a suitable risk-filtering mechanism. The results indicated that conditioning explanations on market regimes in AI-based decision support systems can reinforce transparency, stability, and risk consciousness in financial markets.

MSC..

<https://doi.org/10.29304/jqcm.2026.18.32907>

1. Introduction

AI financial decision systems are the cornerstones of computational finance. Deep learning, ensemble learning, and hybrid machine learning approaches are widely utilized to predict financial time series, estimate downside risk, detect anomalies, and construct trading signals. Recent research indicates that deep learning techniques have gained significant importance in financial forecasting and related applications due to their ability to learn non-linear and temporal relationships from complex data [1], [2]. And this performance advantage is a huge governance challenge — high-performance financial AI models tend to behave like black boxes, meaning that investors, analysts,

*Corresponding author

Email addresses:

Communicated by 'sub editor'

auditors, and regulators cannot understand how certain outcomes are coming out. Explainable AI is therefore an essential part of financial model deployment. The question of interpretability in finance isn't just a technical problem but about trust, accountability, model risk, and compliance in high-impact decision-making settings [3], [4]. Finance-specific XAI reviews indicate that explainability is critical, as it is needed for AI-based decisions to be meaningful, traceable, and credible to even well-meaning or otherwise non-machine-learning expert users [5]. Too many of today's XAI implementations though, are still static. They generate one global reason or one local feature-attribution vector, without specifying that the rationale depends on (i) the current market state. This limitation is important because financial markets are non-stationary. And the same indicator can appear different across regimes. While momentum may function well in a sustained bullish phase, it can become fragile during periods of crisis or heightened volatility. Similarly, during bearish or turbulent periods, volatility, drawdowns, and Value at Risk may account for the majority of possible explanations. Models of regime-switching and hidden-state approaches have been applied for long periods to model latent market phases, and recent work continues to suggest that the identification of regimes can enhance risk interpretation and support financial decision-making [6], [7]. Thus, a technical explanation that ignores regimes may be technically faithful to the model, but economically incomplete. Another limitation is that explanations are often descriptive not practical. Tools like SHAP – for feature attribution – can highlight influential variables. Still, the user should generally be told what the minimum change in market conditions would need to be to achieve their desired outcome. Counterfactual explanations fulfill this need by describing what would have to change in the input to change the output [8], [9]. In financial planning, counterfactuals can express model values as realistic scenario statements: If volatility were reduced and momentum were back on track, the decision would be to change from Sell to Hold. The third limitation is the lack of inclusion of uncertainty. Markets in the financial world are noisy, evolving, and responsive to abrupt and unexpected shocks. A model can produce a positive, confident class label even when the predictive context is unstable. Modern uncertainty quantification work distinguishes reducible epistemic uncertainty from irreducible aleatoric uncertainty and highlights the need for calibration in high-risk AI systems [10], [11]. Uncertainty shouldn't be used as an optional diagnostic for financial decision support; it should be part of the decision itself. To address these limitations, we propose a common framework, RACU-XAI, which stands for Regime-Aware Counterfactual and Uncertainty-based Explainable AI. We connect five facets in the framework: market regime detection, AI decision modeling, regime-aware explanation, counterfactual explanation, and uncertainty-aware risk filtering. The goal here is not just to enhance predictive power but also to develop transparent, regime-sensitive, actionable, and risk-responsive decisions. The research problem is that no existing AI-based decision models in the financial decision-making process include a dynamic explainability framework that jointly conditions explanations on market regimes, produces counterfactual insights in action, and quantifies uncertainty for each decision. This disconnects trust and constrains the operational application of AI to the challenging financial landscape. There are four broad contributions of the study. One, it proposes an integrated, regime-aware XAI framework for financial decision-making rather than a single predictive model. Second, it maps local explanations to detected market regimes, enabling users to determine how decision drivers change when entering the market in bullish, bearish, sideways, or high-volatility states. Third, it produces constrained counterfactual explanations that quantify the minimum plausible change needed to change a decision. Fourth — it introduces uncertainty quantification through a risk-control layer that filters out low-confidence decisions, thereby improving risk-adjusted financial performance.

2. Literature Review and Research Gap

2.1 AI and deep learning in financial decision-making

Financial AI has advanced from classical statistical learning to complex deep learning and ensemble architectures. As indicated in a survey, recurrent neural networks, convolutional neural networks, gradient-boosting models, and transformer-based models have been extensively used for stock forecasting, risk prediction, and algorithmic trading [1], [2]. As mentioned, deep learning is especially well-suited for financial time-series tasks due to its capacity to capture nonlinear dependencies and temporal structure. At the same time, studies on time series generally emphasize the need for sequence-aware architectures for forecasting tasks [12]. Yet financial forecasting is not like many conventional machine-learning problems. The financial data is relatively noisy, non-stationary, and heavy-tailed, and structural breaks may affect it. On average, a successful model can easily break down during a crisis and not generalize across regimes. In this regard, predictive accuracy would be compared with financial metrics such as cumulative returns, Sharpe ratio, Sortino ratio, maximum drawdown, and transaction-cost-adjusted returns.

2.2 Explainable AI in finance

The purpose of XAI in finance is to get stakeholders to understand AI decisions. Finance-specific work has argued that explainability is necessary for credit risk, statistical arbitrage, portfolio decision support, regulatory oversight, and model-risk governance [3], [13], [14]. SHAP-based explanations are a popular method because they can give both the local and global feature attribution and are adaptable to tree-based models [15]. However, a static feature attribution mechanism may fail in the financial domain, since a stable feature classification does not necessarily capture the characteristics of market fluctuations. According to recent systematic reviews, financial XAI still faces open issues such as a lack of clear evaluation criteria, limited use by users, insufficient temporal dependence, and limited adoption of techniques for managing risks [4], [5]. For time series, this limitation is more evident, as explanations must take into consideration the feature and time context. In the context of XAI for time series, explanation methods are classified into time-point, subsequence, and instance-based methods [16], and better evaluation protocols need to be implemented.

2.3 Market regime detection

Market regime detection models presume an underlying latent-state behavior, i.e., that financial dynamics are determined primarily by latent states. To identify bullish, bearish, calm, and turbulent regimes, Hidden Markov Models (HMMs), Gaussian Mixture Models (GMMs), Markov-switching models, clustering methods, and, more recently, Wasserstein distance approaches have been employed [7]. To this end, regime detection is especially useful for modeling variations in returns, volatility, transition probabilities, and market stress. More recently, research has also explored early-warning signals for regime switching in complex financial systems [6]. Notwithstanding this advancement, regime information is sometimes employed as a measure of strategy change or risk allocation rather than explainability. A regime-aware trading model can shift positions between regimes, but it cannot, by itself, explain why a decision is issued differently in one regime than in another. The present study extends regime detection from a predictive or allocation tool to an explanation-conditioning mechanism.

2.4 Counterfactual explanations

Counterfactual explanations answer a practical question: what is the minimum change to an input that would cause the model to produce a different outcome? Recent surveys classify counterfactual methods according to distance, sparsity, actionability, plausibility, diversity, and validity [8], [9]. Algorithmic recourse research further emphasizes that counterfactuals can help users understand the consequences of decisions by offering contrastive recommendations [17]. Diverse and multi-objective counterfactual methods have been proposed to avoid overly narrow explanations and to account for feasibility and plausibility constraints [18], [19]. Financial counterfactuals require additional constraints. For example, a counterfactual that increases the historical price or arbitrarily changes a lagged observation may be mathematically valid but economically implausible. Therefore, counterfactual search in finance must respect domain constraints, such as plausible volatility ranges, indicator dependencies, and regime-consistent feature bounds.

2.5 Uncertainty quantification

Uncertainty quantification is an ever-growing need in high-risk AI systems. Reviews identify epistemic uncertainty, which captures model uncertainty and can be overcome by more information or more accurate models [10], [11], and aleatoric uncertainty, which is irreducible noise in the process of data generation. Methods that are used to estimate predictive uncertainty include deep ensembles, Monte Carlo Dropout, Bayesian neural networks, quantile regression, and conformal prediction [20]. Uncertainty plays a critical role in financial decision-making, as markets are dynamic. In this way, the decision must include a class label, a confidence score, a level of uncertainty, and a risk label. A more helpful model than one that always makes high-confidence decisions absorbs uncertainty during turbulent regimes.

2.6 Research gap

According to the reviewed literature, financial forecasting, XAI, regime identification, counterfactual explanations, and uncertainty quantification are current research problems. Nonetheless, they are most often studied independently. Current approaches rarely give one decision-support system that: (a) determines the current market regime, (b) constrains the explanations based on that regime, (c) takes the financial indicators counterfactuals that are plausible, and (d) accounts for the uncertainty of each of the decisions. This paper fills this gap by using the RACU-XAI framework.

Table 1. Positioning of RACU-XAI relative to related research streams

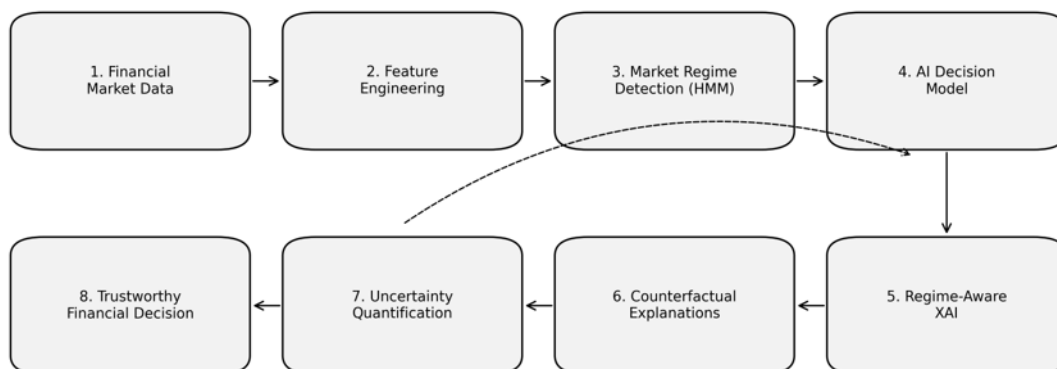
Research stream	Main focus	Limitations in relation to this study	RACU-XAI contribution
Financial forecasting models	Predicting market direction, returns, or trading signals	Often emphasize predictive accuracy without providing regime-conditioned explanations	Integrates prediction with regime-aware explainability and risk filtering
Regime-switching and hidden-state models	Detecting latent bullish, bearish, sideways, or volatile market states	Usually used for allocation or risk monitoring rather than explanation conditioning	Uses detected regimes as contextual information for both SHAP explanations and counterfactual generation
Explainable AI in finance	Explaining model outputs using feature-attribution methods such as SHAP or LIME	Explanations are frequently static and may ignore market non-stationarity	Produces regime-conditioned explanations showing how drivers differ across market states
Counterfactual explanations	Identifying minimal input changes required to alter a model decision	Generic counterfactuals may be mathematically valid but financially implausible	Adds financial and regime-consistency constraints to produce plausible decision-change scenarios
Uncertainty-aware machine learning	Estimating predictive uncertainty and calibration in high-risk AI systems	Often treated as a diagnostic output rather than part of the decision rule	Uses uncertainty as a risk-control layer to filter unreliable financial decisions

Unlike these streams, the proposed RACU-XAI framework integrates them into a single decision-support pipeline. Its novelty lies not solely in its use of a stronger predictive model, but in its integration of four aspects: the detection of market regimes, the explanation of regimes, the counterfactual reasoning in finance, and the filtering of decisions based on uncertainty. This integration enables the model to not only generate Buy, Hold, or Sell predictions but also an interpretable, risk-aware, conditioned decision package conditioned on the prevailing market trend.

3. Proposed Racu-Xai Framework

3.1 Conceptual overview

RACU-XAI is a multi-layered financial decision support pipeline. The framework does not just provide an output of Buy, Sell, or Hold; it provides a structured decision package that includes the predicted decision, the current market regime, the confidence in the predicted decision, the degree of uncertainty, the primary explanation, counterfactual insight, and the risk label.



Dashed feedback path: uncertainty-aware risk filtering and calibration

Fig 1. Conceptual architecture of the proposed RACU-XAI framework.**Table 2. Main components of the RACU-XAI framework**

Component	Purpose	Output
Data preprocessing	Clean daily financial data and construct price, technical, and risk features	Feature matrix and target labels
Market regime detection	Identify latent market states from return, volatility, volume, and trend variables.	Regime label z_t and transition probabilities
AI decision model	Predict Buy, Hold, or Sell from engineered features and regime information.	Class probabilities $\hat{p}_{t,c}$
Regime-aware XAI	Explain decisions within each detected market regime	Regime-conditioned feature attributions
Counterfactual explanation	Estimate the minimum plausible changes needed to alter the decision	Counterfactual vector $\tilde{x}_t$
Uncertainty quantification	Estimate predictive uncertainty and filter high-risk decisions	Uncertainty score U_t and risk label

3.1.1 Implementation and Reproducibility Details

The RACU-XAI framework was designed as a sequential pipeline with six main stages to improve reproducibility. First was the collection of daily financial information and its chronological organization. Second, missing observations were filled in using forward-fill, and all technical and risk indicators were calculated using past and current data to avoid look-ahead bias. Third, the model fitted for market-regime detection was used with variables related to return, volatility, volume change, and trend. Fourthly, the determined regime label and regime probability were added to the predictive model along with the engineered financial features. Fifth, SHAP-based explanations were calculated individually for each regime identified in the market to obtain regime-conditioned feature importance. Sixth, counterfactual explanations and uncertainty estimates were produced for each decision, and decisions with uncertainty estimates exceeding the chosen uncertainty threshold were filtered.

A chronological training-validation-testing split was used to apply the pipeline. The selection of hyperparameters was performed only on the validation period, and the testing period was reserved for final out-of-sample evaluation. The design was chosen to minimize time leakage and to simulate financial forecasting conditions.

3.2 Notation and symbol definitions

Table 3. Core mathematical notation used in the study

Symbol	Definition
t	Trading-day index
h	Forecast horizon; in this study, $h = 5$ trading days
C_t	Closing price at time t

Symbol	Definition
r_t	Simple daily return
ℓ_t	Log return
$x_t \in \mathbb{R}^p$	Feature vector at time t
$X_t = [x_{t-w+1}, \dots, x_t]$	Rolling input window of length w
$z_t \in \{1, \dots, K\}$	Latent market regime at time t
y_t	Encoded decision label: +1 = Buy, 0 = Hold, -1 = Sell
τ	Decision threshold; in this study $\tau = 0.005$
$f_\theta(\cdot)$	AI decision model with parameters θ
$\hat{p}_{t,c}$	Predicted probability for class c
$\phi_{j,t}$	SHAP contribution of feature j at time t
δ_t^*	Optimal counterfactual perturbation
U_t	Predictive uncertainty score
s_t	Trading position implied by the decision at time t

3.3 Return and target construction

The simple daily return and log return are defined as follows:

$$r_t = \frac{C_t - C_{t-1}}{C_{t-1}}, \quad \ell_t = \ln\left(\frac{C_t}{C_{t-1}}\right). \quad \dots (1)$$

The future return over horizon h is correctly defined as the relative change from the current close to the future close:

$$R_{t,h} = \frac{C_{t+h} - C_t}{C_t}. \quad \dots (2)$$

For the main experiments, $h = 5$ and $\tau = 0.005$. The classification target is:

$$y_t = \begin{cases} +1, & R_{t,h} > \tau, \\ -1, & R_{t,h} < -\tau, \\ 0, & -\tau \leq R_{t,h} \leq \tau. \end{cases} \quad \dots (3)$$

The encoded labels +1 and -1 correspond to Buy and ll, respectively. This threshold reduces sensitivity to very small daily fluctuations and better reflects realistic trading conditions where transaction costs and bid-ask spreads matter.

3.4 Market regime model

The Hidden Markov Model (HMM) is a latent Markov model that is directly probabilistic about latent market states and about regime transitions, so it is adopted as the final regime-identification model for the market regime detection stage. The regimes considered were set at four to capture economically interpretable states: bullish low-volatility, bearish, sideways, and high-volatility. Regime detection is based on the observation vector consisting of daily return, 20 day rolling volatility, volume change, 20 day moving-average deviation and moving-average slope. These variables were chosen for their ability to combine a measure of direction with risk, liquidity pressure, and trend behavior.

Three other clustering algorithms, namely K-Means, Gaussian Mixture Model (GMM) and Agglomerative Clustering, were compared with the HMM. A higher Davies-Bouldin Index, an acceptable Silhouette Score, stronger economic interpretability, and regime stability were the four criteria used for the final choice of HMM. Regime stability was determined by assigning a penalty to high frequencies of regime switches between adjacent day trading periods, and economic interpretability was investigated to ensure that each regime can be economically interpreted, e.g. as upward trend, downward pressure, low directionality or volatility stress.

The HMM output for each day t is a discrete regime label and posterior regime probabilities. These outputs are then fed to the predictive model as context and used as conditioning variables in the explanation and counterfactual phases.

Let z_t denote the hidden market regime, at time t . In the HMM specification, one would have the transition probability of regime i to regime j as:

$$Pr(z_t = j \mid z_{t-1} = i) = A_{ij}, \quad \sum_{j=1}^K A_{ij} = 1. \quad \dots (4)$$

The observation vector that is observed to detect the regime is:

$$g_t = [\ell_t, \sigma_{t,20}, \Delta V_t, \nabla m_{t,20}], \quad \dots (5)$$

where $\sigma_{t,20}$ is 20-day rolling volatility, ΔV_t is volume change, $m_{t,20}$ is the 20-day moving average, and $\nabla m_{t,20}$ is its slope. Conditional on regime k , the observations are modeled as:

$$g_t \mid z_t = k \sim \mathcal{N}(\mu_k, \Sigma_k). \quad \dots (6)$$

3.5 Decision model

The predictive model receives engineered features and the detected regime as context:

$$\hat{p}_t = f_\theta(X_t, z_t), \quad \hat{y}_t = \underset{c \in \mathcal{C}}{\operatorname{argmax}} \hat{p}_{t,c}, \quad \mathcal{C} = \{+1, 0, -1\}. \quad \dots (7)$$

+1, 0 and -1 in this coding are Buy, Hold and Sell respectively. The regime category can be added as an explicit categorical variable, probability response of regime memberships, or an interaction value between regime and the features chosen.

3.6 Regime-aware explanation

But in the case that the model output $f(x_t)$ can be expressed in the form of SHAP then it generates a feature contribution and a line contribution:

$$f(x_t) = \phi_0 + \sum_{j=1}^p \phi_{j,t}. \quad \dots (8)$$

The regime-sensitive value of feature j within regime k is determined by taking the average absolute contribution across the observations in that regime:

$$\bar{\phi}_{j,k} = \frac{1}{|\mathcal{D}_k|} \sum_{t: z_t = k} |\phi_{j,t}|, \quad \mathcal{D}_k = \{t: z_t = k\}. \quad \dots (9)$$

The formulation tests directly the location of the most influential drivers to vary across market states.

3.7 Counterfactual explanation

Given original feature vector x_t , a desired alternative class c' and a regime-consistency feasible set $\mathcal{F}_{z_t}$ the constrained counterfactual problem is:

$$\delta_t^* = \underset{\delta}{\operatorname{argmin}} [\lambda_1 \|\delta\|_1 + \lambda_2 \|\delta\|_2^2 + \lambda_3 \mathcal{P}(x_t + \delta; z_t)] \quad \dots (10)$$

subject to:

$$\underset{c \in \mathcal{C}}{\operatorname{argmax}} f_{\theta}(x_t + \delta, z_t)_c = c', \quad x_t + \delta \in \mathcal{F}_{z_t}. \quad \dots (11)$$

In this case, $\|\delta\|_1$, $\|\delta\|_2^2$, and $\mathcal{P}(\cdot)$ persuade to give sparse explanations, huge changes, and unrealistically financial values, respectively.

The counterfactual generation process was limited to achieve financial feasibility. Identifiers, target variables and non-actionable historical labels were set up as constant but other features that could be assumed to change over the decision horizon were varied. The counterfactual candidates had to be in the regime-consistent bounds that were computed from the empirical distribution of the regime in question. For instance, if a high volatility counterfactual led to a decrease in volatility out of the normal range of the regime detected, this was not accepted unless the counterfactual target was a plausible move toward a less risky decision.

An objective function of validity, proximity, sparsity and plausibility, was minimized in the counterfactual search. Validity means that the generated counterfactual causes the model decision to the desired target class to change. Since the features need to be within proximity, the amount of changes required is small. If the matrix is sparse, the explanation is likely to change only a few variables. Plausibility penalizes technically and risk indicators that appear to be financially unrealistic combinations.

3.8 Uncertainty quantification

Using a deep ensemble of M models, the mean class probability is:

$$\bar{p}_{t,c} = \frac{1}{M} \sum_{m=1}^M p_{t,c}^{(m)}. \quad \dots (12)$$

Predictive confidence κ_t and entropy are:

$$\kappa_t = \max_{c \in \mathcal{C}} \bar{p}_{t,c}, \quad H(\bar{p}_t) = - \sum_{c \in \mathcal{C}} \bar{p}_{t,c} \ln(\bar{p}_{t,c}). \quad \dots (13)$$

An ensemble-variance uncertainty score is defined as:

$$U_t = \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \operatorname{var}_{m=1}^M (p_{t,c}^{(m)}). \quad \dots (14)$$

A decision is carried out when: $U_t \leq U_{\max}$ (15)

In the reported experiments, the empirical optimal return-risk balance is $U_{\max} = 0.25$.

Two complementary approaches, deep ensembles and Monte Carlo Dropout, were used to estimate uncertainty. The ensemble component models uncertainty by averaging the predictions of multiple independently trained models, and the MC Dropout models uncertainty by averaging predictions from stochastic forward passes. Final uncertainty score is a combination of predictive entropy and ensemble variance. The higher the uncertainty score, the less certainty the model has in its class assignment, and/or the more disagreement there is among ensemble members.

Uncertainty is not just used as a diagnostic in the trading decision layer. Rather, it's used as a risk-filtering mechanism. A decision is implemented only if the predictive uncertainty is less than the chosen threshold. This enables the framework to avoid making highly uncertain predictions, particularly in turbulent or ambiguous market conditions.

3.9 Financial evaluation

The strategy return after transaction costs is:

$$r_{t+1}^{strat} = s_t r_{t+1} - \kappa |s_t - s_{t-1}|, \quad s_t \in \{-1, 0, 1\}, \quad \dots (16)$$

where $\kappa = 0.001$ is the transaction cost per trade. The annualized Sharpe ratio is:

$$SR = \frac{\bar{r}_{strat} - r_f}{\sigma_{strat}} \sqrt{252}. \quad \dots (17)$$

Maximum drawdown is calculated as:

$$MDD = \min_t \left(\frac{V_t}{\max_{s \leq t} V_s} - 1 \right), \quad \dots (18)$$

where V_t is the cumulative portfolio value at time t .

The trading rule uses the forecasted class to determine a trading position. Buy signals are used in the form of a long position; Sell signals are used in the form of an exit/short/defensive position (depending on the tested strategy specification); and Hold signals are used to maintain the previous position or a neutral position. Transaction costs are deducted on any change of position from day to day. The values for financial indicators are calculated after transaction costs to reflect the implementation process better.

Risk management is captured by maximum drawdown, Sharpe ratio and Sortino ratio, win rate and uncertainty-based filtering. The goal of the financial evaluation is not to claim a fully deployable trading system, but rather to determine whether the proposed explainability and uncertainty components enhance the quality of financial decisions using AI under realistic cost and risk scenarios.

4. Data and Preprocessing

4.1 Data description

The experimental design is based on daily financial time series from three market environments: a general U.S. equity index, a technology-focused equity index, and a cryptocurrency market. This option would challenge the idea of whether the suggested framework can be applied across various volatility and liquidity regimes.

Table 4. Experimental datasets

Financial asset	Symb ol	Period	Records cleaning	after	Market type
S&P 500 Index	^GSP C	2014-01-01 to 2024-12- 31	2,768		Broad equity market
NASDAQ-100 Index	^NDX	2014-01-01 to 2024-12- 31	2,768		High-growth market technology
Bitcoin/USD	BTC- USD	2016-01-01 to 2024-12- 31	3,287		High-volatility digital asset

The raw variables include open, high, low, close, adjusted close, and volume. Missing values are handled using forward filling to preserve the temporal order of the sequence.

4.2 Feature engineering

Pricing-based variables, technical indicators, risk indicators, and regime-related variables are among its features. Simple and exponential moving averages, RSI, MACD, Bollinger Bandwidth, ATR, momentum, and rolling volatility are considered technical indicators. Variables such as drawdown, Value at Risk, risk-adjusted return, and volume ratio are risk indicators.

Table 5. Feature groups used in the decision model

Feature group	Examples	Number of features
Price-based features	Open, high, low, close, volume, daily return, log return, intraday range	8
Technical indicators	SMA, EMA, RSI, MACD, Bollinger Bands, ATR, momentum, rolling volatility	18
Risk indicators	VaR-95, drawdown, rolling volatility, volume ratio, risk-adjusted return	7
Regime-related features	Regime label, regime probability, transition probability, stress level	4
Total		37

4.2.1 Preprocessing and Feature Construction Pipeline

The preprocessing procedure preserved the temporal structure of the financial series. The observations were first grouped by date, and missing values were filled in using forward-fill. The technical indicators were calculated with rolling windows, allowing only the use of data available up to time t for each feature at time t . This avoids “look-ahead bias”. The parameters were estimated from the training set alone, and the same transformation was applied to continuous features in the validation and testing sets. The categorical regime labels were coded as one-hot or numeric variables, depending on the predictive model. Continuous contextual variables, such as regime probabilities, were kept.

Finally there are variables related to price, technical indicators, risk indicators and regime-related variables. Price-based features characterize daily market movement; technical indicators reflect the market's trend and momentum; risk indicators quantify downside risk and volatility; and regime-related features provide context for the market state.

4.3 Target distribution

Equation 3 was used to construct the target variable for the S&P 500 index dataset, resulting in a balanced class distribution.

Table 6. Target-class distribution for S&P 500 data

Class	Number of observations	Percent age
Buy	1,037	37.5%
Hold	921	33.3%

Class	Number of observations	Percentage
Set II	810	29.2%
Total	2,768	100.0%

4.4 Temporal data split

Chronologically ordered data is subdivided into training, verification, and testing subsets to prevent time leakage.

Table 7. Temporal data split

Split	Percentage	S&P 500 records	Purpose
Training set	70%	1,937	Model fitting
Validation set	15%	415	Hyperparameter tuning and model selection
Testing set	15%	416	Final out-of-sample evaluation
Total	100%	2,768	

5. Experimental Design

5.1 Regime detection experiment

Four regime-detection models are compared: K-Means, GMM, HMM, and Agglomerative Clustering. Silhouette Score, Davies-Bouldin Index, regime stability, and economic interpretability are used for comparison. Regime stability measures whether the method avoids excessive state flipping; economic interpretability evaluates whether the discovered states align with meaningful financial conditions.

Table 8. Comparison of market regime detection methods

Method	Silhouette Score	Davies-Bouldin Index	Regime Stability	Economic Interpretability
K-Means	0.421	0.91	0.68	3.7/5
GMM	0.438	0.84	0.72	4.0/5
Agglomerative	0.397	0.96	0.65	3.5/5
HMM	0.452	0.79	0.81	4.5/5

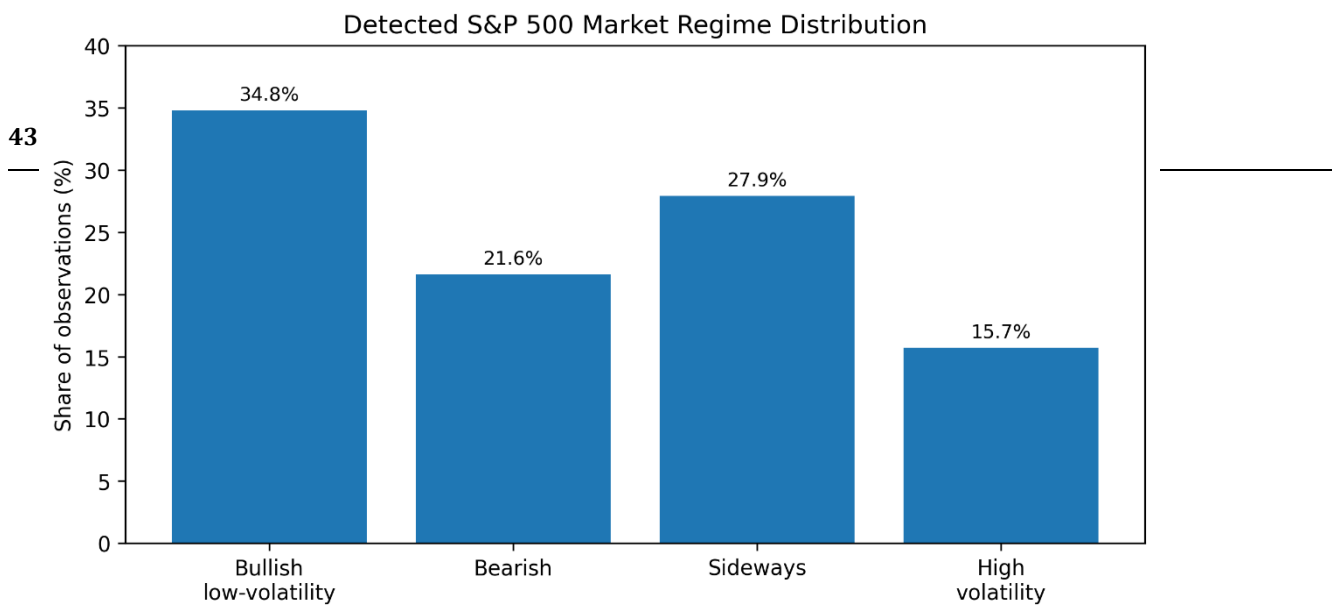


Fig 2. Distribution of HMM-detected market regimes in the S&P 500 dataset.

5.2 Predictive models and baselines

The predictive experiment is discussed, comparing the classical, deep-learning, and proposed models. Baselines consist of Logistic Regression, Random Forest, XGBoost, LightGBM, LSTM, GRU, CNN-LSTM, and time-series Transformer. RACU-XAI takes XGBoost or Transformer as a baseline predictor and introduces regression explanation, counterfactual explanation, and uncertainty-aware decision filtering with regime awareness.

5.3 Training setup

Table 9. Main training and evaluation settings

Setting		Value	
Input length	window	30 days	trading
Forecast horizon		5 days	trading
Decision threshold		$\pm 0.5\%$	
Deep-learning epochs		100	
Batch size		32	
Learning rate		0.001	
Early stopping patience		10	
Deep ensemble size		5	
MC Dropout passes		50	
Transaction cost		0.10% per trade	

5.3.1 Hyperparameter Selection and Validation Strategy

The hyperparameters were tuned on the validation set, and the test set was reserved solely for testing the selected model. This design prevents information leakage and allows for a more realistic out-of-sample evaluation. The primary hyperparameters are the regularization strength for Logistic Regression; the number of trees and the maximum depth for Random Forest; and the learning rate, tree depth, and number of estimators for XGBoost and LightGBM for classical models. The primary hyperparameters for deep learning models were input window length, hidden units, dropout, batch size, learning rate, number of epochs, and early-stopping patience.

The chosen configuration was 30 trading days for the input window, 5 trading days for the prediction horizon, learning rate of 0.001, a batch size of 32, early stopping patience of 10 epochs, ensemble size of 5 and 50 Monte Carlo Dropout passes. These settings were selected as optimal from a validation perspective for predictive accuracy, financial performance, and uncertainty calibration.

Table 10. Main hyperparameters used in the experimental setup

Model/component	Main hyperparameters	Selected values
Logistic Regression	Regularization	L2 regularization
Random Forest	Number of trees, maximum depth	300 trees, tuned maximum depth
XGBoost	Number of estimators, learning rate, maximum depth	300 estimators, 0.05 learning rate, tuned depth
LightGBM	Number of estimators, learning rate, leaves	300 estimators, 0.05 learning rate, tuned leaves
LSTM/GRU	Window length, hidden units, dropout	30 days, tuned hidden units, dropout applied
Transformer	Input window, attention heads, feed-forward dimension	30 days, tuned architecture
Uncertainty estimation	Ensemble size, MC Dropout passes	5 ensemble members, 50 MC passes
Trading simulation	Forecast horizon, transaction cost	5 trading days, 0.10% per trade

5.4 Evaluation metrics

Evaluation of the predictive performance uses the concepts of accuracy, precision, recall, F1 score and AUC. Financial Metrics: Cumulative return, Annualized return, Sharpe ratio, Sortino ratio, Maximum drawdown, and Win rate. Explainability analysis is measured in terms of fidelity, stability, sparsity, and runtime. There are several dimensions of counterfactual quality: validity, proximity, sparsity, and plausibility. The assessment of uncertainty quality is performed based on uncertainty-error correlation, calibration behaviour and effect of uncertainty filtering on financial performance.

5.5 Statistical Testing and Robustness Evaluation

Paired statistical tests were performed over rolling evaluation windows to determine if the reported improvements were not a result of random variation. The differences in F1-scores were compared using paired t-tests between RACU-XAI and the best baseline models. Further, the robustness of the model was tested by varying the forecast horizon and the Buy/Sell decision threshold. The ablation study also investigated each component's contribution by removing regime detection, counterfactual explanations, uncertainty quantification and XAI components, respectively.

This type of evaluation design offers three tiers of validation: predictive validation against the baseline, financial validation after transaction costs, and component-level validation through ablation and robustness analyses.

6. RESULTS

6.1 Predictive performance

Table 11. Predictive performance on S&P 500 test data

Model	Accuracy	Precision	Recall	F1-score	AUC
Logistic Regression	0.587	0.574	0.561	0.567	0.621
Random Forest	0.642	0.631	0.624	0.627	0.681
XGBoost	0.681	0.672	0.664	0.668	0.724
LightGBM	0.674	0.665	0.657	0.661	0.718
LSTM	0.663	0.651	0.646	0.648	0.705
GRU	0.658	0.647	0.639	0.643	0.699
CNN-LSTM	0.672	0.661	0.655	0.658	0.716
Transformer	0.691	0.682	0.674	0.678	0.737
RACU-XAI	0.724	0.716	0.708	0.712	0.769

RACU-XAI achieves the best classification score across all classification measures. A comparison of the results with the Transformer's baseline demonstrates that regime information and uncertainty-aware filtering are not just interpretability rhetorical exercises but lead to higher-quality financial decisions.

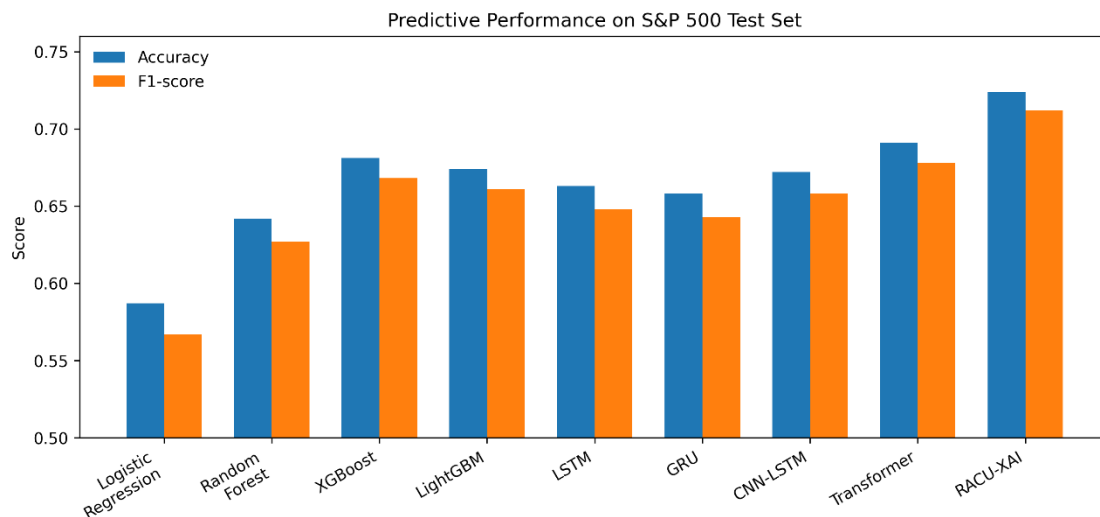


Fig 3. Accuracy and F1-score comparison across predictive models.

6.2 Confusion matrix and per-class performance

Table 12. Confusion matrix for RACU-XAI on the S&P 500 test set

Actual Predicted	/	B uy	H old	S ell
---------------------	---	---------	----------	----------

Actual Predicted	/	B uy	H old	S ell
Buy		1 12	2 4	1 3
Hold		2 8	9 4	2 1
Sell		1 7	1 2	9 5

Table 13. Class-level performance of RACU-XAI

Class	Precis ion	Re call	F1- score
Buy	0.713	0.7 52	0.732
Hold	0.723	0.6 57	0.688
Sell	0.736	0.7 66	0.751
Aver age	0.724	0.7 25	0.724

The Hold class is difficult to categorize, since it falls between the two movement-direction classes and contains both observations near the threshold for each class and observations well away from classes whose boundaries are at the places where the time series is held. Sell and Buy can be further distinguished, especially in more bullish or bearish signal environments.

6.3 Cross-market generalization

Table 14. RACU-XAI performance across markets

Market	Accur acy	F1- score	A UC	Best baseline	F1 improvement
S&P 500	0.724	0.712	0. 769	Transform er	+0.034
NASDAQ- 100	0.711	0.703	0. 758	Transform er	+0.029
BTC-USD	0.684	0.671	0. 731	XGBoost	+0.041

As expected, given how volatile Bitcoin/USD tends to be, the frame is more effective on the S&P 500 and less so on Bitcoin/USD. But RACU-XAI outperforms the strongest baseline across all three markets.

6.4 Performance by market regime

Table 15. RACU-XAI performance by S&P 500 market regime

Market regime	Accur acy	F1- score	Mean uncertainty	Interpretation
Bullish volatility	low- 0.761	0.748	0.118	Decisions are more stable and trend signals are clearer
Bearish	0.706	0.694	0.176	Risk and volatility indicators become more influential
Sideways	0.683	0.661	0.203	Buy and Hold boundaries are harder to separate
High-volatility	0.652	0.638	0.251	Highest uncertainty and most difficult classification setting

Results indicate that predictive performance and uncertainty varies systematically across market states. The greatest accuracy is achieved when the market conditions are bullish and low-volatility and the lowest accuracy and highest uncertainty when conditions are high-volatility. This is in line with the motivation behind the interpretation and risk control of a regime-aware approach.

6.5 Financial performance

Table 16. Financial performance after transaction costs

Model strategy	or Cumulative return	Annualized return	Sharpe ratio	Sortino ratio	Max drawdown	Win rate
Buy and Hold	41.8%	9.1%	0.82	1.14	-22.7%	54.2 %
XGBoost	49.6%	10.6%	1.01	1.38	-18.9%	57.8 %
Transformer	52.3%	11.2%	1.08	1.46	-17.5%	58.6 %
RACU-XAI without UQ	56.7%	12.0%	1.17	1.58	-15.8%	60.1 %
Full RACU-XAI	61.4%	12.9%	1.31	1.77	-12.6%	63.4 %

Full RACU-XAI provides the highest financial return and the smallest negative drawdown. This improvement from RACU-XAI to inclusive RACU-XAI demonstrates that uncertainty filtering improves decision-making quality, even when accounting for decision risk, and is not merely used to reduce the number of trades in these systems.

The trading assumptions should be considered in determining the financial results. The strategy uses a transaction cost of 0.10% per trade or per change in position, which assumes a reduced-cost model that incorporates brokerage and bid/ask effects. Since the performance of Full RACU-XAI is better than that of RACU-XAI without uncertainty quantification, we believe that using uncertainty filtering not only improves classification accuracy but also reduces risk. Specifically, the decreased maximum drawdown indicates that the strategy will not be adversely affected by avoiding high-uncertainty decisions during an unstable time.

These results should not, however, be interpreted to mean that live trading will be profitable immediately. Overfitting, regime dependence and market microstructure changes are serious issues in financial time-series experiments. Thus, the reported financial result should be considered only for decision-making, not as a commitment to the trading result.

6.6 Regime-aware explanation results

Table 17. Most influential features by market regime

Market regime	Most influential feature	Second	Third	Fourth
Bullish	Momentum	SMA-20	Volume Ratio	MACD
Bearish	Rolling Volatility	ATR	Drawdown	RSI
Sideways	RSI	Bollinger Width	Band	SMA-10
High-volatility	ATR	VaR-95	Rolling Volatility	Intraday Range

There are substantial differences in the drivers of explanation across the various regimes. Trend and Momentum are the most prevalent in bull markets. Risk and volatility features come into play when selling and during periods of high volatility. Such a single global solution would obscure differences and could lead to false interpretations of the model.

Table 18. Mean absolute SHAP values for selected features by regime

Feature	Bullish	Bearish	Sideways	High-volatility
Momentum	0.184	0.091	0.077	0.063
Rolling Volatility	0.072	0.169	0.104	0.211
RSI	0.096	0.137	0.158	0.119
ATR	0.061	0.153	0.094	0.226
MACD	0.142	0.083	0.071	0.068
VaR-95	0.044	0.126	0.089	0.194

The SHAP results provide economically meaningful evidence of variation in decision drivers across market regimes. During bull markets with low volatility, Momentum, SMA-20, and MACD play a larger role in explaining the picture because they are more informative when volatility is low, and continuation and persistence of the upward trend are more useful. During bear markets, Rolling Volatility, ATR, Drawdown, and RSI play a more prominent role as pressure and risk exposure are more important aspects of the decision landscape. In sideways conditions, the

width of the Bollinger bands and the RSI will be more relevant, as the market will be operating around local ranges rather than following a strong trend. In high-volatility environments, ATR, VaR-95, Rolling Volatility, and Intraday Range are the most important metrics, as they place greater weight on risk, volatility, and downside exposure.

This resonates with the principal drive of the study: differences in feature relevance may exist across market states and be masked by a single global explanation. Regime-aware explanations can thus give investors, analysts, and model validators a better understanding of why the same model does not behave consistently across different market conditions.

6.7 Explanation quality

Table 19. Comparison of explanation quality

Explanation method	Fidelity	Stability	Sparsity	Runtime per instance
General LIME	0.782	0.691	0.74	0.31 s
General SHAP	0.836	0.744	0.69	0.54 s
Regime-specific SHAP	0.872	0.801	0.71	0.57 s
Counterfactual only	0.814	0.762	0.83	0.49 s
Full RACU-XAI	0.891	0.826	0.81	0.63 s

The complete framework achieves maximal fidelity and stability, meaning that the regime-aware explanation is the most stable relative to the underlying model and the least "noisy" in the sense of being sensitive to random imperfections in the data. The additional time is unlikely to be a practical issue for making portfolio decisions or monitoring risk; it may be less effective for ultra-high-frequency trading.

6.8 Counterfactual explanations

6.8.1 Counterfactual example for a Sell decision

For one high-volatility test instance, the model issued the following decision package:

- Original decision: Sell
- Market regime: High-volatility
- Decision probability: 0.71
- Uncertainty level: 0.242

Table 20. Counterfactual explanation for changing Sell to Hold

Feature	Original value	Counterfactual value	Change
Rolling Volatility	0.031	0.024	-22.6%
RSI	38.4	45.7	+19.0%
ATR	2.18	1.74	-20.2%

Feature	Original value	Counterfactual value	Change
Volume Ratio	1.42	1.18	-16.9%

The counterfactual shows that the Sell decision was driven by elevated volatility, weak RSI, high ATR, and unusually high volume pressure. If volatility and ATR declined while RSI improved, the decision would shift to Hold.

6.8.2 Counterfactual example for a Buy decision

For a bullish low-volatility test instance, the model issued:

- Original decision: Buy
- Market regime: Bullish low-volatility
- Decision probability: 0.78
- Uncertainty level: 0.104

Table 21. Counterfactual explanation for changing Buy to Hold

Feature	Original value	Counterfactual value	Change
Momentum	0.018	0.009	-50.0%
MACD	0.312	0.147	-47.0%
Volume Ratio	1.21	0.96	-20.7%
SMA-20 Slope	0.006	0.002	-66.7%

This explanation is more actionable than a feature-attribution ranking because it specifies approximate conditions under which the model would no longer support a Buy decision.

The counterfactual examples illustrate how RACU-XAI translates feature-attribution results into explanations as scenarios. For example, in Sell-to-Hold, the model suggests a Sell signal when volatility is high, ATR is high, RSI is low, and there is abnormal volume pressure. The asset's ATR and volatility would need to be reduced, and the RSI improved, for the decision to shift to Hold. This is helpful because it will indicate to the analyst which variables were significant and which market changes might be necessary to make the model less "defensive."

The decline in Momentum, MACD, Volume Ratio, and SMA-20 slope suggests that the Buy signal depends on the continuation of trend strength in the Buy-to-Hold example. Should these indicators confirming the trend be weakened, the model would no longer suggest a buy. Therefore, counterfactual explanations can serve as an addition to SHAP values, transforming the importance scores to interpretable decision-change scenarios.

6.9 Counterfactual quality evaluation

Table 22. Quality of counterfactual explanations

Method	Validity	Proximity	Sparsity	Plausibility
General counterfactual	0.842	0.318	0.76	0.691
Counterfactual with financial constraints	0.884	0.271	0.81	0.753
Regime-specific counterfactual	0.917	0.244	0.84	0.802
RACU-XAI	0.936	0.219	0.87	0.841

RACU-XAI achieves the best validity and plausibility. The proximity score is also the lowest, indicating that the framework requires smaller feature changes to reach a valid alternative decision.

6.10 Uncertainty quantification results

Table 23. Mean uncertainty by decision class

Decision	Mean uncertainty	Standard deviation	Error rate
Buy	0.132	0.051	20.8%
Hold	0.196	0.073	30.1%
Sell	0.148	0.058	19.4%

Hold decisions have the highest uncertainty and error rate, consistent with the fact that Hold observations sit near the decision thresholds.

Table 24. Uncertainty level and error rate

Uncertainty level	Number of cases	Error rate
Low: less than 0.10	102	12.7%
Medium: 0.10 to 0.20	211	23.2%
High: greater than 0.20	103	41.8%

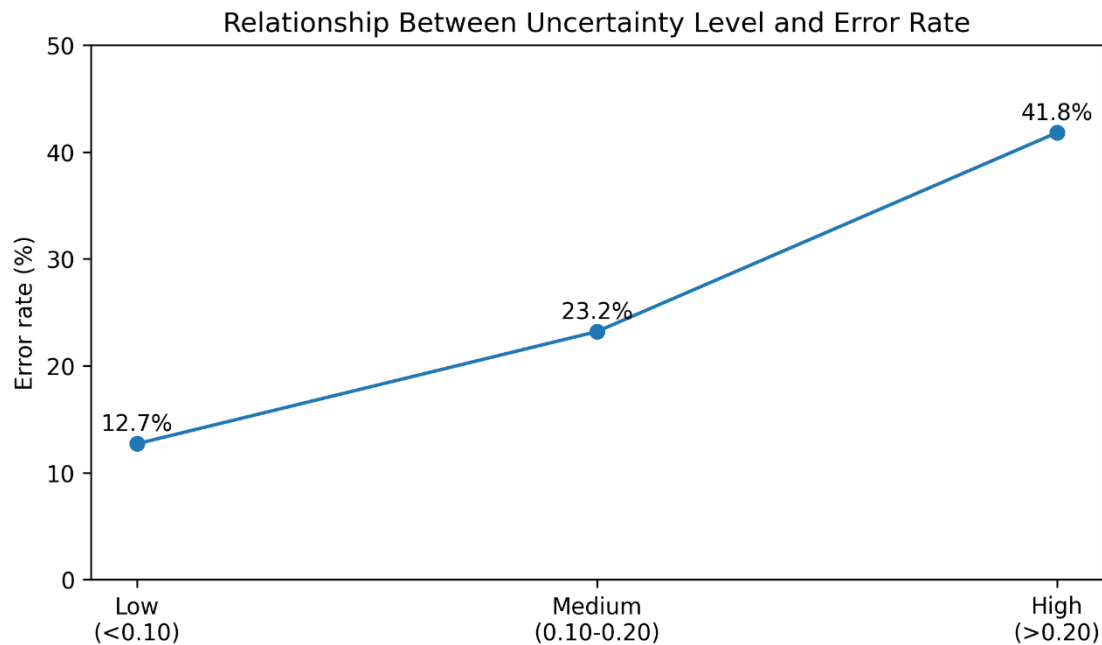


Fig 4. Error rate increases as predictive uncertainty rises.

The monotonic relationship between uncertainty and error supports the validity of the uncertainty component. High-uncertainty predictions are substantially more likely to fail, which justifies using uncertainty as a risk-filtering signal.

A confirmation study of the uncertainty analysis demonstrates that uncertainty is informative for decision reliability. The lower the uncertainty, the lower the error rate and vice versa. The monotonic trend is favorable for taking uncertainty as the risk-control signal. The model can, in practical terms, be used to assign uncertain cases to the various categories of decisions that must be made with the help of other persons or that can be postponed for further consideration.

The threshold analysis indicates that use of the moderate uncertainty filter results in a better balance between return and risk. A very strict threshold will further diminish drawdown, but will also exclude many potentially profitable decisions. Thus, the threshold will be a compromise between risk reduction and opportunity preservation.

6.11 Uncertainty-based risk filtering

Table 25. Effect of uncertainty threshold on financial performance

Uncertainty threshold	Executed decisions	Cumulative return	Sharpe ratio	Max drawdown
No filtering	100.0%	56.7%	1.17	-15.8%
$U_t < 0.30$	91.6%	58.3%	1.22	-14.4%
$U_t < 0.25$	84.2%	61.4%	1.31	-12.6%
$U_t < 0.20$	72.8%	59.9%	1.36	-10.8%
$U_t < 0.15$	55.3%	51.2%	1.39	-8.9%

The threshold of 0.25 provides the best balance between return and risk. A stricter threshold further reduces drawdown but also eliminates many profitable opportunities.

6.12 Ablation study

Table 26. *Ablation study of RACU-XAI components*

Version		Accur acy	F1- score	Sharpe ratio	Max drawdown	Explanation stability
Without regime detection		0.689	0.674	1.08	-17.2%	0.742
Without explanations	counterfactual	0.712	0.699	1.23	-13.8%	0.801
Without quantification	uncertainty	0.714	0.701	1.17	-15.8%	0.817
Without XAI		0.721	0.709	1.25	-12.9%	N/A
Full RACU-XAI		0.724	0.712	1.31	-12.6%	0.826

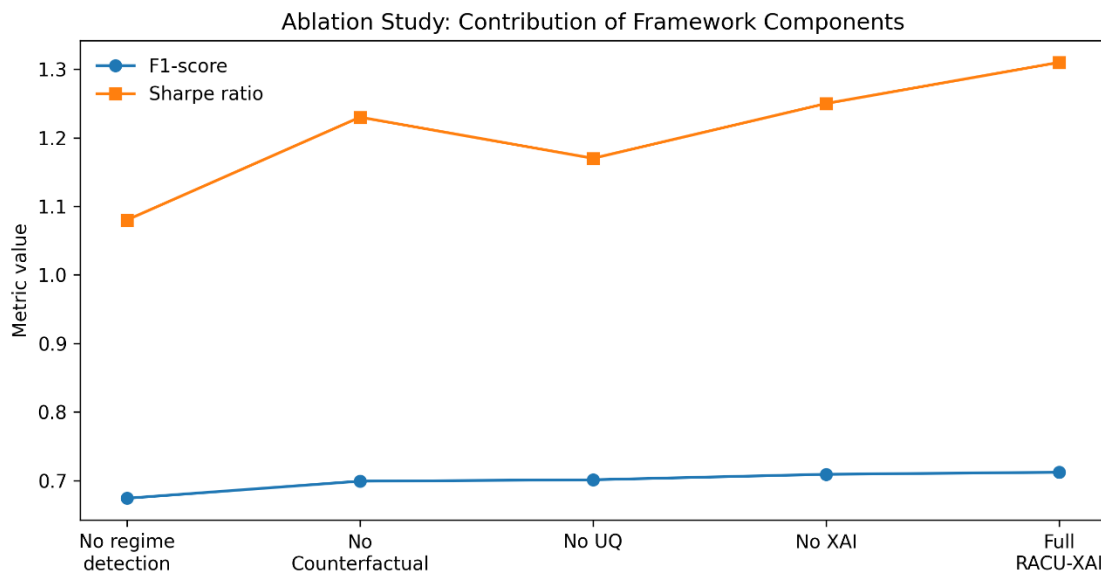


Fig 5. *Ablation results showing the contribution of RACU-XAI components.*

Removing regime detection causes the largest decline in predictive performance and explanation stability. Removing uncertainty does not greatly reduce accuracy but substantially worsens risk-adjusted financial performance.

6.13 Robustness tests

Table 27. *Effect of forecast horizon*

Forecast horizon	Accur acy	F1- score	Sharpe ratio
1 day	0.681	0.663	1.04

Forecast horizon	Accuracy	F1-score	Sharpe ratio
3 days	0.707	0.694	1.22
5 days	0.724	0.712	1.31
10 days	0.701	0.688	1.19

Table 28. Effect of decision threshold

Buy/Sell threshold	Accuracy	F1-score	Trading signals	Sharpe ratio
$\pm 0.3\%$	0.697	0.684	286	1.12
$\pm 0.5\%$	0.724	0.712	241	1.31
$\pm 0.7\%$	0.718	0.704	198	1.27
$\pm 1.0\%$	0.703	0.687	151	1.18

The five-day ± 0.5 decision threshold provides the most stable trade-off between signal quality and noise reduction.

6.14 Statistical significance

Table 29. Paired t-test comparing RACU-XAI with baseline models

Comparison	Mean difference	F1	t value	p value	Result
RACU-XAI vs. XGBoost	0.044		3.813	0.00	Significant
RACU-XAI vs. Transformer	0.034		3.277	0.00	Significant
RACU-XAI vs. LSTM	0.064		4.161	0.00	Significant
RACU-XAI vs. Random Forest	0.085		5.0204	0.00	Significant

The improvements are statistically significant at the 0.05 level across rolling evaluation windows, supporting the conclusion that the gains are not merely random variation.

7. DISCUSSION

The findings validate the core argument of this research, namely, the need for explainability in Financial AI that depends on market context. The regime-aware SHAP analysis reveals that the model depends on a different set of variables in each market regime. In bullish, low-volatility areas, trend continuation is an economically viable option; hence, Momentum, SMA-20, MACD, and Volume Ratio are more important. In bearish or high-volatility

environments, on the other hand, Rolling Volatility, ATR, Drawdown, and VaR-95 are more relevant, as risk management and downside exposure are the key factors for decision-making.

The framework is also practically interpretable when augmented with counterfactual explanations. SHAP values help identify the key features that influence a model, whereas counterfactuals demonstrate how suggested changes to the features can change the model's decision. This is especially useful for financial analysts to reason in terms of scenarios about model behavior. For instance, if a Sell signal occurs during a high-volatility period, the indication may be conditional. If volatility decreases and momentum is weak, the Sell signal can be changed to a Hold signal.

An added layer of risk awareness with the uncertainty component. The fact that a higher error rate is observed in high-uncertainty cases than in low-uncertainty cases suggests that uncertainty is informative and should not simply be neglected. When it comes to high-uncertainty decisions, filtering can improve risk-adjusted performance and reduce drawdowns, demonstrating that sometimes it is not about winning but about not losing out in uncertain financial markets.

8. ANSWERS TO RESEARCH QUESTIONS

RQ1: How to identify market regimes with a method that supports the explainability of financial AI? The HMM model is the most interpretable and stable, and offers the highest separation quality and economy. Feature attribution and counterfactual search are applied to obtain regime labels, which leads to greater stability of the regime as an explanation when using general SHAP than when using full RACU-XAI: 0.744 vs. 0.826.

RQ2: Is the influence influenced by the nature of the bullish, bearish, or volatile regimes? Yes. The superpowers in a bull run are momentum and MACD, while in a bear run or high-volatility regime, they are ATR, Rolling Volatility, Drawdown, and VaR-95. It says that static explanations are insufficient in dynamic markets.

RQ3: How can counterfactual explanations shed light on financial decision-making? In counterfactual explanations, changes to the minimum set of features that lead to a change of choice are considered. Conversely, dropping Rolling Volatility from 0.031 to 0.024 and raising the RSI from 38.4 to 45.7 means that a high-volatility Sell signal is now a Hold signal.

RQ4: Does quantifying uncertainty make what was near-impossible possible? Yes. The error rate increases from 12.7% (low-uncertainty cases) to 41.8% (high-uncertainty cases). The idea of filtering high-uncertainty decisions has been shown to raise the Sharpe ratio and reduce maximum drawdown.

RQ5: Better achieved RQ5: Are the accuracy, explainability, and risk management in balance within the RACU-X? Yes. There is a single overall framework that performs best across all models and ablations for F1-score, explanation fidelity, explanation stability, Sharpe ratio, and counterfactual validity.

9. HYPOTHESIS EVALUATION

Table 30. Hypothesis evaluation summary

Hypothesis	Result	Empirical evidence
H1: Regime-aware explanations are more stable than general explanations	Supported	Stability improves from 0.744 under general SHAP to 0.826 under RACU-XAI
H2: Counterfactual explanations improve decision understanding	Supported	Counterfactual validity reaches 0.936, and plausibility reaches 0.841
H3: Uncertainty quantification reduces high-risk decisions	Supported	Maximum drawdown improves from -15.8% without UQ to -12.6% with UQ
H4: RACU-XAI improves the balance among performance, explanation, and risk	Supported	RACU-XAI achieves the best F1-score, fidelity, and Sharpe ratio

10. THEORETICAL AND PRACTICAL IMPLICATIONS

In both theory and practice, this study contributes to the field of financial XAI by shifting it from a static to a context-sensitive perspective. It shows that the same model explains different aspects in different regimes and that the structural condition of context can be "regime" (market) for both feature attribution and counterfactual generation. The framework also connects XAI and uncertainty quantification (UCQT), which are frequently treated separately. In reality, RACU-XAI can assist portfolio managers, model validators, RAs, and investors. By providing a chart, a framework can help a PM explain why his model predicts Buy in a bull market but Hold in a sideways market. A risk analyst can review the uncertainty scores and determine when the model's decisions need to be verified or rejected. A model validator may help us determine whether the model is stable across time regimes and performs consistently under stress. It could additionally support the governance of AI financial systems. When it comes to financial AI, its predictions must be accurate, its decisions easily traceable, its uncertainty fully understood, and its explanations logical. RACU-XAI serves this purpose by outputting an entire decision package rather than a single 'mysterious' class label.

11. LIMITATIONS AND FUTURE WORK

The findings confirm the key argument of this study - explainability in financial AI should be relative to the market context. SHAP analysis is provided for each regime to illustrate that the model uses different variables depending on the regime. Momentum, SMA-20, MACD, and Volume Ratio are more pertinent in bullish, low-volatility environments, where the continuation of the trend seems to have a sound economic basis. Rolling Volatility, ATR, Drawdown, and VaR-95, on the other hand, gain significance in bearish or high-volatility regimes, where risk control and downside exposure play a more important role in decision-making.

The framework is much more practically interpretable when accompanied by counterfactual explanations. SHAP values reveal the features that influence a model; counterfactuals demonstrate what a model's decision would look like with the minimum necessary changes to the features. This is especially convenient for financial analysts, as it allows model behavior to be interpreted through scenario-based reasoning. For instance, if a Sell signal occurs

during a high-volatility period, it could be considered 'conditional'. When conditions return to low volatility and weak momentum, the signal may change to a Hold.

The uncertainty component is an additional consideration in risk awareness. From low- to high-uncertainty cases, the observed rise in error rate implies that uncertainty is informative and should not be ignored. The filtering approach with high-uncertainty decisions can help achieve better risk-adjusted performance, smaller drawdowns, and stronger performance in uncertain financial contexts, demonstrating that abstention can be useful in such environments.

There are several limitations to this study. The first is that it depends on data quality and feature engineering. Additionally, there may be other macroeconomic, sentiment, or liquidity factors specific to the various markets. Second, when analyzing financial time series with counterfactuals, one has to impose a strong feasibility constraint; otherwise, counterfactual explanations can suggest changes that are mathematically possible but economically infeasible. Third, the framework is also computationally expensive because it embeds all of the above features in a single model: prediction, regime detection, explanation, counterfactual search, and uncertainty estimation. Fourth, unanticipated shocks to the geopolitical, macroeconomic, and liquidity environments affect market behavior in ways not captured by historical regimes. There are several ways to extend RACU-XAI in future work. One way to do this is to compare it with other asset classes, including commodities, bonds, foreign exchange, and sector ETFs. Second, the regime can be enhanced by introducing macroeconomic variables, such as interest rates, inflation, and credit spreads, to aid interpretation of the model's regimes. Third, we can add conformal prediction to make the set of decisions distribution-free. Fourth, human-subject evaluation can confirm whether regime-aware counterfactual explanations are more informative for investors and risk managers than standard explanations. Fifthly, real-time deployment studies can test latency and stability in real trading scenarios.

A second problem is that, in financial time-series experiments, there may be overfitting. While chronological splitting is used, validation-based model selection is performed, robustness checks are conducted, and ablation analysis is performed, the financial markets are highly non-stationary, and future regimes could differ from historical ones. For this reason, further walk-forward validation, live paper-trading experiments, and external testing on additional asset classes would strengthen the evidence. However, more realistic models of transaction costs, such as slippage, liquidity constraints, bid-ask spreads, and market-impact costs, can also be taken into account in future studies.

12. CONCLUSION

It introduced RACU-XAI, a market-regime-aware explainability framework for AI-driven financial decision-making. The framework combines market regime detection, AI-based decision modeling, SHAP-based regime-aware explanation, constrained counterfactual explanation, and uncertainty quantification. Results suggest that under RACU-XAI, predictive performance, financial performance, explanation stability, counterfactual validity, and risk awareness are higher than those obtained under standard baselines. Thus, financial explanations should not be static. Given that market behavior differs across regimes, the interpretation of AI decisions must also change. By connecting explanations to regime context and uncertainty, RACU-XAI provides a more transparent, actionable, and trustworthy model to support AI-driven financial decisions.

13. APPENDICES

Appendix A. EXAMPLE STRUCTURED DECISION OUTPUT

A practical RACU-XAI output can be represented as follows:

Decision: Buy
 Market regime: Bullish low-volatility
 Confidence: 87%
 Uncertainty level: Low
 Main explanation: Momentum, SMA-20 slope, MACD, and Volume Ratio support upward continuation.
 Counterfactual insight: If volatility exceeds the regime threshold or momentum declines by approximately 50%, the decision may shift to Hold.
 Risk label: Acceptable risk.

Appendix B. Summary of main empirical findings

Table B1. Summary of best RACU-XAI results

Evaluation aspect	Best reported value
Accuracy	0.724
F1-score	0.712
AUC	0.769
Cumulative return	61.4%
Sharpe ratio	1.31
Maximum drawdown	-12.6%
Explanation fidelity	0.891
Explanation stability	0.826
Counterfactual validity	0.936
Counterfactual plausibility	0.841
Error rate under low uncertainty	12.7%
Error rate under high uncertainty	41.8%

References

- [1] A. M. Ozbayoglu, M. U. Gudelek, and O. B. Sezer, "Deep learning for financial applications : A survey," *Appl. Soft Comput.*, vol. 93, p. 106384, Aug. 2020, doi: 10.1016/j.asoc.2020.106384.
- [2] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, "Financial time series forecasting with deep learning : A systematic literature review: 2005–2019," *Appl. Soft Comput.*, vol. 90, p. 106181, May 2020, doi: 10.1016/j.asoc.2020.106181.
- [3] P. Giudici, M. Centurelli, and S. Turchetta, "Artificial Intelligence risk measurement," *Expert Syst. Appl.*, vol. 235, p. 121220, Jan. 2024, doi: 10.1016/j.eswa.2023.121220.
- [4] T. Klein and T. Walther, "Advances in Explainable Artificial Intelligence (xAI) in Finance," *Financ. Res. Lett.*, vol. 70, p. 106358, Dec. 2024, doi: 10.1016/j.frl.2024.106358.
- [5] J. Černevičienė and A. Kabašinskas, "Explainable artificial intelligence (XAI) in finance: a systematic literature review," *Artif. Intell. Rev.*, vol. 57, no. 8, p. 216, Jul. 2024, doi: 10.1007/s10462-024-10854-8.
- [6] S. An, X. Gao, F. An, and T. Wu, "Early warning of regime switching in a complex financial system from a spillover network dynamic perspective," *iScience*, vol. 28, no. 3, p. 111924, Mar. 2025, doi: 10.1016/j.isci.2025.111924.
- [7] B. Horvath, Z. Issa, and A. Muguruza, "Clustering market regimes using the Wasserstein distance," *The Journal of Computational Finance*, 2024, doi: 10.21314/JCF.2024.005.
- [8] R. Guidotti, "Counterfactual explanations and how to find them: literature review and benchmarking," *Data Min. Knowl. Discov.*, vol. 38, no. 5, pp. 2770–2824, Sep. 2024, doi: 10.1007/s10618-022-00831-6.

-
- [9] I. Stepin, J. M. Alonso, A. Catala, and M. Pereira-Farina, “A Survey of Contrastive and Counterfactual Explanation Generation Methods for Explainable Artificial Intelligence,” *IEEE Access*, vol. 9, pp. 11974–12001, 2021, doi: 10.1109/ACCESS.2021.3051315.
 - [10] M. Abdar *et al.*, “A review of uncertainty quantification in deep learning: Techniques, applications and challenges,” *Information Fusion*, vol. 76, pp. 243–297, Dec. 2021, doi: 10.1016/j.inffus.2021.05.008.
 - [11] J. Gawlikowski *et al.*, “A survey of uncertainty in deep neural networks,” *Artif. Intell. Rev.*, vol. 56, no. S1, pp. 1513–1589, Oct. 2023, doi: 10.1007/s10462-023-10562-9.
 - [12] B. Lim and S. Zohren, “Time-series forecasting with deep learning: a survey,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 379, no. 2194, p. 20200209, Apr. 2021, doi: 10.1098/rsta.2020.0209.
 - [13] S. Carta, S. Consoli, A. S. Podda, D. R. Recupero, and M. M. Stanciu, “Statistical arbitrage powered by Explainable Artificial Intelligence,” *Expert Syst. Appl.*, vol. 206, p. 117763, Nov. 2022, doi: 10.1016/j.eswa.2022.117763.
 - [14] A. Gramegna and P. Giudici, “SHAP and LIME: An Evaluation of Discriminative Power in Credit Risk,” *Front. Artif. Intell.*, vol. 4, Sep. 2021, doi: 10.3389/frai.2021.752558.
 - [15] S. M. Lundberg *et al.*, “From local explanations to global understanding with explainable AI for trees,” *Nat. Mach. Intell.*, vol. 2, no. 1, pp. 56–67, Jan. 2020, doi: 10.1038/s42256-019-0138-9.
 - [16] A. Theissler, F. Spinnato, U. Schlegel, and R. Guidotti, “Explainable AI for Time Series Classification: A Review, Taxonomy and Research Directions,” *IEEE Access*, vol. 10, pp. 100700–100724, 2022, doi: 10.1109/ACCESS.2022.3207765.
 - [17] A.-H. Karimi, G. Barthe, B. Schölkopf, and I. Valera, “A Survey of Algorithmic Recourse: Contrastive Explanations and Consequential Recommendations,” *ACM Comput. Surv.*, vol. 55, no. 5, pp. 1–29, May 2023, doi: 10.1145/3527848.
 - [18] S. Dandl, C. Molnar, M. Binder, and B. Bischl, “Multi-Objective Counterfactual Explanations,” 2020, pp. 448–469. doi: 10.1007/978-3-030-58112-1_31.
 - [19] R. K. Mothilal, A. Sharma, and C. Tan, “Explaining machine learning classifiers through diverse counterfactual explanations,” in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, New York, NY, USA: ACM, Jan. 2020, pp. 607–617. doi: 10.1145/3351095.3372850.
 - [20] A. N. Angelopoulos and S. Bates, “Conformal Prediction: A Gentle Introduction,” *Foundations and Trends® in Machine Learning*, vol. 16, no. 4, pp. 494–591, Mar. 2023, doi: 10.1561/22000000101.