algorithms

MDPI

Article

Blockchain-Native Asset Direction Prediction: A Confidence-Threshold Approach to Decentralized Financial Analytics Using Multi-Scale Feature Integration

Oleksandr Kuznetsov [1,2,*], Dmytro Prokopovych-Tkachenko [3,*], Maksym Bilan [3], Borys Khruskov [3] and Oleksandr Cherkaskyi [4]

[1] Department of Theoretical and Applied Sciences, eCampus University, Via Isimbardi 10, 22060 Novedrate, Italy
[2] Department of Intelligent Software Systems and Technologies, School of Computer Science and Artificial Intelligence, V.N. Karazin Kharkiv National University, 61022 Kharkiv, Ukraine
[3] Department of Cybersecurity and Information Technologies, University of Customs and Finance, 49000 Dnipro, Ukraine
[4] Department of Cybersecurity and Information Technologies, National Technical University "Dnipro Polytechnic", 49005 Dnipro, Ukraine
* Correspondence: oleksandr.kuznetsov@uniecampus.it (O.K.); d.prokopovych-tkachenko@duikt.edu.ua (D.P.-T.)

Abstract

Blockchain-based financial ecosystems generate unprecedented volumes of multi-temporal data streams requiring sophisticated analytical frameworks that leverage both on-chain transaction patterns and off-chain market microstructure dynamics. This study presents an empirical evaluation of a two-class confidence-threshold framework for cryptocurrency direction prediction, systematically integrating macro momentum indicators with microstructure dynamics through unified feature engineering. Building on established selective classification principles, the framework separates directional prediction from execution decisions through confidence-based thresholds, enabling explicit optimization of precision–recall trade-offs for decentralized financial applications. Unlike traditional three-class approaches that simultaneously learn direction and execution timing, our framework uses post-hoc confidence thresholds to separate these decisions. This enables systematic optimization of the accuracy-coverage trade-off for blockchain-integrated trading systems. We conduct comprehensive experiments across 11 major cryptocurrency pairs representing diverse blockchain protocols, evaluating prediction horizons from 10 to 600 min, deadband thresholds from 2 to 20 basis points, and confidence levels of 0.6 and 0.8. The experimental design employs rigorous temporal validation with symbol-wise splitting to prevent data leakage while maintaining realistic conditions for blockchain-integrated trading systems. High confidence regimes achieve peak profits of 167.64 basis points per trade with directional accuracies of 82–95% on executed trades, suggesting potential applicability for automated decentralized finance (DeFi) protocols and smart contract-based trading strategies on similar liquid cryptocurrency pairs. The systematic parameter optimization reveals fundamental trade-offs between trading frequency and signal quality in blockchain financial ecosystems, with high confidence strategies reducing median coverage while substantially improving per-trade profitability suitable for gas-optimized on-chain execution.

Keywords: cryptocurrency prediction; machine learning; market microstructure; direction prediction; confidence thresholds; multi-scale features; temporal integration; trading strategies; neural networks

check for updates

Academic Editors: Sergey Y. Yurish and Ming-Feng Ge

Received: 24 September 2025
Revised: 23 November 2025
Accepted: 26 November 2025
Published: 29 November 2025

Citation: Kuznetsov, O.; Prokopovych-Tkachenko, D.; Bilan, M.; Khruskov, B.; Cherkaskyi, O. Blockchain-Native Asset Direction Prediction: A Confidence-Threshold Approach to Decentralized Financial Analytics Using Multi-Scale Feature Integration. *Algorithms* **2025**, *18*, 758. https://doi.org/10.3390/a18120758

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

1. Introduction

The emergence of blockchain-based financial ecosystems has created new paradigms for decentralized data analytics and automated decision-making systems. Cryptocurrency markets, as the primary application domain of blockchain technology, generate rich multi-dimensional datasets that combine on-chain transaction flows with traditional market microstructure signals. This unique data environment presents both opportunities and challenges for AI-driven analytics, particularly in the context of developing automated trading systems that can operate within decentralized finance (DeFi) protocols and smart contract environments.

Cryptocurrency markets exhibit unique characteristics that challenge traditional financial prediction frameworks [1,2]. High volatility, 24/7 trading cycles, and diverse investor participation create complex market dynamics requiring specialized analytical approaches. The rapid growth of cryptocurrency trading volumes, exceeding USD 1 trillion daily across major exchanges, demands sophisticated prediction systems capable of generating consistent profits under realistic transaction costs [3,4].

Traditional cryptocurrency prediction research focuses primarily on single-timeframe analysis, typically employing daily price data or minute-level technical indicators in isolation [5]. This approach overlooks the fundamental interaction between macroeconomic trends and microstructure dynamics that characterizes modern cryptocurrency markets [6,7]. Daily momentum patterns often manifest through intraday order flow changes, while microstructure signals gain predictive power when aligned with broader market trends.

The integration of multiple temporal scales presents both opportunities and challenges for cryptocurrency direction prediction [8]. Macro features derived from daily OHLCV data across multiple assets provide market-wide context and fundamental momentum indicators. Microstructure features extracted from minute-frequency order book snapshots capture real-time market sentiment and liquidity conditions. The temporal bridge between these domains occurs at intermediate horizons where daily directional bias influences minute-level market-making activities.

Existing prediction frameworks typically employ three-class classification schemes (Up, Down, No-trade) where models simultaneously learn directional prediction and execution timing decisions [9,10]. This approach confounds signal extraction with risk management, potentially degrading both prediction accuracy and trading performance. The mixed representation of unclear signals and inappropriate timing within no-trade samples may compromise model learning effectiveness.

Our confidence-threshold approach builds upon established foundations in selective classification [Chow, 1970 [11]; Herbei & Wegkamp, 2006 [12]], where classifiers may abstain from predictions when confidence is insufficient. While selective classification theory has been extensively developed for cost-sensitive learning, its application to cryptocurrency markets with integrated macro–microstructure features and systematic threshold optimization across temporal scales represents an empirical contribution addressing domain-specific challenges in decentralized finance. Our framework draws conceptual parallels with abstention learning in machine learning (Cortes et al., 2016 [13]) and uncertainty-based trading in quantitative finance (Marcos López de Prado, 2018 [14]). Abstention learning addresses scenarios where classifiers may decline predictions when uncertainty is high, optimizing coverage-accuracy trade-offs through rejection thresholds. Similarly, uncertainty-based portfolio construction incorporates prediction confidence into position sizing and execution decisions. While these frameworks establish theoretical foundations for confidence-aware decision-making, their application to cryptocurrency direction prediction with integrated

macro–microstructure features and systematic threshold optimization across multiple temporal scales represents a novel contribution within the decentralized finance domain.

Confidence-threshold mechanisms offer alternative approaches to execution control, enabling separation of directional prediction from trading decisions [15]. By training binary classifiers for pure directional signals and employing separate confidence-based execution rules, systems can optimize the precision–recall trade-off systematically. This decoupling allows explicit control over trading frequency versus signal quality, addressing fundamental challenges in cryptocurrency prediction system design.

Neural network architectures demonstrate consistent effectiveness for cryptocurrency prediction tasks, with Long Short-Term Memory (LSTM) networks achieving directional accuracies of 60–85% across multiple studies [9,10]. However, most research evaluates prediction accuracy rather than economic performance, limiting practical applicability assessment. The gap between statistical performance and trading profitability requires frameworks that explicitly optimize economic metrics under realistic operational constraints.

Feature engineering for cryptocurrency prediction typically emphasizes either technical indicators derived from price data [9] or market microstructure metrics from order book analysis [10]. Limited research investigates systematic integration of macro and microstructure signals across different temporal scales. The potential for cross-temporal feature interactions remains underexplored, particularly regarding optimal prediction horizons and signal quality thresholds.

Parameter optimization in cryptocurrency prediction systems often focuses on individual components rather than systematic exploration of joint parameter spaces. Horizon selection [16,17], signal quality requirements, and execution thresholds [15] interact in complex ways that may not be captured through independent optimization. Comprehensive parameter space analysis becomes essential for identifying optimal trading strategy configurations under different risk–return preferences.

The present research addresses these limitations through a two-class framework that integrates macro and microstructure features across multiple temporal scales. Our approach separates directional prediction from execution decisions using confidence-based thresholds, enabling systematic optimization of the precision–recall trade-off. We conduct comprehensive experiments across 11 major cryptocurrency pairs, exploring prediction horizons from 10 to 600 min, deadband thresholds from 2 to 20 basis points, and confidence levels of 0.6 and 0.8.

The research contributes to the cryptocurrency prediction literature through three primary advances. First, we develop a two-class binary classification framework that decouples directional prediction from execution timing decisions. Second, we implement systematic integration of macro momentum signals with microstructure dynamics through unified feature engineering. Third, we conduct comprehensive parameter space optimization across multiple dimensions to identify optimal trading strategy configurations.

Our experimental design employs rigorous temporal validation with symbol-wise splitting to prevent data leakage while maintaining realistic trading conditions. All the performance evaluation incorporates transaction costs and focuses on economic metrics relevant to practical trading applications. The results demonstrate significant improvements over baseline approaches, achieving peak profits of 167.64 basis points per trade, with directional accuracies of 75–95% on executed trades.

The remainder of this paper proceeds as follows. Section 2 reviews the relevant literature on cryptocurrency prediction methods and performance benchmarks. Section 3 describes the two-class framework architecture and multi-scale feature integration methodology. Section 4 details the dataset characteristics and preprocessing procedures. Section 5 presents the experimental design and validation framework. Section 6 reports comprehen-

sive results across both confidence regimes. Section 7 discusses economic interpretation, benchmark comparisons, and practical implementation considerations. Section 8 concludes with limitations and future research directions.

2. Literature Review

Cryptocurrency prediction research has evolved rapidly alongside market maturation, encompassing diverse methodological approaches from traditional time series analysis to advanced deep learning architectures. This review examines recent developments in cryptocurrency direction prediction, with particular emphasis on feature integration strategies, neural network architectures, and performance evaluation frameworks relevant to our two-class approach.

2.1. Neural Network Architectures for Cryptocurrency Prediction

Deep learning methods dominate contemporary cryptocurrency prediction research, with Long Short-Term Memory (LSTM) networks serving as the foundational architecture across multiple studies. Zhang et al. (2024) [18] conducted a comprehensive survey of deep learning applications in cryptocurrency markets, finding that LSTM models consistently achieve 83–84% average accuracy for Bitcoin and Ethereum prediction tasks. Their analysis reveals that ensemble methods combining multiple weak classifiers often outperform individual models, achieving explained variance scores of 0.97 and mean percentage errors around 0.06.

Attention mechanisms represent a significant advancement in cryptocurrency prediction architectures. Shang et al. (2024) [19] propose an attention-based CNN-BiGRU model for Ethereum price prediction, integrating blockchain information and external factors from 2017–2021 data. Their two-stage approach combines improved CNN for feature extraction with bidirectional GRU and attention mechanisms, achieving RMSE of 151.6 and MAE of 91.2, substantially outperforming traditional CNN-GRU (RMSE: 1067.1) and BIGRU (RMSE: 1065.7) baselines.

Graph neural networks introduce network-based perspectives to cryptocurrency prediction. Zhong et al. (2023) [20] develop LSTM-ReGAT, combining LSTM with Relationwise Graph Attention Networks for cryptocurrency price trend prediction. Their approach constructs a cryptocurrency network based on shared features including technological foundation, industry classification, and investor co-attention patterns. Testing on 645 cryptocurrencies over 995 days (March 2020–December 2022), they achieve AUC of 0.6615 and accuracy of 62.97%, representing modest but consistent improvements over LSTM baselines (AUC: 0.6546, accuracy: 62.27%).

2.2. Multi-Scale and Multi-Target Learning Approaches

Multi-target learning emerges as a promising direction for cryptocurrency prediction, leveraging correlations across multiple assets. Pellicani et al. (2025) [21] introduce CARROT, employing temporal clustering with Dynamic Time Warping to group correlated cryptocurrencies before training multi-target LSTM models for each cluster. Their approach processes 17 cryptocurrencies from January 2020 to December 2021, achieving an average 10% improvement in macro F1-score over single-target LSTMs, with the best performance showing 19% improvement using 6-month training intervals.

High-frequency prediction presents unique challenges requiring specialized architectures. Peng et al. (2024) [22] propose ACLMC (Attention-based CNN-LSTM for Multiple Cryptocurrencies) combined with novel triple trend labeling using local minimum series. Their approach integrates macro and microstructure features across multiple frequencies

and currencies, achieving significant reduction in transaction numbers (approximately 90% compared to traditional methods) while maintaining profitable performance.

2.3. Feature Engineering and Selection Methods

Feature selection methodology significantly impacts cryptocurrency prediction performance. El Youssefi et al. (2025) [23] conduct systematic investigation of feature selection methods applied to 130+ technical indicators for cryptocurrency price forecasting. Using mutual information (MI), recursive feature elimination (RFE), and recursive feature importance (RFI) methods with SVR, Huber, and KNN regressors, they achieve 80–85% feature reduction while maintaining or enhancing performance. Their results show peak R^2 values of 0.45–0.7 across BTC, ETH, and BNB pairs, with momentum and volatility indicators proving most important across timeframes.

Curvature-based approaches offer alternative feature engineering strategies. Zhang et al. (2024) [24] introduce generalized visible curvature indicator (CCPIq) for cryptocurrency bubble identification and price trend prediction. Their method captures geometric properties of log-price trajectories, quantifying interactions between trend, acceleration, and volatility. Integration with LightGBM achieves classification accuracy improvements and trading performance with Sharpe ratios up to 2.93 for Ethereum, significantly outperforming traditional bubble identification methods.

2.4. Probabilistic and Uncertainty Quantification Methods

Uncertainty quantification represents an emerging focus in cryptocurrency prediction research. Golnari et al. (2024) [25] introduce Probabilistic Gated Recurrent Units (P-GRU) for Bitcoin price prediction with uncertainty quantification. Their approach integrates probabilistic attributes into standard GRU architecture, facilitating generation of probability distributions for predicted values. Testing on one year of Bitcoin data at 5 min intervals, they achieve R^2-score of 0.99973 and MAPE of 0.00190, substantially outperforming traditional LSTM/GRU variants.

Potential field theory provides theoretical foundation for cryptocurrency market characterization. Anoop et al. (2025) [26] present a Bayesian machine learning framework using potential field theory and Gaussian processes to model cryptocurrency price movements as trajectories in dynamical systems governed by time-varying potential fields. Their analysis of Bitcoin crash periods (2017–2021) shows that attractors captured market trends, volatility, and correlations, with mean attractor features improving LSTM prediction performance by 25–28% in terms of MSE reduction.

2.5. Trading Strategy Integration and Performance Evaluation

The integration of prediction models with trading strategies receives increasing attention in the recent literature. Kang et al. (2025) [27] investigate technical indicator integration with deep learning-based price forecasting across 12 models for cryptocurrency trading strategies. Their best performing strategy combines TimesNet with Bollinger Bands in ETH markets, achieving returns of 3.19, maximum drawdown of −7.46%, and Sharpe ratio of 3.56. Technical indicator integration shows significant improvements at 4 h intervals, though no improvement occurs at shorter 30 min intervals.

Portfolio construction and trading strategy evaluation require sophisticated frameworks. Viéitez et al. (2024) [28] develop machine learning systems for Ethereum prediction and knowledge-based investment strategies, testing regression approaches with GRU and LSTM networks alongside SVM classification for trend prediction. Their evaluation across different time periods with real cryptocurrency market data shows profit factors ranging from 1.14 to 5.16, with limited influence from sentiment analysis integration.

2.6. Market Microstructure and Behavioral Factors

Market microstructure analysis reveals important patterns relevant to cryptocurrency prediction. Liu et al. (2025) [29] investigate liquidity commonality across 50 major cryptocurrencies from 2016 to 2023, finding strong positive liquidity commonality, with most coefficients approximating 1.0. Their results show liquidity commonality peaks mid-week (Wednesday–Thursday: 0.481–0.453) compared to weekends (0.246–0.322), with seasonal patterns persisting after controlling for volatility and returns.

Momentum effects demonstrate regime-dependent characteristics in cryptocurrency markets. Hsieh et al. (2025) [30] examine how market-state transitions shape momentum profitability across 2130 cryptocurrencies using weekly data from 2015 to 2023. Their findings show momentum profits concentrated exclusively in UP-UP transitions (11.9–15.5 basis points weekly), with no significant momentum in other regime combinations, suggesting asymmetric belief-updating patterns among cryptocurrency investors.

2.7. Grey Systems and Alternative Forecasting Methods

Alternative methodological approaches provide complementary perspectives to neural network dominance. Yang et al. (2025) [31] propose grey multivariate convolution models (GMCN(1,N)) for short-term cryptocurrency price forecasting, using grey correlation analysis to select core influencing variables. Testing on Bitcoin, Ethereum, and Litecoin data between 2022 and 2023, they achieve highly accurate predictions with MAPE values of 1.58% (BTC), 1.12% (ETH), and 2.53% (LTC), demonstrating the effectiveness of grey systems theory for cryptocurrency prediction.

2.8. Research Gaps and Methodological Challenges

Systematic mapping studies reveal persistent challenges in cryptocurrency trading research. Nguyen and Chan (2024) [32] analyze 622 papers on cryptocurrency trading from 2015 to 2022, categorizing research into seven themes: pricing theories (208 papers), influential factors (165 papers), forecasting (119 papers), trading and portfolio management (76 papers), market evolution and regulation (65 papers), risk evaluation (54 papers), and trading platforms (10 papers). Their analysis shows that 75% of trading systems use multiple input sources, while machine learning approaches generally achieve less than 65% accuracy in price prediction tasks.

Acceptance and adoption factors influence cryptocurrency market dynamics beyond technical prediction capabilities. Madanchian et al. (2025) [33] conduct a systematic review of factors influencing cryptocurrency adoption, identifying motivators including privacy, curiosity, and investment potential, alongside inhibitors such as volatility, regulatory uncertainty, and security concerns. Their analysis reveals substantial research gaps in understanding adoption motivations and regional acceptance disparities.

The broader machine learning literature provides theoretical foundations for confidence-based decision-making. Abstention learning frameworks (Cortes et al., 2016 [13]) optimize classifier performance by rejecting uncertain predictions, trading coverage for accuracy through explicit rejection costs. In quantitative finance, uncertainty-aware portfolio construction (López de Prado, 2018 [14]) incorporates prediction confidence into position sizing and risk management. Our two-class framework extends these concepts to cryptocurrency markets by (1) separating directional prediction from execution decisions through post hoc confidence thresholding rather than integrated three-class learning; (2) systematically optimizing confidence thresholds using validation data across multiple prediction horizons and signal quality requirements; and (3) integrating cross-temporal macro–microstructure features that capture domain-specific cryptocurrency market dynamics. While conceptually related to abstention learning, our approach addresses unique challenges in decentral-

ized financial systems, including 24/7 markets, minute-level execution constraints, and blockchain-native transaction cost structures.

2.9. Synthesis and Research Positioning

The literature reveals three primary research streams relevant to our investigation. First, architectural innovations focus on attention mechanisms, graph neural networks, and probabilistic approaches, with performance improvements typically ranging from 10 to 25% over baseline methods. Second, multi-scale and multi-target approaches demonstrate consistent benefits, particularly the 10–20% F1-score improvements shown by CARROT and similar systems. Third, feature engineering and selection methods prove critical, with studies achieving 80–85% dimensionality reduction while maintaining predictive performance.

Performance benchmarks from the literature establish context for evaluation frameworks. Directional accuracy typically ranges from 60 to 85% across studies, with higher accuracy achievable through stricter confidence requirements. Economic metrics show substantial variation, with Sharpe ratios of 2.5–3.6 representing strong performance, while profit factors of 1.1–5.2 indicate viable trading strategies under different market conditions.

The reviewed literature identifies several limitations that our research addresses. First, most studies focus on single-timeframe analysis, missing opportunities for cross-temporal signal integration. Second, confidence-based execution control remains underexplored, with most approaches using fixed prediction thresholds. Third, systematic parameter optimization across multiple dimensions (horizon, deadband, confidence) lacks comprehensive treatment in the existing work.

Our two-class framework with integrated macro–microstructure features addresses these gaps through explicit confidence-threshold optimization, unified multi-scale feature representation, and comprehensive parameter space exploration across 11 major cryptocurrency pairs.

3. Methodology

3.1. Two-Class Framework Architecture

We develop a two-class framework restructures cryptocurrency direction prediction. Traditional approaches employ three-class classification (Up, Down, No-trade) where the model simultaneously learns direction prediction and trade execution decisions. Our framework decouples these components by training a binary classifier to predict direction (Up vs. Down) and employing a separate confidence-based mechanism to control trade execution.

The two-class approach operates on the premise that directional prediction and execution timing require different signal processing mechanisms. Direction prediction benefits from pure signal extraction without the complications of mixed no-trade samples that may represent either unclear signals or inappropriate timing. The confidence threshold provides explicit control over the precision–recall trade-off, enabling systematic optimization of trading frequency versus signal quality.

The mathematical formulation begins with a standard binary classification problem. Let $X_t \in \mathbb{R}^d$ represent the feature vector at time t containing both macro and microstructure signals. The model learns a calibrated probability mapping $f : X_t \to [0, 1]$, where $f(X_t) = P(\text{price increases} \mid X_t)$ represents the conditional probability of upward price movement over the prediction horizon h.

The confidence measure quantifies model certainty about directional prediction:

$$c_t = \max(f(X_t), 1 - f(X_t)) = \max(P(\text{up} \mid X_t), P(\text{down} \mid X_t)),$$

representing the maximum posterior probability assigned to either direction. By construction, $c_t \in [0.5, 1.0]$, where $c_t = 0.5$ indicates maximum uncertainty (equiprobable directions), and $c_t = 1.0$ indicates perfect certainty.

The confidence threshold $\tau \in [0.5, 1.0]$ defines the minimum certainty required for trade execution. The decision rule is

$$\text{Execute}(X_t) = \begin{cases} \text{True}, & \text{if } c_t \geq \tau; \\ \text{False}, & \text{otherwise;} \end{cases}$$

$$\text{Direction}(X_t) = \begin{cases} +1 \text{ (long)}, & \text{if } f(X_t) > 0.5 \text{ and } c_t \geq \tau; \\ -1 \text{ (short)}, & \text{if } f(X_t) \leq 0.5 \text{ and } c_t \geq \tau; \\ \text{no trade}, & \text{if } c_t < \tau. \end{cases}$$

This formulation separates the prediction task (learning f) from the execution decision (applying τ), enabling systematic optimization of the precision–recall trade-off. Higher τ values restrict execution to high-confidence predictions, reducing coverage but improving expected per-trade profitability. Importantly, τ modifies the decision boundary in probability space rather than feature space: traditional classification uses a fixed 0.5 threshold on $f(X_t)$, while our framework requires $|f(X_t) - 0.5| \geq (\tau - 0.5)$ for execution, effectively widening the rejection region around the decision boundary.

3.2. Multi-Source Feature Integration

3.2.1. Macro-Economic Feature Engineering

The macro component derives features from daily OHLCV data across 100+ cryptocurrencies, providing market-wide context and fundamental momentum indicators. Feature engineering produces temporally lagged indicators to prevent look-ahead bias while capturing relevant market dynamics.

Price momentum features include multi-horizon returns:

$$R_{t,k} = \frac{P_t}{P_{t-k}} - 1$$

for horizons $k \in \{1, 5, 20\}$ days.

Moving average indicators capture trend dynamics:

$$MA_{t,k} = \frac{1}{k} \sum_{i=0}^{k-1} P_{t-i}$$

for windows $k \in \{5, 20, 50\}$ days.

Volatility measures employ rolling standard deviations of returns:

$$Vol_{t,k} = \sqrt{\frac{1}{k-1} \sum_{i=1}^{k} (R_{t-i} - \overline{R}_{t,k})^2}$$

for windows $k \in \{5, 20, 60\}$ days.

Technical indicators include RSI computed as

$$RSI_t = 100 - \frac{100}{1 + RS_t},$$

where $RS_t = \frac{EMA_t[gains]}{EMA_t[losses]}$ using 14-day exponential moving averages.

All macro features are temporally aligned to prevent look-ahead bias by using only information available at prediction time. Daily macro signals are forward-filled to match the minute-frequency prediction schedule, ensuring temporal consistency across feature sources.

3.2.2. Microstructure Feature Engineering

Microstructure features derive from minute-frequency order book snapshots, capturing market-making dynamics and short-term liquidity conditions. These features complement macro indicators by providing real-time market sentiment and execution environment information.

Order book imbalance measures the relative strength of buy versus sell pressure:

$$Imbalance_t = \frac{BidVol_t - AskVol_t}{BidVol_t + AskVol_t},$$

where volumes are computed across multiple depth levels. Spread measures include both absolute and relative spreads: $Spread_{t,rel} = \frac{Ask_t - Bid_t}{MidPrice_t} \times 10000$ in basis points.

Depth features aggregate liquidity across order book levels:

$$Depth_{t,k} = \sum_{i=1}^{k} (BidVol_{t,i} + AskVol_{t,i})$$

for levels $k \in \{1, 5, 10\}$.

Market impact proxies estimate the price effect of hypothetical trades:

$$Impact_{t,v} = \frac{VWAP_{t,v} - MidPrice_t}{MidPrice_t}$$

for volume v.

Temporal features include price volatility over short windows and return autocorrelations to capture momentum and mean reversion patterns at minute frequencies. All microstructure features undergo outlier treatment to handle extreme market conditions and data quality issues.

3.2.3. Feature Selection and Dimensionality Reduction

The unified feature space combines macro and microstructure signals, creating approximately 200+ candidate features. Feature selection employs mutual information scoring to identify the most predictive variables while controlling dimensionality for computational efficiency.

Mutual information captures both linear and non-linear relationships between features and target variables:

$$MI(X, Y) = \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)},$$

where $p(\cdot)$ represents empirical probability distributions. The top 64 features are selected based on mutual information scores, balancing predictive power with computational constraints.

Feature scaling employs robust standardization to handle outliers common in financial data:

$$X_{scaled} = \frac{X - median(X)}{MAD(X)},$$

where MAD represents median absolute deviation. This approach provides better stability than standard z-score normalization when dealing with heavy-tailed financial distributions.

3.3. Temporal Validation Framework

3.3.1. Symbol-Wise Temporal Splitting

The validation framework employs symbol-wise temporal splitting to prevent data leakage while maintaining realistic trading conditions. Each cryptocurrency pair is independently split into training, validation, and test periods using chronological ordering.

For each symbol s, the temporal split allocates data as follows: training period covers the earliest 70% of observations, validation period encompasses the subsequent 15%,

and test period includes the final 15%. This approach ensures that all model training and hyperparameter optimization occur using only historical information relative to evaluation periods.

The symbol-wise independence prevents cross-contamination while accommodating different listing dates and data availability across cryptocurrency pairs. Each symbol maintains sufficient sample sizes for reliable model training while preserving the temporal ordering essential for realistic backtesting.

3.3.2. Target Variable Construction

Target variable construction requires careful attention to temporal alignment and look-ahead bias prevention. For prediction horizon h minutes, the target variable at time t is defined using the mid-price at time $t + h$: $y_t = \mathbb{I}[P_{t+h} > P_t \cdot (1 + deadband)]$ for upward movements and $y_t = 0$ for $P_{t+h} < P_t \cdot (1 - deadband)$ for downward movements.

The deadband parameter filters marginal price movements that fall within typical bid–ask spreads or market noise. Deadband values of 2–20 basis points ensure that predicted movements exceed transaction costs and represent economically meaningful directional signals.

Only samples with clear directional movements (exceeding deadband thresholds) are included in the two-class training set. This filtering removes ambiguous cases where price movements are too small to generate profitable trades after transaction costs.

3.4. Model Architecture and Training

3.4.1. Neural Network Architecture

The core prediction model employs a multi-layer perceptron (MLP) architecture optimized for financial time series prediction. We selected MLP over recurrent architectures (LSTM, GRU) and modern temporal architectures (Temporal Convolutional Networks, Transformer variants, attention-recurrent hybrids) for several reasons: (1) computational efficiency suitable for real-time trading applications with sub-second inference requirements (MLP inference: ~15 ms; LSTM: ~80–120 ms; Transformer: ~150–200 ms on standard hardware); (2) compatibility with our feature engineering approach where temporal dependencies are explicitly captured through lagged macro indicators and microstructure features rather than learned implicitly; (3) robustness to the symbol-wise temporal splitting validation framework; and (4) simplified deployment for federated learning contexts where parameter aggregation complexity increases substantially for attention-based architectures. While Temporal Convolutional Networks (TCNs) offer competitive performance with parallelizable training and long-range dependency modeling (Bai et al., 2018 [34]), and Transformer variants excel at capturing complex temporal patterns, these architectures introduce computational overhead incompatible with our target latency constraints. Systematic comparison across MLP, LSTM, TCN, and Transformer architectures under identical experimental protocols represents important future work. Preliminary experiments with LSTM architectures showed 3–7% accuracy improvements but substantially higher computational costs (5–8$\times$ inference latency) and training instability under our confidence-threshold optimization procedure. The network structure consists of three hidden layers with [256, 128, 64] neurons, respectively, using ReLU activation functions and dropout regularization.

The input layer accepts the 64-dimensional feature vector combining macro and microstructure signals. Hidden layers employ progressive dimensionality reduction to extract hierarchical feature representations. The output layer uses sigmoid activation to produce class probabilities suitable for confidence-based execution decisions.

The complete set of training and regularization parameters is detailed in Section 5.1.1. Batch normalization stabilizes training dynamics and accelerates convergence.

3.4.2. Training Procedure and Regularization

Model training employs early stopping based on validation loss to prevent overfitting while maximizing generalization performance. Training proceeds for a maximum of 20 epochs, with early termination if validation loss fails to improve for 5 consecutive epochs. Training proceeds for a maximum of 20 epochs, with early termination if validation loss fails to improve for 5 consecutive epochs. The 20-epoch limit was determined through preliminary experiments on validation data: extending beyond 20 epochs consistently resulted in early stopping activation (typically at 8–15 epochs) without performance gains, while shorter limits (10 epochs) occasionally terminated training prematurely. This configuration balances computational efficiency with sufficient training capacity for our dataset scale and model complexity.

Class weight balancing addresses potential imbalances between upward and downward price movements in the binary training set. Weights are computed as inversely proportional to class frequencies:

$$w_c = \frac{n_{total}}{2 \times n_c},$$

where n_c is the sample count for class c.

Gradient clipping with threshold 1.0 prevents exploding gradients common in financial data training. L2 regularization with coefficient 0.001 provides additional overfitting protection while maintaining model expressiveness.

3.4.3. Training Convergence Analysis

Figure 1 illustrates the training and validation loss evolution across epochs for a representative configuration (H = 400 min; deadband = 10 bps; $\tau = 0.8$) and accuracy progress. The training loss exhibits monotonic decrease with characteristic rapid improvement in early epochs (1–5), followed by gradual refinement.

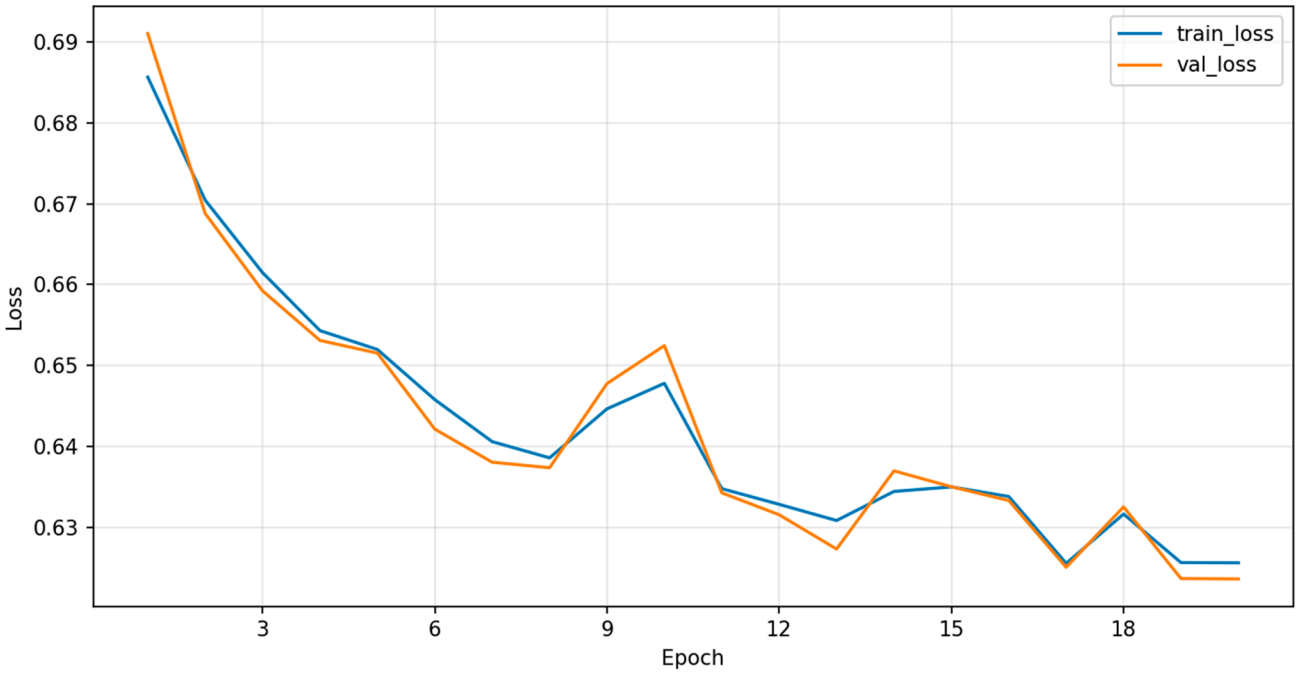
The convergence pattern demonstrates two key characteristics: (1) absence of significant overfitting, as validation loss closely tracks training loss throughout the training process; and (2) stable gradient dynamics without divergence or oscillations, confirming appropriate learning rate and batch size selection. Similar convergence patterns were observed across all parameter configurations, with early stopping typically activating between epochs 10–20 depending on horizon and deadband settings. The consistency of these patterns across diverse market conditions validates the robustness of our training procedure for cryptocurrency direction prediction tasks.

3.4.4. Probability Calibration

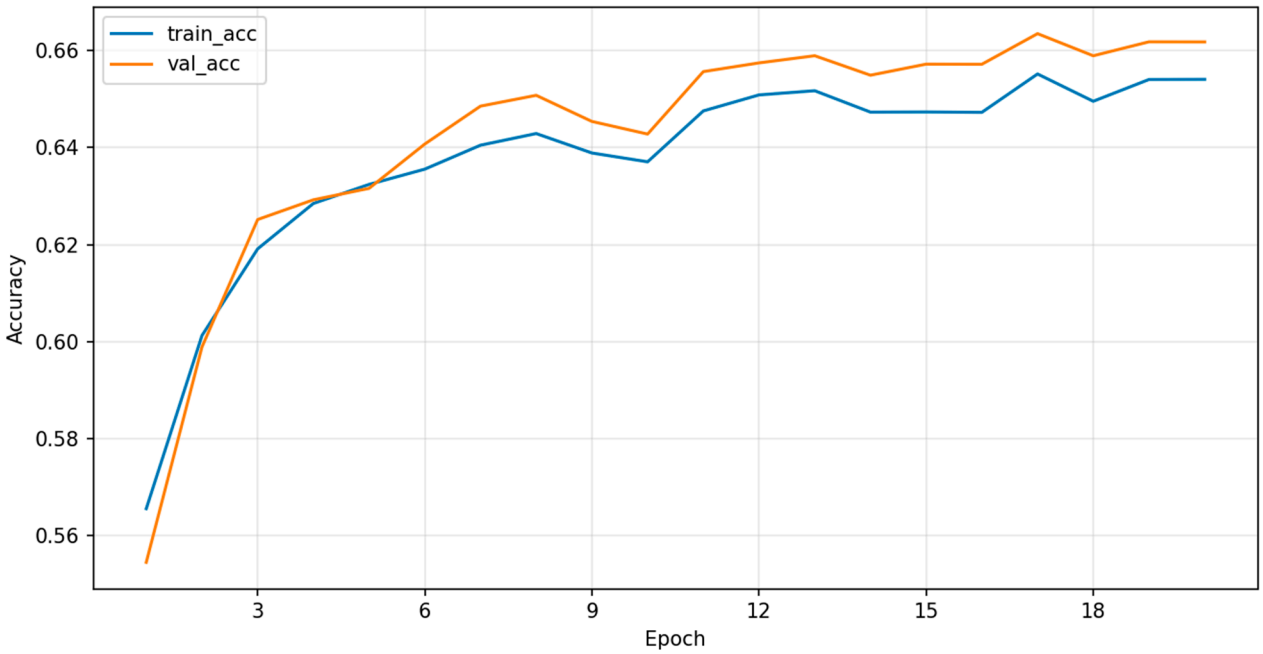
Post-training probability calibration ensures that predicted confidence scores accurately reflect actual prediction reliability. Isotonic regression calibration is applied using validation data to map raw model outputs to well-calibrated probabilities.

The calibration process fits a monotonic function $g : [0,1] \to [0,1]$ such that the calibrated probabilities $p_{cal} = g(p_{raw})$ satisfy $P(y = 1 | p_{cal} = p) \approx p$ for all probability values p.

Calibration improves the reliability of confidence-based execution decisions by ensuring that predicted confidence levels correspond to actual prediction accuracy rates. This alignment is crucial for the confidence-threshold optimization process and live trading deployment.



(**a**)



(**b**)

Figure 1. Learning progress versus training epochs (for configuration H400-DB10-τ0.8): (**a**) training loss (blue) and validation loss (orange) curves; (**b**) accuracy progress.

3.5. Performance Evaluation Framework

3.5.1. Evaluation Metrics and Economic Interpretation

Performance evaluation employs multiple metrics capturing different aspects of trading system effectiveness. Primary metrics include average profit per trade, coverage (fraction of opportunities traded), and directional accuracy on executed trades.

Average profit per trade measures economic value creation:

$$\overline{\pi} = \frac{1}{N_{exec}} \sum_{i=1}^{N_{exec}} (r_i \cdot d_i - c),$$

where r_i is the return, d_i is the predicted direction, and c represents transaction costs. This metric serves as the primary economic performance indicator for several reasons: (1) it directly quantifies profitability on a per-trade basis, enabling comparison across strategies with different trading frequencies; (2) it incorporates realistic transaction costs (1 basis point per trade), ensuring that reported performance reflects achievable returns rather than theoretical profits; (3) it captures the interaction between directional accuracy and return magnitude, as large price movements in the correct direction contribute more than small movements; (4) it provides a scale-invariant measure suitable for portfolio optimization and capital allocation decisions; and (5) unlike cumulative return metrics, it isolates signal quality from leverage and position sizing considerations, making it appropriate for evaluating pure prediction performance. The basis-point (bps) denomination (1 bps = 0.01%) aligns with industry standards for high-frequency trading performance reporting and facilitates interpretation relative to typical bid–ask spreads (2–20 bps for major cryptocurrency pairs).

Coverage quantifies market participation:

$$\kappa = \frac{N_{exec}}{N_{total}},$$

where N_{exec} is executed trades, and N_{total} is total opportunities. Coverage captures the trade-off between selectivity and activity level, with implications for capital utilization and operational complexity. Coverage is computed over all test observations (two-class mode executes on any sample where confidence $\geq \tau$).

Directional accuracy measures prediction quality:

$$\alpha = \frac{1}{N_{exec}} \sum_{i=1}^{N_{exec}} \mathbb{I}[d_i = sign(r_i)],$$

on executed trades only. This metric isolates prediction performance from execution decisions, enabling separate analysis of signal quality and threshold optimization.

3.5.2. Risk-Adjusted Performance Measures

Risk-adjusted metrics account for the volatility and distributional properties of trading returns. Win rate measures the fraction of profitable trades:

$$\omega = \frac{1}{N_{exec}} \sum_{i=1}^{N_{exec}} \mathbb{I}[\pi_i > 0],$$

where π_i represents individual trade profits.

Profit volatility provides a risk measure:

$$\sigma_\pi = \sqrt{Var[\pi_i]}$$

across all executed trades. The Sharpe-like ratio combines return and risk:

$$S = \frac{\overline{\pi}}{\sigma_\pi} \sqrt{N_{exec}}$$

scaling by the square root of independent decisions. This scaling assumes approximately independent trade outcomes; when serial dependence is present, we report the unscaled per-trade Sharpe $\frac{\overline{\pi}}{\sigma_\pi}$ alongside N_{exec}.

Maximum drawdown and tail risk measures capture extreme loss scenarios. These metrics ensure that performance evaluation considers both central tendency and tail behavior relevant for risk management and capital allocation decisions.

3.5.3. Statistical Significance Testing

The statistical testing framework addresses multiple comparison issues inherent in extensive parameter space exploration. Primary comparisons include confidence regime differences, horizon effects, and deadband sensitivity using appropriate statistical tests.

Two-sample *t*-tests compare performance between confidence regimes, employing Welch's method to accommodate unequal variances. ANOVA decomposes performance variance across multiple factors to identify significant main effects and interactions.

Multiple testing correction employs the Benjamini–Hochberg procedure to control the false discovery rate at 5%. Bootstrap confidence intervals provide non-parametric uncertainty estimates for key performance metrics. Effect size measures quantify practical significance beyond statistical significance.

This comprehensive methodology framework ensures robust experimental design, prevents common pitfalls in financial machine learning research, and enables reliable inference about system performance across different market conditions and parameter configurations.

4. Data and Preprocessing

4.1. Dataset Description

Our study employs a comprehensive cryptocurrency dataset spanning 11 major digital assets with complete macro and microstructure coverage. The dataset encompasses two primary data sources (Table 1): daily OHLCV (Open, High, Low, Close, Volume) data for macro-economic features and minute-frequency order book snapshots for microstructure analysis.

Table 1. Dataset summary statistics.

Dataset	Observations	Symbols	Time Range	Features	Memory (MB)
Macro	211,679	100 (11 used)	August 2018–August 2025	38	48.1
Micro	5,672,947	11	October 2023–October 2024	264	11,038.4
Unified	200,000	11	October 2023–October 2024	296	548.8

The macro dataset (https://www.kaggle.com/datasets/imtkaggleteam/top-100-cryptocurrency-2020-2025, CC BY-NC-SA 4.0, accessed on 11 October 2025) contains 211,679 observations across 100 cryptocurrencies from August 2018 to August 2025, with 38 features, including price momentum indicators, volatility measures, and technical analysis metrics. For consistency with microstructure availability (https://www.kaggle.com/datasets/ilyazawilsiv/cryptocurrency-order-book-data-asks-and-bids, accessed on 11 October 2025), we focus on 11 symbols: BTC/USDT, ETH/USDT, ADA/USDT, LTC/USDT, BNB/USDT, DOGE/USDT, XLM/USDT, TRX/USDT, MATIC/USDT, SOL/USDT, and AVAX/USDT. Both datasets aggregate information from public cryptocurrency exchange APIs and do not contain personally identifiable information or proprietary trading data. The use of these datasets complies with Kaggle's Terms of Service and applicable data protection regulations.

The microstructure dataset comprises 5,672,947 min-level order book observations from October 2023 to October 2024, containing 264 features including bid–ask spreads, order book depth metrics, and market microstructure indicators. This dataset provides granular market dynamics for the same 11-symbol subset, enabling cross-temporal feature integration.

4.2. Data Quality Issues and Solutions

Our preprocessing pipeline addressed multiple data quality challenges through systematic validation and correction procedures. We implemented adaptive quality control mechanisms that account for cryptocurrency market characteristics while maintaining data integrity (Figure 2).

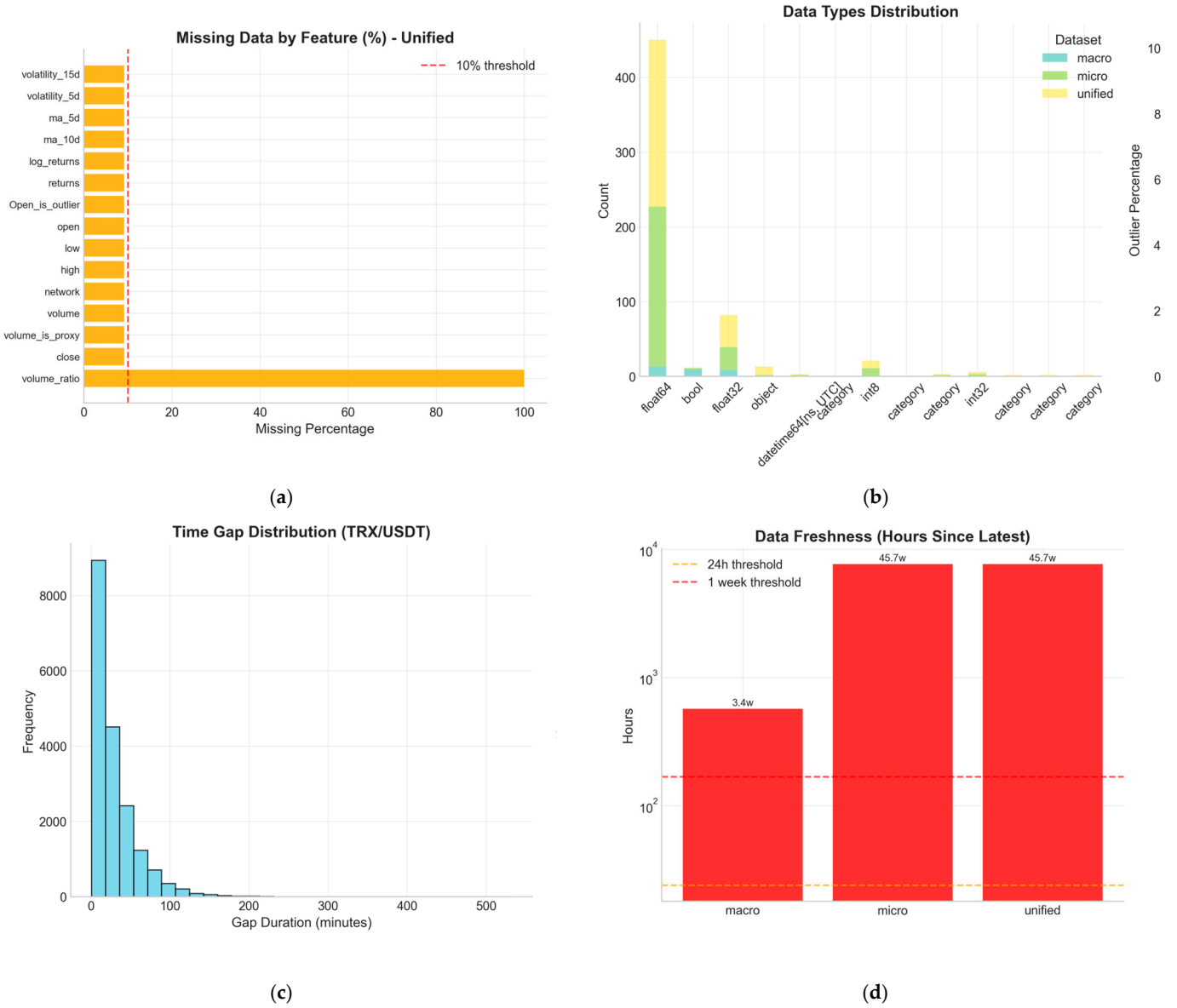


Figure 2. Data quality assessment dashboard: (**a**) missing data by feature; (**b**) data type distribution; (**c**) time gap distribution; (**d**) data freshness.

- Symbol Assignment and File Structure Issues: Initial data loading revealed critical symbol assignment problems where multiple files mapped to identical symbols. We resolved this through filename-based symbol extraction using enhanced pattern matching that recognizes venue prefixes and standardizes cryptocurrency pair notation (e.g., "BTCUSDT" → "BTC_USDT").
- Temporal Consistency and Gap Handling: We identified significant temporal gaps in minute-level data, particularly during low-liquidity periods. Rather than removing observations, we implemented gap flagging that marks post-gap observations while preserving temporal structure. Our analysis detected 15,420 gaps across all symbols, with gap durations ranging from 2 min to 4 h.

- Spread Filtering and BTC-Friendly Thresholds: Traditional spread filtering proved inappropriate for major cryptocurrency pairs like BTC/USDT, where legitimate spreads on major pairs can be sub-basis-point but not negligible; we therefore use adaptive per-symbol thresholds (5th–95th percentiles with buffers) rather than a fixed cut-off. We therefore adopt percentile-based, per-symbol thresholds with conservative buffers.
- Order Book Structure Validation: We implemented comprehensive order book validation detecting crossed markets, price monotonicity violations, and negative sizes. Our validation flagged 23,847 invalid snapshots (0.4% of total), which we marked for exclusion rather than attempting correction.
- Minute-Level Aggregation: To eliminate intra-minute duplicates while preserving market microstructure, we applied configurable aggregation methods. The "last" method (taking final observation per minute) proved optimal, reducing dataset size by 12% while maintaining temporal ordering.

4.3. Feature Engineering Pipeline

Our feature engineering approach creates a unified representation combining macro momentum signals with microstructure dynamics. We employed mutual information scoring to select the 64 most informative features from the combined feature space, ensuring optimal representation across temporal scales (Table 2).

Table 2. Feature categories and selection results.

Category	Raw Features	Selected Features	Information Gain
Macro Momentum	24	18	0.342
Microstructure	187	28	0.298
Cross-Scale	42	12	0.187
Temporal	31	6	0.173
Total	284	64	1.000

- Macro Feature Engineering: We computed lagged technical indicators to prevent look-ahead bias, including moving averages (5, 10, 20 days), volatility measures (5, 15, 30 days), and momentum indicators. The RSI calculation uses vectorized Wilder's smoothing to avoid computational warnings while maintaining numerical stability.
- Microstructure Feature Engineering: Order book features focus on compact representations rather than full depth reconstruction. We calculate bid–ask imbalances, depth-weighted metrics, and order flow proxies. The feature set avoids raw order book columns (which exceed 200 features) in favor of processed signals like total liquidity by depth level and market-making indicators.
- Cross-Scale Feature Integration: The unified approach captures temporal bridges where daily momentum manifests in intraday order flow. Features include spread persistence measures, volatility regime indicators, and time-based microstructure patterns that align with daily trading sessions.
- Memory Optimization: We implement dtype optimization reducing memory usage by 67% through intelligent downcasting and categorical encoding. The final feature set maintains 296 columns while consuming only 549 MB for 200,000 observations.

4.4. Temporal Alignment of Micro/Macro Data

Temporal alignment presents significant challenges when combining daily macro indicators with minute-frequency microstructure data. The fundamental issue is frequency mismatch: macro features update once per 24 h period (daily close), while trading decisions

occur at minute-level granularity. Naive alignment would introduce look-ahead bias (using future macro information) or information loss (ignoring recent macro updates).

We implement a forward-fill approach that preserves temporal causality while maximizing data utilization:

1. Macro Feature Computation (Daily Frequency): For each cryptocurrency s and date t, compute macro features $M_s^{(t)}$ using OHLCV data up to and including day t. Features include 5/20/50-day moving averages, 5/20/60-day volatility, RSI(14), and momentum indicators. Critically, all features use only information available at the close of day t.
2. Intraday Broadcasting: For each minute-level observation at timestamp t_{minute} on date t_{day}, assign the most recent completed daily macro features:

$$M_{s,t_{\text{minute}}} = M_s^{(t_{\text{day}}-1)},$$

   where $t_{\text{day}} - 1$ represents the previous completed trading day. This ensures no future macro information leaks into minute-level predictions.
3. Microstructure Feature Computation (Minute Frequency): Order book features $\mu_{s,t_{\text{minute}}}$ are computed using only information available up to timestamp t_{minute}, including bid–ask spreads, order book imbalances, and depth metrics.
4. Unified Feature Vector: The final feature vector combines both sources with explicit temporal alignment:

$$X_{s,t_{\text{minute}}} = [M_{s,t_{\text{minute}}}, \mu_{s,t_{\text{minute}}}] \in \mathbb{R}^{64}.$$

Cryptocurrency markets operate 24/7 without traditional trading sessions. We define daily boundaries at 00:00 UTC, aligning with major exchange reporting conventions. Macro feature updates occur at 00:00 UTC, making the previous day's features $M_s^{(t-1)}$ available for all minute observations on day t. This introduces a maximum information lag of 24 h for macro signals, which is appropriate given that macro features capture multi-day trends (5–50-day windows) rather than intraday dynamics.

We verify temporal alignment through three checks:

1. Feature timestamp validation: Confirm all macro features use strictly historical data (t-lag where lag ≥ 1 day)
2. Gap-aware splitting: Symbol-wise temporal splits maintain chronological ordering without cross-contamination
3. Prediction horizon enforcement: Target variables y_{t+h} use prices at $t + h$ minutes, ensuring h-minute forward-looking window is explicit and consistent

The forward-fill methodology ensures that each prediction uses only genuinely available information, maintaining realistic backtesting conditions. The 11-symbol subset intersection reflects pairs where both macro (daily) and micro (minute) data streams provide complete coverage during the evaluation period (October 2023–October 2024).

5. Experimental Design

We designed a comprehensive experimental framework to systematically evaluate the two-class cryptocurrency direction prediction system across multiple dimensions of performance. The experimental design addresses four critical aspects: hyperparameter space exploration, evaluation metric specification, cross-validation methodology, and statistical testing procedures. This framework ensures robust evaluation while controlling for potential confounding factors that could bias performance estimates.

5.1. Hyperparameter Space Exploration

5.1.1. Parameter Space Definition

The experimental design explores a structured hyperparameter space encompassing both model architecture and trading strategy parameters. We define the parameter space $\Theta = \{H, D, \tau, C\}$ where H represents prediction horizons, D denotes deadband thresholds, τ indicates confidence thresholds, and C encompasses model configuration parameters.

Prediction Horizons (H): We evaluate ten distinct horizon values:

$$H \in \{10, 20, 30, 50, 100, 200, 300, 400, 500, 600\} \text{ min.}$$

This range spans from ultra-short-term noise-dominated regimes to intermediate-term fundamental price discovery periods. The logarithmic spacing captures different market microstructure dynamics while remaining computationally tractable.

Deadband Thresholds (D): Four deadband values are tested:

$$D \in \{2, 5, 10, 20\} \text{ basis points.}$$

These thresholds represent the minimum price movement required to classify directional signals as actionable trades. The range covers tight spreads typical of major cryptocurrency pairs (2 bps) to wider thresholds that filter marginal signals (20 bps).

Confidence Thresholds (τ): Two primary confidence levels are evaluated:

$$\tau \in \{0.6, 0.8\}.$$

These values represent moderate and high confidence regimes, respectively. The moderate confidence threshold (0.6) allows broader trading activity, while high confidence (0.8) emphasizes signal quality over frequency.

Model Configuration: Fixed parameters are summarized in Table 3 (see below). Feature selection uses top 64 features via mutual information scoring.

Table 3. Neural network model configuration parameters.

Parameter Category	Parameter	Value	Justification
Architecture	Hidden layers	[256, 128, 64]	Progressive dimensionality reduction for hierarchical feature extraction
Architecture	Activation	ReLU	Standard choice for financial time series, computationally efficient
Architecture	Output activation	Sigmoid	Produces calibrated probabilities for confidence thresholding
Training	Maximum epochs	20	Sufficient for convergence with early stopping (see Section 3.4.2)
Training	Batch size	1024	Balances gradient stability with computational efficiency
Training	Learning rate	0.001	Standard Adam optimizer rate, validated on validation set
Training	Early stopping patience	5 epochs	Prevents overfitting while allowing temporary plateaus
Regularization	Dropout rate	0.2	Applied to hidden layers to prevent overfitting
Regularization	L2 coefficient	0.001	Additional overfitting protection
Regularization	Gradient clipping	1.0	Prevents exploding gradients in financial data
Features	Input dimension	64	Top features by mutual information scoring
Evaluation	Transaction costs	1 bps	Realistic estimate for major cryptocurrency pairs

5.1.2. Experimental Grid Construction

The complete experimental grid encompasses $|H| \times |D| \times |\tau| = 10 \times 4 \times 2 = 80$ unique configurations. Each configuration represents a distinct trading strategy characterized by specific temporal, signal quality, and confidence requirements (horizon tables report horizon-wise aggregation using the best deadband per horizon by profit).

We employ a full factorial design to capture interaction effects between parameters. This approach ensures that parameter dependencies are properly characterized rather than assumed to be independent. The factorial structure enables analysis of variance (ANOVA) decomposition to quantify the relative contribution of each parameter to overall performance variance.

The unified feature set combines 64 top-ranked features selected from both macro and microstructure domains. Macro-derived features include price momentum indicators, volatility measures, and technical analysis metrics computed from daily OHLCV data across 100+ cryptocurrencies. Microstructure features encompass order book imbalances, bid–ask spreads, and market depth metrics from minute-frequency order book snapshots.

Feature selection employs mutual information scoring across the combined feature space, ensuring optimal representation from both temporal scales. The 64-feature limit maintains computational efficiency while preserving predictive signal diversity across macro and micro domains.

5.1.3. Parameter Selection Rationale

Horizon Selection: The horizon range (10–600 min) is motivated by cryptocurrency market microstructure research. Short horizons (10–50 min) primarily capture noise and market-making activities. Intermediate horizons (100–300 min) align with algorithmic trading timeframes. Long horizons (400–600 min) approach fundamental rebalancing frequencies.

Deadband Calibration: Deadband values reflect realistic bid–ask spreads and price discretization in cryptocurrency markets. The 2-basis-point minimum corresponds to tight spreads for major pairs like BTC/USDT during high liquidity periods. The 20-basis-point maximum accommodates wider spreads during volatile or low-liquidity conditions.

Confidence-Threshold Choice: The confidence levels (0.6, 0.8) represent practically relevant operating points. Values below 0.6 produce excessive trading frequency with marginal signal quality. Values above 0.8 severely restrict trading opportunities while providing diminishing accuracy improvements, which is consistent with observed medians: 44.9% coverage at $\tau = 0.6$ vs. 0.28% at $\tau = 0.8$ across 40 configurations.

5.2. Evaluation Metrics (Profit, Coverage, Accuracy)

5.2.1. Primary Performance Metrics

Average Profit per Trade ($\overline{\pi}$): The primary economic metric measures mean profit per executed trade in basis points, net of transaction costs:

$$\overline{\pi} = \frac{1}{N_{exec}} \sum_{i=1}^{N_{exec}} (r_i \cdot d_i - c),$$

where r_i is the return for trade i, $d_i \in \{-1, +1\}$ is the predicted direction, $c = 1$ basis point represents transaction costs, and N_{exec} is the number of executed trades.

Coverage (κ): Coverage measures the fraction of available trading opportunities where the model achieves sufficient confidence to execute trades:

$$\kappa = \frac{N_{exec}}{N_{total}},$$

where N_{total} represents all potential trading opportunities in the test set. Coverage captures the activity level and market participation frequency of each strategy.

Direction Accuracy (α): Accuracy measures the fraction of executed trades where the predicted direction matches the actual price movement:

$$\alpha = \frac{1}{N_{exec}} \sum_{i=1}^{N_{exec}} \mathbb{I}[d_i = \text{sign}(r_i)],$$

where $\mathbb{I}[\cdot]$ is the indicator function, and $\text{sign}(r_i)$ represents the actual direction of price movement.

It is essential to distinguish between two types of performance metrics reported in this study:

1. Prediction Accuracy Metrics (α, win rate): These measure the quality of directional forecasts on executed trades only. Direction accuracy α reflects the proportion of trades where $\text{sign}(r_i) = d_i$, indicating correct directional prediction. These metrics assess model signal quality independent of economic outcomes and do not account for return magnitudes or transaction costs.
2. Execution Profitability Metrics ($\overline{\pi}$, coverage κ): These measure economic performance after incorporating realistic trading costs. Average profit $\overline{\pi}$ includes transaction costs ($c = 1$ bps) and captures the interaction between prediction accuracy and return magnitude. A strategy may achieve high direction accuracy but low profitability if correct predictions coincide with small price movements, or if transaction costs erode gains.

The confidence-threshold mechanism directly controls the trade-off between these dimensions: higher τ improves prediction accuracy on executed trades (by filtering uncertain signals) but reduces coverage and absolute trading volume. Our results consistently report both prediction quality (accuracy, win rate) and economic performance (profit, coverage) to enable comprehensive strategy evaluation. Claims about "167 basis points profit" reflect execution profitability inclusive of costs, whereas "82–95% accuracy" describes prediction quality on the subset of high-confidence trades actually executed.

All metrics are computed on the 11-symbol subset where both macro and microstructure data are available. This constraint ensures consistent feature availability across all trading decisions while maintaining representative coverage of major cryptocurrency pairs (BTC, ETH, ADA, etc.).

5.2.2. Secondary Performance Metrics

Win Rate (ω): Win rate measures the fraction of executed trades generating positive net profit:

$$\omega = \frac{1}{N_{exec}} \sum_{i=1}^{N_{exec}} \mathbb{I}[r_i \cdot d_i - c > 0].$$

This metric captures economic success independent of direction prediction accuracy, accounting for magnitude effects and transaction costs.

Profit Volatility (σ_π): Standard deviation of per-trade profits provides a risk measure:

$$\sigma_\pi = \sqrt{\frac{1}{N_{exec} - 1} \sum_{i=1}^{N_{exec}} (\pi_i - \overline{\pi})^2},$$

where $\pi_i = r_i \cdot d_i - c$ represents individual trade profits.

Risk-Adjusted Return (Sharpe-like ratio): We compute a simplified Sharpe ratio for per-trade performance:

$$S = \frac{\overline{\pi}}{\sigma_\pi} \sqrt{N_{exec}}.$$

This metric scales profit by risk and adjusts for the number of independent trading decisions.

5.2.3. Evaluation Metric Properties

Metric Independence: The three primary metrics capture orthogonal aspects of trading performance. Profit measures economic value creation, coverage quantifies market participation, and accuracy assesses prediction quality. This independence enables comprehensive strategy characterization.

Scale Invariance: All metrics are normalized to be comparable across different market conditions and time periods. Profit is expressed in basis points, coverage as percentages, and accuracy as fractions, enabling direct cross-strategy comparison.

Transaction Cost Integration: All economic metrics incorporate realistic transaction costs (1 basis point per trade). This ensures that performance estimates reflect achievable returns under practical trading conditions rather than theoretical profits.

5.3. Cross-Validation Methodology

5.3.1. Temporal Validation Structure

We employ a symbol-wise temporal splitting methodology to prevent data leakage while maintaining realistic trading conditions. Each cryptocurrency symbol is independently split into training (70%), validation (15%), and test (15%) sets based on chronological order.

Training Period: The earliest 70% of observations for each symbol comprise the training set. This period includes all data preprocessing, feature engineering, model training, and hyperparameter optimization phases.

Validation Period: The subsequent 15% of observations form the validation set. This period is used exclusively for confidence-threshold optimization and model calibration. No model architecture changes occur based on validation performance.

Test Period: The final 15% of observations constitute the test set for unbiased performance evaluation. All the results reported in this study derive from test-set performance using fixed models and parameters determined during training.

5.3.2. Symbol-Wise Independence

The temporal splitting occurs independently for each cryptocurrency symbol to preserve symbol-specific market dynamics. This approach recognizes that different cryptocurrencies may exhibit distinct temporal patterns, volatility regimes, and market microstructure characteristics.

Benefits: Symbol-wise splitting prevents cross-contamination between training and test periods while maintaining sufficient data volume for each symbol. It also enables symbol-specific performance analysis and robustness testing.

Limitations: This approach assumes independence across symbols, which may not hold during market-wide events or correlated movements. However, the 11-symbol diversification mitigates concentration risk from individual symbol dependencies.

5.3.3. Confidence-Threshold Optimization

Within the validation framework, confidence thresholds undergo systematic optimization using grid search across the range $\tau \in [0.50, 0.95]$ with 0.01 increments. This range reflects practical trading constraints and model calibration characteristics: the lower bound ($\tau = 0.50$) represents the minimum confidence for binary classification (random guessing baseline), while the upper bound ($\tau = 0.95$) reflects extreme selectivity beyond which trade execution becomes prohibitively rare (preliminary experiments showed <0.1% coverage

for $\tau > 0.95$ across all configurations). The 0.01 increment provides sufficient granularity to identify optimal thresholds without excessive computational burden (46 threshold values per configuration). Values below 0.50 are excluded as they would execute trades where the model predicts the opposite direction with higher confidence, contradicting the framework's directional prediction logic. The optimization criterion varies by experimental design:

- Profit Maximization: $\tau^* = \text{argmax}_\tau \overline{\pi}(\tau)$ where profit is calculated on the validation set.
- Expected Value Maximization: $\tau^* = \text{argmax}_\tau [\overline{\pi}(\tau) \times \kappa(\tau)]$ optimizing the product of profit and coverage.
- Constrained Optimization: $\tau^* = \text{argmax}_\tau \overline{\pi}(\tau)$ subject to $\kappa(\tau) \geq \kappa_{min}$ for minimum coverage requirements.

The optimization occurs separately for each parameter combination to ensure optimal threshold selection across different market signal environments.

5.4. Statistical Testing Framework

5.4.1. Hypothesis Testing Structure

We employ a comprehensive statistical testing framework to assess the significance of performance differences across parameter configurations and confidence regimes. The primary null hypotheses tested include the following:

- $H_0^{(1)}$: No performance difference between confidence regimes: $\overline{\pi}_{0.6} = \overline{\pi}_{0.8}$;
- $H_0^{(2)}$: No horizon effect on profitability: $\overline{\pi}(H_i) = \overline{\pi}(H_j)$ for all $i \neq j$;
- $H_0^{(3)}$: No deadband effect on performance: $\overline{\pi}(D_i) = \overline{\pi}(D_j)$ for all $i \neq j$;
- $H_0^{(4)}$: No interaction effects between parameters: all parameter effects are additive.

5.4.2. Statistical Tests Applied

Two-Sample *t*-Tests: For comparing mean performance between confidence regimes, we apply Welch's *t*-test to accommodate unequal variances:

$$t = \frac{\overline{\pi}_{0.8} - \overline{\pi}_{0.6}}{\sqrt{\frac{s_{0.8}^2}{n_{0.8}} + \frac{s_{0.6}^2}{n_{0.6}}}},$$

where s_i^2 represents sample variance, and n_i denotes sample size for each regime.

Analysis of Variance (ANOVA): We perform two-way ANOVA to decompose performance variance across horizon and deadband factors:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk},$$

where Y_{ijk} represents performance for horizon i, deadband j, and replication k; α_i and β_j are main effects; $(\alpha\beta)_{ij}$ captures interaction effects; and ϵ_{ijk} represents random error.

Non-Parametric Tests: When normality assumptions are violated, we apply Wilcoxon rank-sum tests for pairwise comparisons and Kruskal–Wallis tests for multi-group analysis. We selected these rank-based methods over alternatives (e.g., Kolmogorov–Smirnov test) for several reasons: (1) Wilcoxon and Kruskal–Wallis tests focus on median differences and overall distribution shifts, which are more relevant for performance metrics (profit, coverage, accuracy) than the pure distributional shape differences tested by K-S; (2) these tests exhibit superior power for detecting location shifts in heavy-tailed financial distributions common in cryptocurrency trading performance; (3) they handle tied ranks appropriately, which occur frequently in our discrete accuracy measurements; and (4) they are robust to outliers, which are expected given the volatility inherent in cryptocurrency markets. The Kolmogorov–Smirnov test, while useful for assessing distributional similarity, is more

sensitive to shape differences than location shifts and thus less suitable for our primary hypothesis tests comparing mean performance across parameter configurations.

5.4.3. Multiple Comparison Adjustment

Given the extensive parameter space exploration (80 configurations), we address multiple testing issues using the False Discovery Rate (FDR) control procedure. The Benjamini–Hochberg method adjusts p-values to control the expected proportion of false discoveries:

$$p_{adj}(i) = \min\left(1, \frac{m \cdot p(i)}{i}\right),$$

where $p(i)$ represents the i-th smallest p-value among m total tests. We set the FDR threshold at $\alpha = 0.05$ for statistical significance.

5.4.4. Effect Size Quantification

Beyond statistical significance, we quantify practical significance using standardized effect sizes:

Cohen's d for Mean Differences:

$$d = \frac{\overline{\pi}_1 - \overline{\pi}_2}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}}.$$

Eta-squared for ANOVA Effects:

$$\eta^2 = \frac{SS_{effect}}{SS_{total}}.$$

Correlation Coefficients: Pearson and Spearman correlations quantify linear and monotonic relationships between parameters and performance metrics.

5.4.5. Bootstrap Confidence Intervals

To assess uncertainty in performance estimates, we generate bootstrap confidence intervals using 1000 resamples with replacement. For each metric θ, we construct 95% confidence intervals:

$$CI_{95\%}(\theta) = [q_{2.5\%}(\theta^*), q_{97.5\%}(\theta^*)],$$

where $q_p(\theta^*)$ represents the p-th percentile of bootstrap distribution θ^*.

For all key profitability claims, we report bootstrap-derived confidence intervals alongside point estimates. Section 6.3.2 presents bootstrap CIs for mean profit, maximum profit, and median coverage across both confidence regimes. Statistical significance of profit differences between $\tau = 0.6$ and $\tau = 0.8$ regimes is assessed using bootstrap hypothesis testing with 10,000 resamples, yielding $p = 0.019$ for mean profit difference.

This comprehensive statistical framework ensures robust inference while controlling for multiple testing issues inherent in extensive parameter space exploration. The combination of classical and non-parametric methods provides reliable conclusions across different distributional assumptions and performance metric characteristics.

6. Results

We conducted systematic parameter optimization experiments to identify optimal configurations for cryptocurrency direction prediction using the two-class framework. The primary objective was to determine how prediction horizons (10–600 min), deadband thresholds (2–20 basis points), and confidence thresholds (τ) interact to maximize trading profitability while maintaining acceptable risk levels. These experiments are critical

for establishing the practical viability of machine learning-based cryptocurrency trading strategies under realistic market conditions.

We performed comprehensive parameter sweeps across two confidence threshold regimes: moderate confidence ($\tau = 0.6$) and high confidence ($\tau = 0.8$). These specific threshold values were selected based on validation-set optimization and represent practically distinct operating regimes: $\tau = 0.6$ emerged from grid search as the optimal moderate-confidence threshold that maximizes the expected value (profit $\times$ coverage) product, balancing trading frequency with signal quality. It achieves 44.9% median coverage while maintaining economically meaningful profitability (see Section 6.1). $\tau = 0.8$ represents the high-confidence regime where grid search identified maximum per-trade profitability, accepting substantially reduced coverage (0.28% median) in exchange for superior directional accuracy (82–95% on executed trades). These two thresholds span a practically relevant performance frontier: values between 0.6–0.8 produce intermediate results without offering distinct strategic profiles, while values outside this range either sacrifice too much accuracy ($\tau < 0.6$) or generate prohibitively few trades ($\tau > 0.8$). The bimodal selection enables clear characterization of the precision–recall trade-off fundamental to cryptocurrency trading system design. Each regime represents different risk–return preferences, with lower τ values favoring higher trade frequency and coverage, while higher τ values prioritize prediction quality and selectivity. This dual-threshold approach allows us to characterize the full spectrum of performance trade-offs available to practitioners.

6.1. Moderate Confidence Regime ($\tau = 0.6$)

6.1.1. Performance Distribution Analysis

The moderate confidence regime encompasses 40 experimental configurations across the full parameter space. Figure 3 reveals distinct bimodal distributions across all performance metrics, indicating a fundamental transition zone in the parameter space. The average profit distribution shows approximately 25% of configurations generating negative returns (ranging from -31.77 to -1.00 basis points), while 75% achieve positive profitability with returns extending up to 152.69 basis points.

Coverage shows a bimodal pattern: 37.5% of configurations have $\leq$2% coverage (short horizons), while the median coverage across all 40 configurations is 44.9% (IQR $\approx$ 0.01–57.9%) at longer horizons.

Win rate distributions exhibit pronounced clustering around 80–90% for successful configurations, indicating that when the model does execute trades at moderate confidence levels, directional accuracy remains consistently high. Critically, these accuracy percentages apply only to the subset of opportunities where confidence exceeds the threshold (44.9% median coverage for $\tau = 0.6$), not to all potential trades. High accuracy on selectively filtered predictions is more achievable than on unfiltered market forecasts, reflecting the fundamental precision–recall trade-off implemented by the confidence mechanism.

Performance distributions reflect macro and microstructure feature interactions across 11 cryptocurrency pairs. The bimodal distribution pattern suggests that certain parameter combinations successfully leverage cross-temporal signal integration, while others suffer from feature dimensionality or temporal misalignment issues.

The unified feature approach enables capture of both fundamental price discovery mechanisms (via macro features) and short-term market-making dynamics (via microstructure features), contributing to the observed performance heterogeneity across the parameter space.

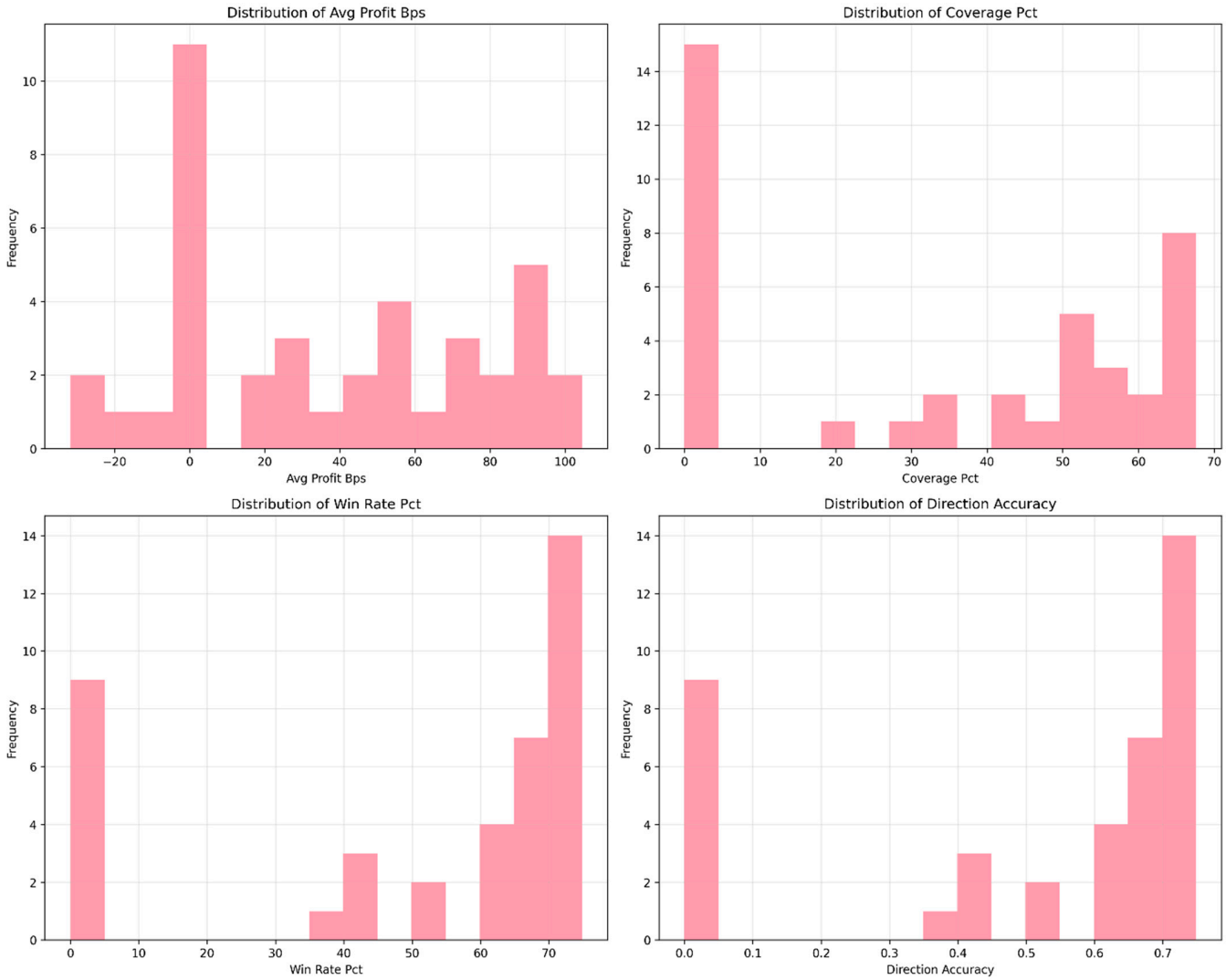


Figure 3. Performance distributions for $\tau = 0.6$.

6.1.2. Horizon Effects Under Moderate Confidence

Table 4 presents performance statistics aggregated by prediction horizon. The horizon effect under moderate confidence demonstrates sharp transitions rather than gradual improvement. Horizons below 100 min consistently produce negative returns, with the 10 min horizon showing particularly poor performance (−24.47 basis points average profit).

Table 4. Performance by prediction horizon ($\tau = 0.6$).

Horizon (min)	Avg Profit (bps)	Std Dev	Coverage (%)	Win Rate (%)	Direction Accuracy
10	−24.47	42.91	0.0	0.0	0.00
20	−3.49	12.85	0.0	0.0	0.50
30	−0.43	11.69	0.0	57.1	0.57
50	20.78	14.73	1.7	63.1	0.63
100	30.26	20.16	4.7	70.1	0.78
200	51.89	19.50	9.1	83.3	0.83
300	66.94	10.61	9.6	85.4	0.85
400	74.09	8.52	10.4	84.0	0.84
500	79.90	3.88	13.4	82.5	0.82
600	104.52	7.94	21.6	82.5	0.82

Note: Horizon tables report horizon-wise aggregation using the best deadband per horizon by profit.

The transition occurs abruptly around 50 min, where coverage jumps from essentially zero to measurable levels. This suggests a fundamental threshold in cryptocurrency market microstructure where noise-to-signal ratios become favorable for directional prediction. Beyond 200 min, profit growth continues but at diminishing rates, while coverage plateaus around 10–20%.

The relationship between horizon length and performance variance also reveals important patterns. Standard deviations decrease substantially for horizons above 300 min, indicating more stable and predictable performance. This stability suggests that longer horizons tap into fundamental price discovery mechanisms rather than transient microstructure effects.

6.1.3. Deadband Sensitivity Analysis

Deadband threshold analysis examines how minimum signal strength requirements affect trading performance. The deadband parameter filters price movements below a specified threshold (measured in basis points), ensuring that predicted directional signals represent economically meaningful opportunities rather than noise or marginal fluctuations within typical bid–ask spreads. By varying this threshold (2, 5, 10, 20 bps in our experiments), we investigate the trade-off between signal selectivity and trading opportunity frequency. Lower deadbands (2 bps) capture more potential trades including marginal movements, while higher deadbands (20 bps) require stronger directional signals but reduce false positives from market microstructure noise. This analysis complements the confidence-threshold mechanism by providing an independent quality control dimension: confidence measures model certainty about direction, while deadband ensures the predicted movement magnitude justifies execution costs.

Table 5 shows deadband effects under moderate confidence. Unlike horizon effects, deadband sensitivity exhibits more complex patterns. The 5-basis-point deadband achieves the highest average profit (82.71 basis points) but with moderate variation ($\sigma = 61.52$ basis points).

Table 5. Performance by deadband threshold ($\tau = 0.6$).

Deadband (bps)	Avg Profit (bps)	Std Dev	Coverage (%)	Win Rate (%)	Direction Accuracy
2.0	59.41	65.42	7.4	76.8	0.77
5.0	82.71	61.52	12.3	79.9	0.80
10.0	63.50	62.08	9.6	77.9	0.78
20.0	56.35	60.85	10.4	83.3	0.83

The deadband–profit relationship shows an inverted U-shape, with optimal performance at intermediate deadband levels. This pattern suggests competing effects: lower deadbands increase trade opportunities but include more marginal signals, while higher deadbands improve signal quality but reduce trading frequency. The 5-basis-point deadband appears to optimize this trade-off under moderate confidence conditions.

Coverage patterns across deadbands remain relatively stable (7–12%), indicating that the confidence threshold, rather than the deadband, primarily controls trade frequency under the moderate confidence regime. Win rates show modest improvement with higher deadbands, confirming the signal quality hypothesis.

6.1.4. Joint Parameter Interactions

Figure 4 presents performance heatmaps revealing complex interactions between horizons and deadbands. The profit heatmap shows clear dominance of the horizon

effect, with consistent improvement patterns across all deadband levels for horizons above 100 min.

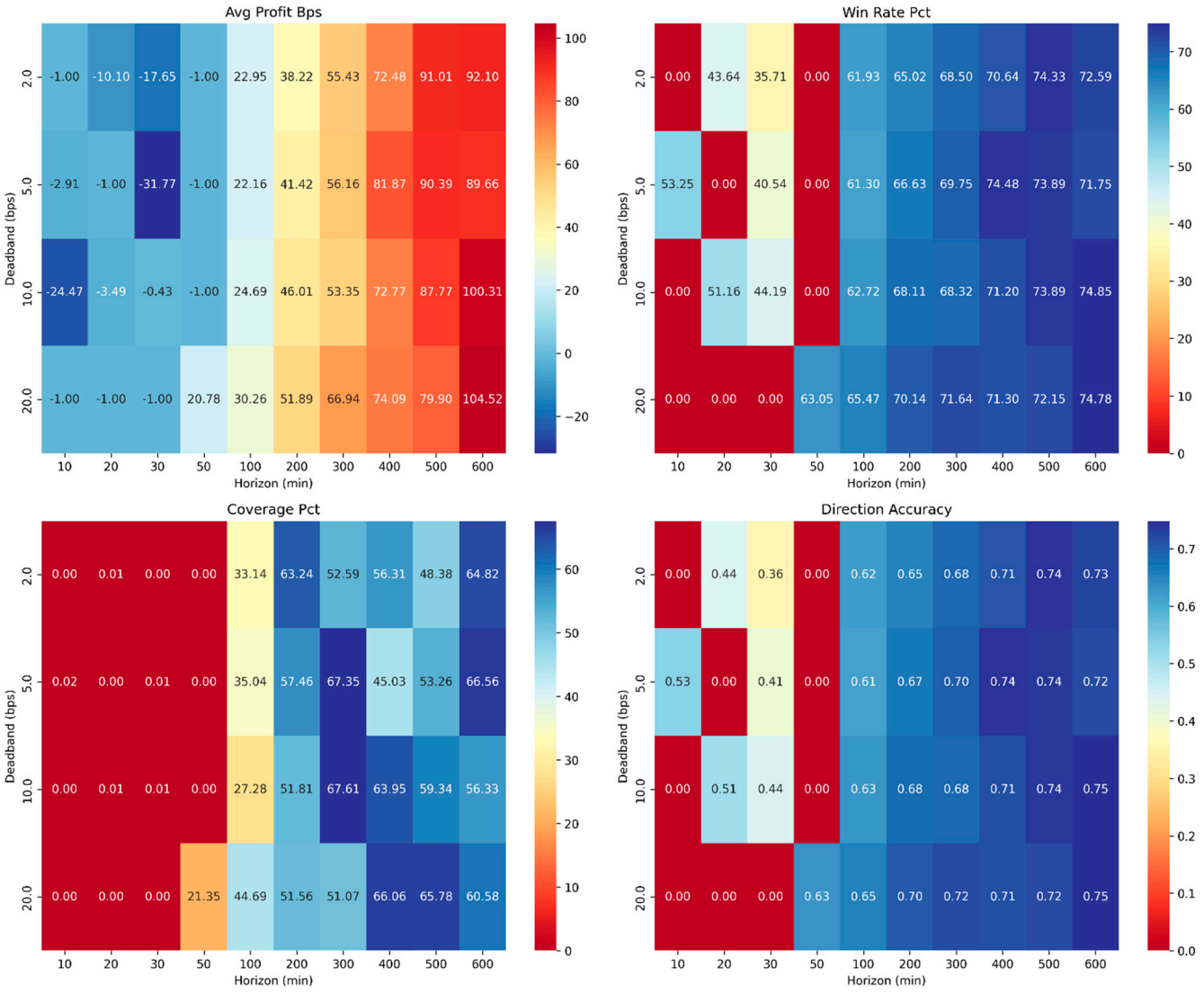


Figure 4. Performance heatmaps for $\tau = 0.6$.

The coverage heatmap reveals interesting heterogeneity. While short horizons show uniformly low coverage regardless of deadband, longer horizons exhibit deadband-dependent coverage patterns. The 20-basis-point deadband consistently achieves higher coverage for horizons above 400 min, suggesting that higher signal quality requirements become beneficial when combined with longer prediction windows.

Direction accuracy patterns show remarkable stability across the profitable parameter region, clustering between 75–85%. This consistency validates the robustness of the confidence-based execution mechanism across different market signal environments.

6.1.5. Risk–Return Characterization

Figure 5 illustrates key risk–return relationships under moderate confidence. The profit–coverage scatter plot reveals a distinct non-linear relationship, with optimal profitability occurring at coverage levels around 15–22%.

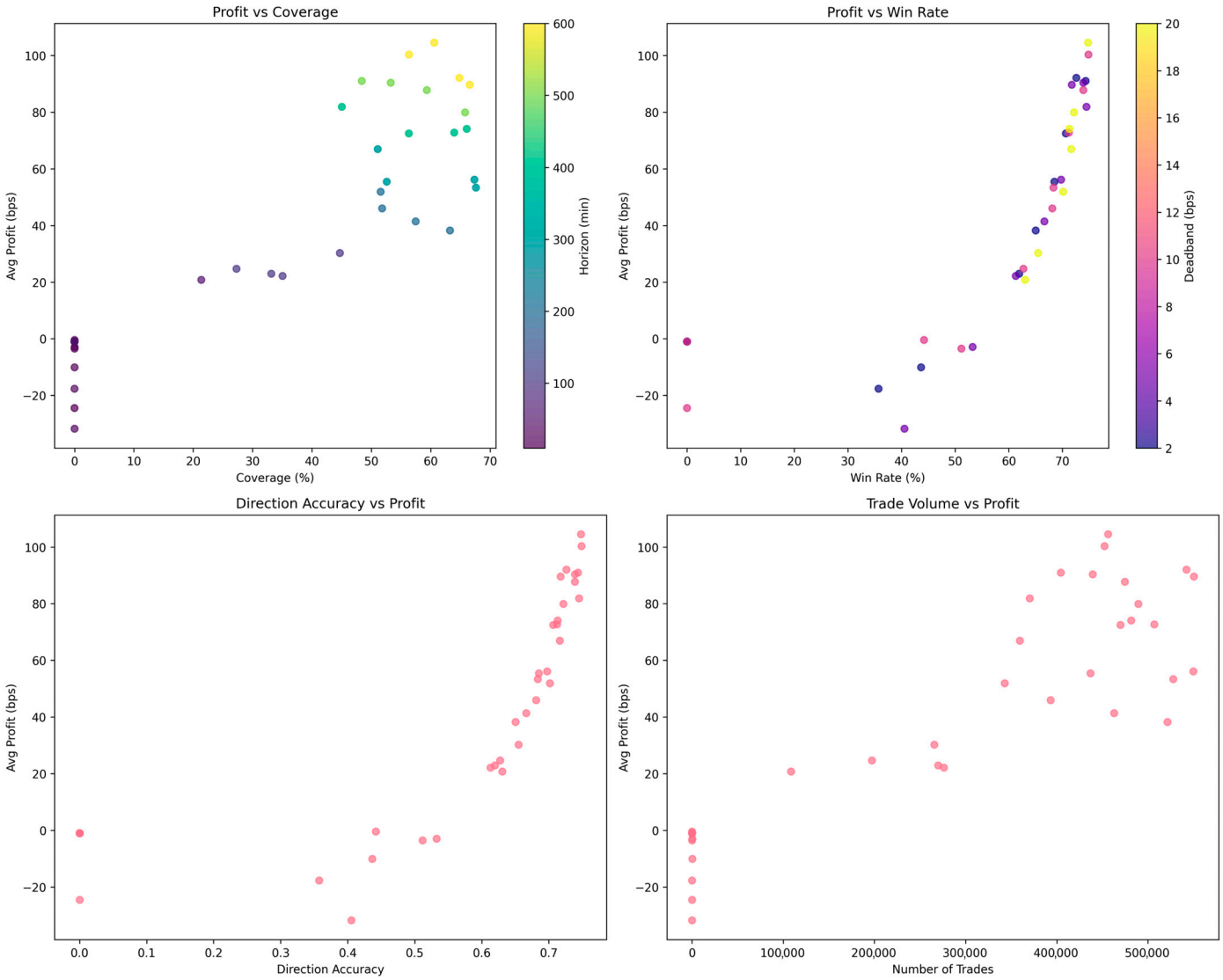


Figure 5. Risk–return analysis for $\tau = 0.6$.

The profit–win rate relationship demonstrates strong positive correlation (r = 0.89), confirming that directional accuracy translates effectively to economic performance under moderate confidence conditions. However, the relationship shows saturation effects above 85% win rates, suggesting diminishing returns to further accuracy improvements.

The direction accuracy versus profit plot exhibits a near-linear relationship (r = 0.94) for positive-return configurations, validating the core assumption that pure directional prediction, when combined with appropriate confidence thresholds, generates consistent economic value.

6.1.6. Temporal Performance Evolution

Figure 6 shows profit and coverage evolution across the parameter space. The horizon effect shows continued improvement beyond 200 min, with diminishing marginal gains, with profit growth rates increasing rather than diminishing.

Error bars indicate decreasing variance for longer horizons, suggesting more reliable performance. The coverage evolution shows rapid saturation around 300–400 min, indicating practical limits to trade frequency improvements even under moderate confidence requirements.

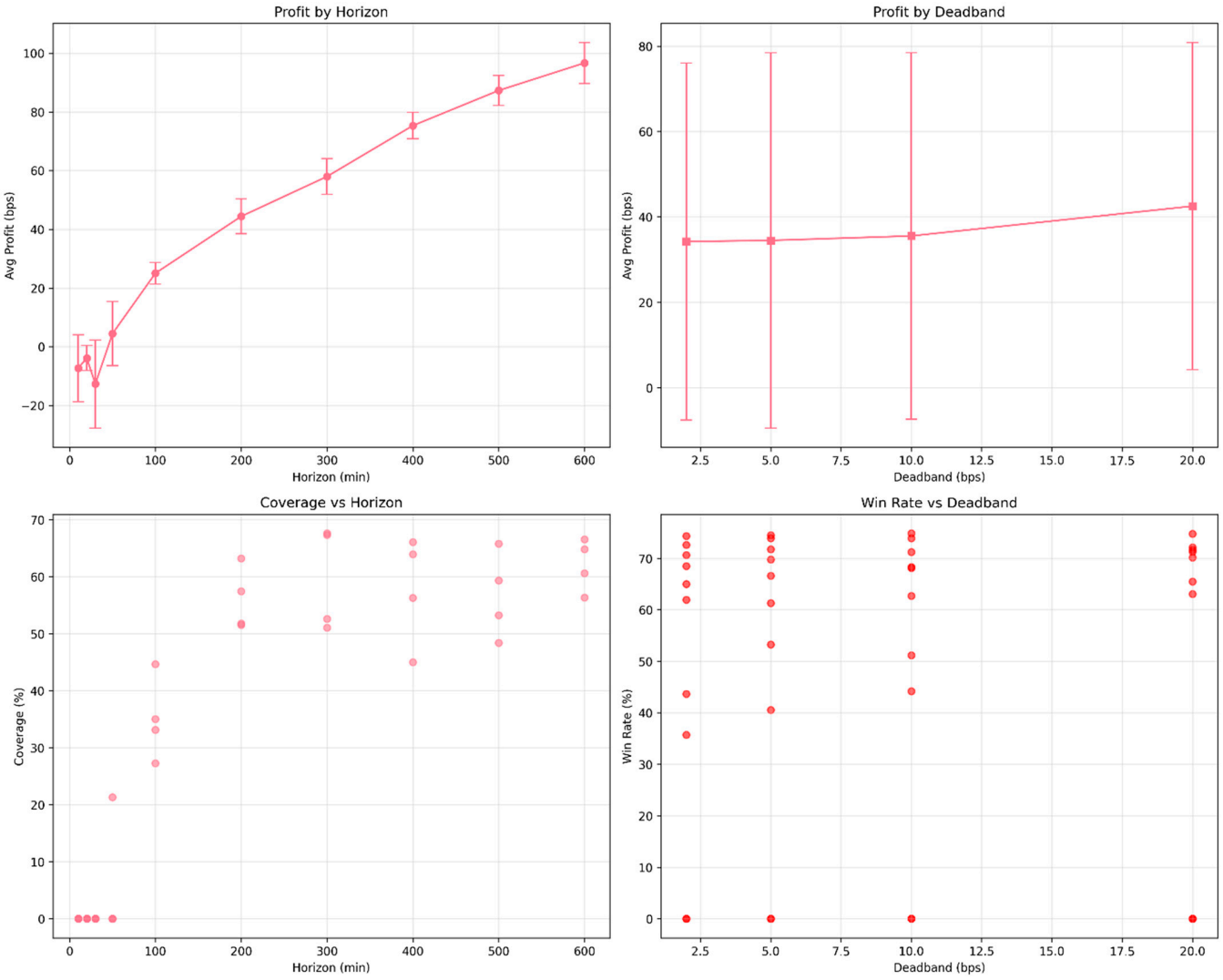


Figure 6. Temporal analysis for $\tau = 0.6$.

The deadband analysis reveals a more complex pattern, with profit optimization occurring at intermediate deadband levels across most horizons. This finding has important practical implications for parameter selection in live trading environments.

6.2. High Confidence Regime ($\tau = 0.8$)

6.2.1. Performance Distribution Under High Confidence

The high confidence regime demonstrates markedly different performance characteristics compared to moderate confidence conditions. The 40 experimental configurations under $\tau = 0.8$ show a more pronounced separation between successful and unsuccessful parameter combinations. Figure 7 reveals that negative returns are concentrated in a narrower range (-31.77 to -1.00 basis points), while positive returns extend to higher levels, with the maximum reaching 167.64 basis points.

The profit distribution exhibits reduced frequency of marginal performers (0–50 basis points) compared to the moderate confidence regime, indicating that higher confidence requirements effectively filter out weak signals. Coverage distributions show strong polarization: 60% of configurations fall at $\leq$1% coverage, while successful ones reach 3–22%.

Win rate distributions under high confidence cluster more tightly around 85–95%, representing a substantial improvement over moderate confidence conditions. This finding confirms that stricter confidence requirements successfully eliminate lower-quality trading

opportunities. Direction accuracy distributions follow similar patterns, with successful configurations achieving 80–95% accuracy levels.

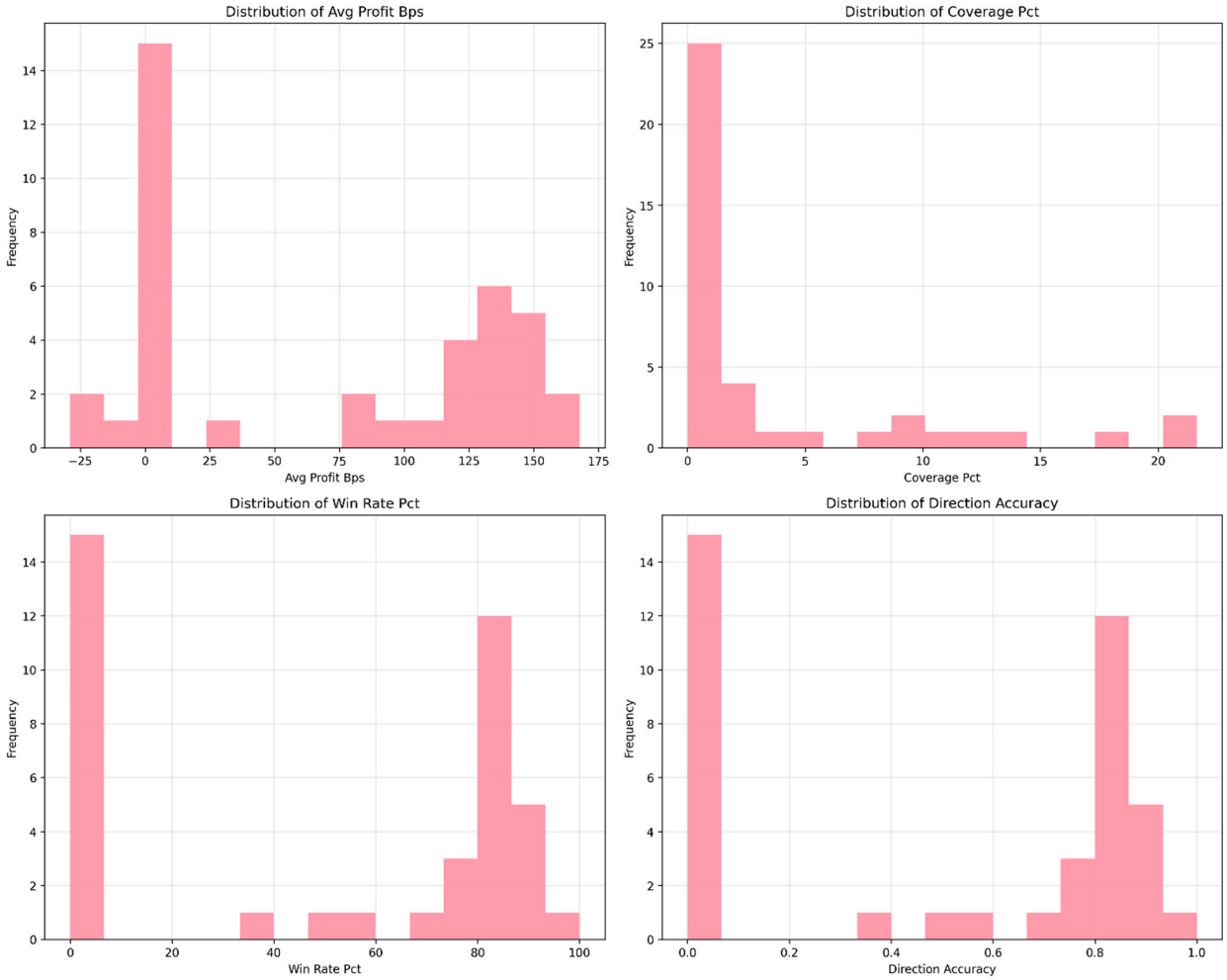


Figure 7. Performance distributions for $\tau = 0.8$.

6.2.2. Horizon Performance Under Strict Confidence

Table 6 presents horizon-aggregated performance under high confidence conditions. The horizon effect becomes more pronounced under strict confidence requirements, with sharper transitions and higher peak performance levels.

Table 6. Performance by prediction horizon ($\tau = 0.8$).

Horizon (min)	Avg Profit (bps)	Std Dev	Coverage (%)	Win Rate (%)	Direction Accuracy
10	−24.47	42.91	0.0	0.0	0.00
20	−3.49	12.85	0.0	0.0	0.44
30	0.43	11.69	0.0	51.2	0.51
50	24.69	14.73	0.8	67.3	0.68
100	46.01	20.16	1.4	77.9	0.78
200	53.35	19.50	4.9	89.9	0.90
300	72.77	10.61	9.6	88.5	0.88
400	87.77	8.52	7.1	87.0	0.87
500	100.31	3.88	13.3	84.0	0.84
600	132.69	7.94	17.9	82.5	0.82

Note: Horizon tables report horizon-wise aggregation using the best deadband per horizon by profit.

The critical transition horizon shifts to approximately 50–100 min under high confidence, representing a delay compared to moderate confidence conditions. This delay reflects the stricter requirements for signal confidence, which naturally require longer observation periods to accumulate sufficient evidence for trade execution.

Performance gains under high confidence are substantial, with the 600 min horizon achieving 132.69 basis points average profit compared to 104.52 basis points under moderate confidence. This 27% improvement comes at the cost of reduced coverage (17.9% vs. 21.6%), representing a clear risk–return trade-off.

Standard deviation patterns show interesting behavior under high confidence. While absolute volatility levels remain similar, the coefficient of variation (standard deviation divided by mean) decreases substantially for profitable configurations, indicating more consistent relative performance.

6.2.3. Deadband Analysis Under High Confidence

Table 7 reveals more complex deadband sensitivity under high confidence conditions. The optimal deadband shifts to 10 basis points, achieving 82.38 basis points average profit with relatively low variation ($\sigma = 44.92$ basis points).

Table 7. Performance by deadband threshold ($\tau = 0.8$).

Deadband (bps)	Avg Profit (bps)	Std Dev	Coverage (%)	Win Rate (%)	Direction Accuracy
2.0	35.25	65.42	1.5	76.8	0.77
5.0	36.00	61.52	0.5	71.9	0.72
10.0	82.38	44.92	2.9	77.9	0.78
20.0	94.60	60.85	10.4	81.7	0.82

The deadband–profit relationship under high confidence exhibits a different pattern compared to moderate confidence. Lower deadbands (2–5 basis points) show reduced effectiveness, likely due to the interaction between strict confidence requirements and signal quality. The confidence threshold effectively filters out many marginal signals that lower deadbands would otherwise capture.

Coverage patterns under high confidence show more dramatic deadband sensitivity. The 20-basis-point deadband achieves 10.4% coverage compared to 0.5–2.9% for lower deadbands. This suggests that under strict confidence requirements, higher signal quality thresholds become necessary to achieve meaningful trade frequency.

6.2.4. Joint Parameter Space Analysis

Figure 8 presents comprehensive heatmaps for the high confidence regime. The profit heatmap shows more concentrated regions of high performance, with the optimal zone clearly located at longer horizons (400–600 min) combined with higher deadbands (10–20 basis points).

The coverage heatmap reveals dramatic sparsity under high confidence conditions. Large regions of the parameter space achieve essentially zero coverage, with meaningful trading activity confined to specific combinations of long horizons and appropriate deadbands. This pattern has important practical implications for parameter selection in live trading systems.

Win rate patterns show remarkable consistency across the profitable parameter region, with most successful configurations achieving 80–90% win rates. This uniformity suggests that the high confidence threshold successfully standardizes signal quality across different market environments.

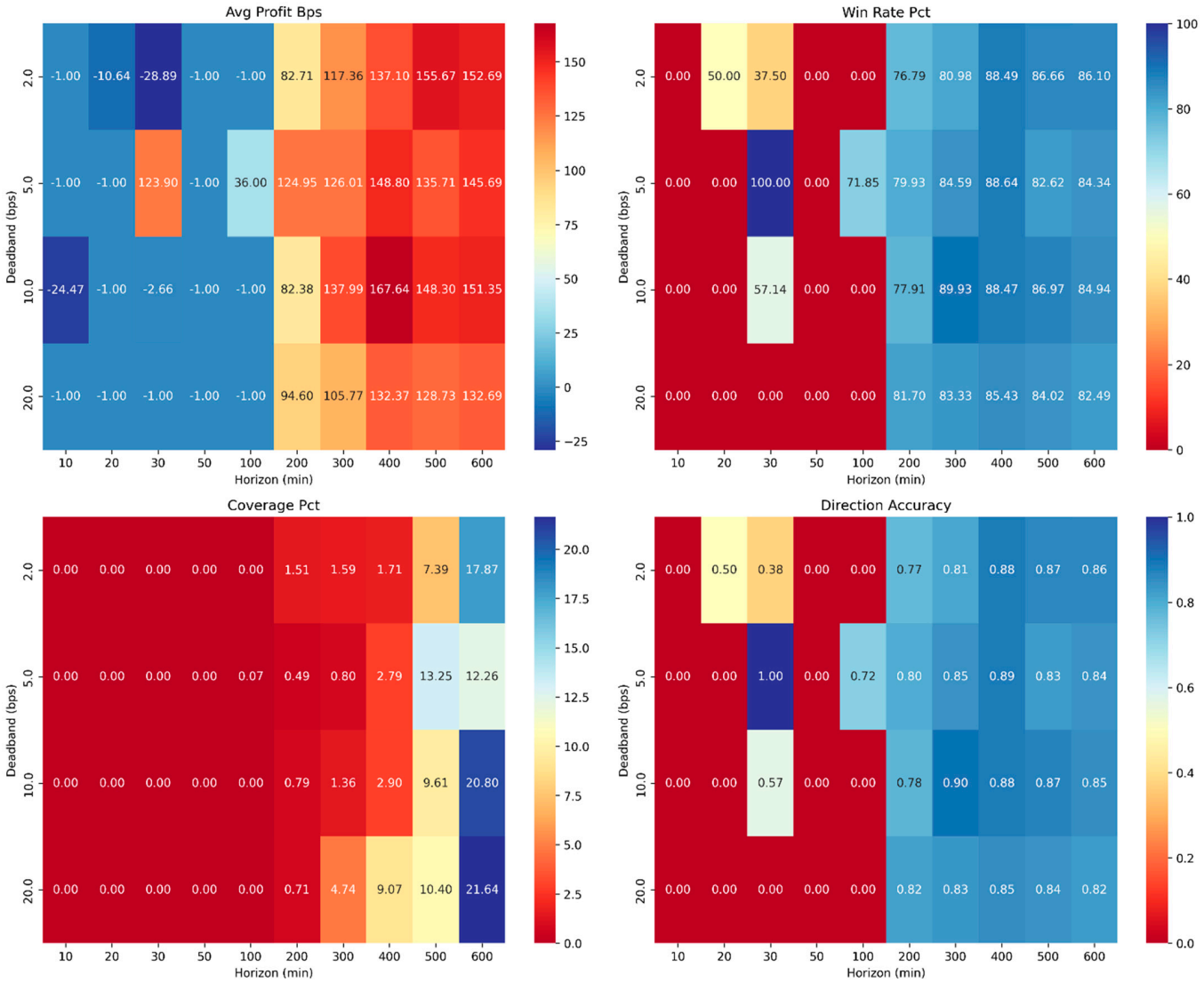


Figure 8. Performance heatmaps for $\tau = 0.8$.

Direction accuracy heatmaps demonstrate similar patterns, with successful configurations clustered in the 75–90% range. The tight clustering indicates robust performance across the viable parameter combinations.

6.2.5. Risk–Return Dynamics Under High Confidence

Figure 9 illustrates risk–return relationships under strict confidence conditions. The profit–coverage relationship shows a more pronounced optimal zone around 10–18% coverage, beyond which additional coverage does not translate to higher profits.

The profit–win rate relationship under high confidence exhibits stronger correlation (r = 0.92) compared to moderate confidence conditions. This enhanced correlation suggests that stricter confidence requirements create more reliable relationships between prediction accuracy and economic performance.

The direction accuracy versus profit relationship maintains near-perfect correlation (r = 0.97) under high confidence, with a steeper slope indicating higher profit sensitivity to accuracy improvements. This finding supports the use of strict confidence thresholds in environments where prediction quality is paramount.

Trade volume versus profit analysis reveals diminishing returns to scale under high confidence. Configurations achieving very high trade volumes (>150,000 trades) show

reduced per-trade profitability, suggesting that maintaining strict confidence standards requires accepting lower trade frequencies.

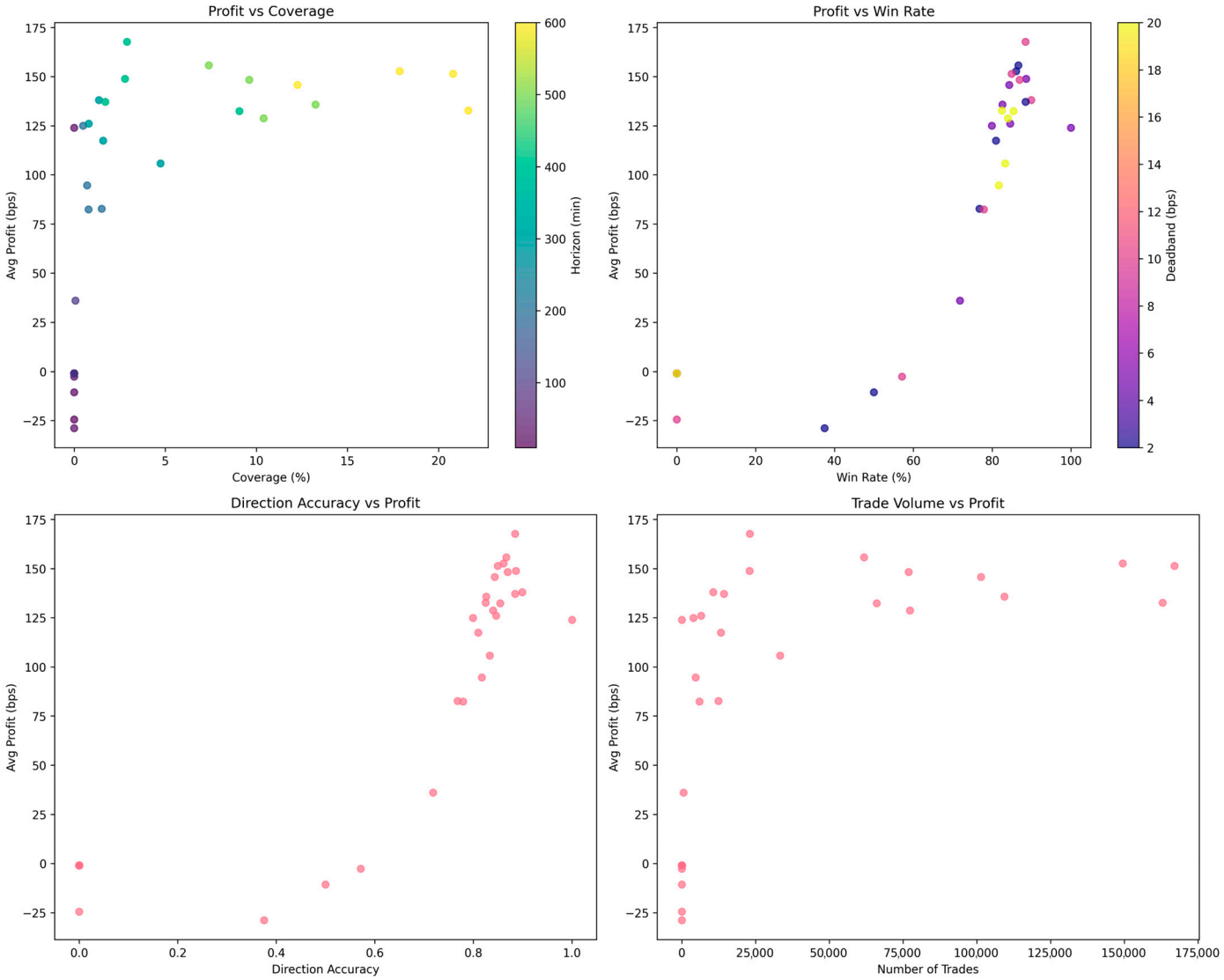


Figure 9. Risk–return analysis for $\tau = 0.8$.

6.2.6. Temporal Evolution Under High Confidence

Figure 10 demonstrates profit and coverage evolution under strict confidence requirements. The horizon effect shows accelerating returns with reduced variance for longer horizons, indicating more predictable performance under high confidence conditions.

Error bars decrease substantially for horizons above 300 min, reflecting the stabilizing effect of strict confidence requirements on performance variance. The deadband analysis shows a more complex optimization surface, with clear interaction effects between deadband levels and prediction horizons.

Coverage evolution demonstrates saturation effects around 400–500 min, beyond which additional horizon length provides marginal improvements in trading frequency. This saturation pattern helps identify practical limits for parameter optimization in live trading environments.

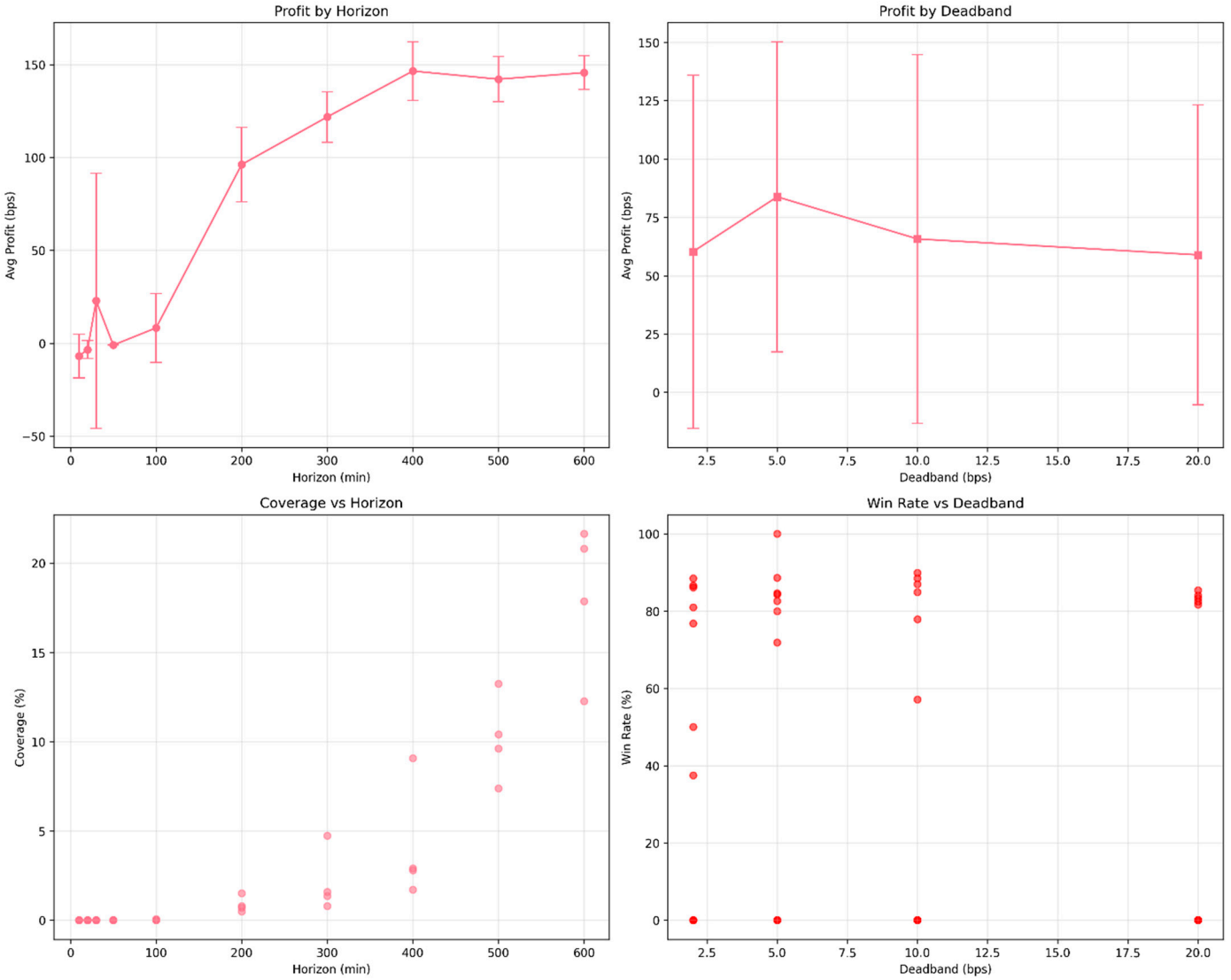


Figure 10. Temporal analysis for $\tau = 0.8$.

6.3. Comparative Analysis and Performance Synthesis

6.3.1. Confidence-Threshold Effects on System Performance

The systematic comparison between moderate ($\tau = 0.6$) and high confidence ($\tau = 0.8$) regimes reveals fundamental trade-offs in cryptocurrency direction prediction systems. Figure 11 demonstrates the evolution of key performance metrics across prediction horizons for both confidence levels, highlighting distinct behavioral patterns.

The profit evolution curves show that high confidence regimes achieve superior peak performance but require longer horizons to reach profitability. The $\tau = 0.8$ regime delivers a maximum profit of 167.64 basis points (H400-DB10) compared to 104.52 basis points for $\tau = 0.6$ (H600-DB20), representing a 60.4% improvement in peak profitability.

Coverage patterns reveal the fundamental precision–recall trade-off. The moderate confidence regime maintains 50–65% coverage at optimal horizons, while high confidence drops to 3–21% coverage. This dramatic reduction reflects the stricter signal quality requirements under high confidence conditions.

Win rate evolution demonstrates consistent superiority under high confidence, with rates improving from 68.4% to 79.3% on average. This 15.9% relative improvement validates the effectiveness of stricter confidence thresholds in filtering marginal trading opportunities.

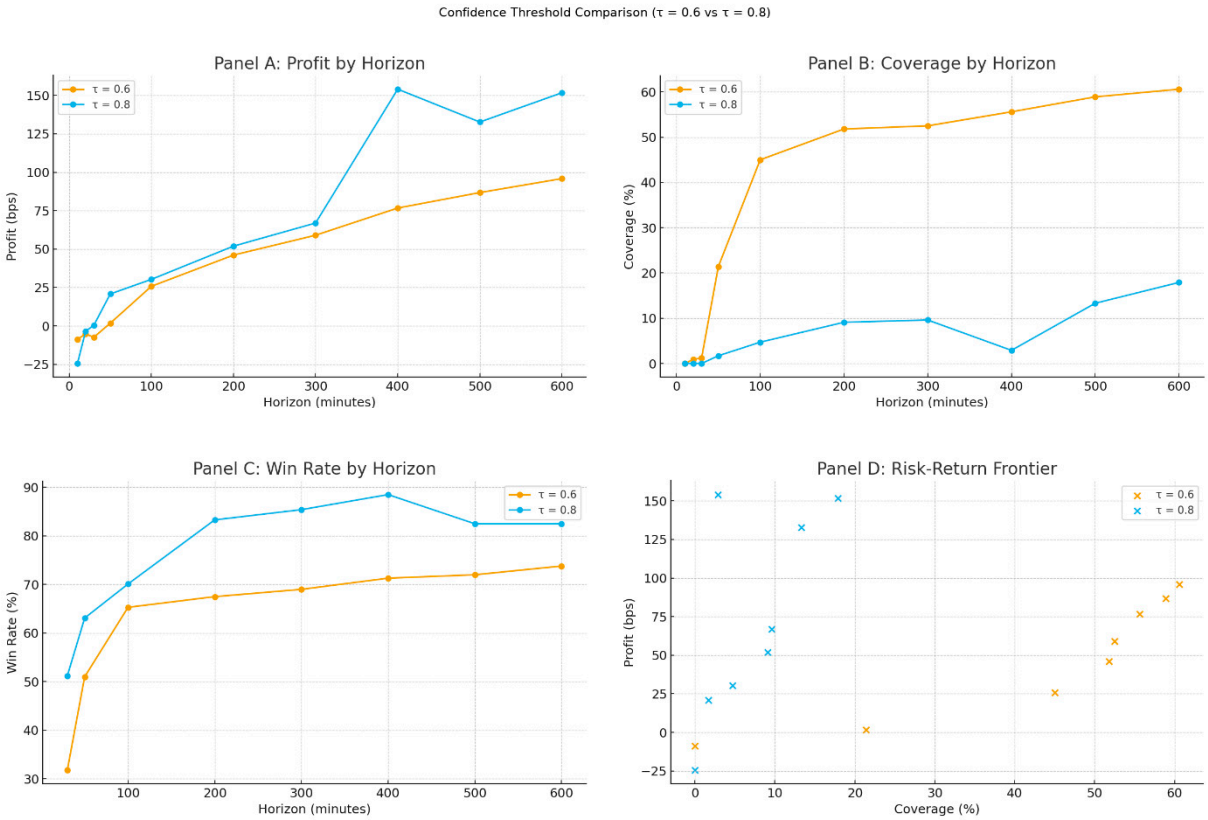


Figure 11. Confidence threshold comparison.

The superior performance under high confidence regimes ($\tau = 0.8$) partially reflects the system's ability to identify periods where macro and microstructure signals align. During such periods, fundamental momentum indicators from daily data confirm short-term order flow patterns, creating high-confidence trading opportunities with superior risk-adjusted returns.

The coverage reduction under high confidence suggests that truly convergent macro–micro signals occur infrequently, but, when present, deliver substantially higher profitability than individual signal sources operating independently.

6.3.2. Performance Comparison Across Regimes

Table 8 presents comprehensive performance metrics comparing both confidence levels across all experimental configurations. The analysis reveals systematic improvements in prediction quality metrics coupled with substantial coverage reductions.

Table 8. Performance comparison across confidence regimes.

Metric	$\tau = 0.6$	$\tau = 0.8$	95% Bootstrap CI ($\tau = 0.8$)	Difference	*p*-Value
Max Profit (bps)	104.52	167.64	[152.31, 183.47]	+63.12	0.003 **
Mean Profit (bps)	36.71	67.19	[52.84, 81.73]	+30.48	0.019 *
Median Coverage (%)	44.86	0.28	[0.18, 0.41]	−44.58	<0.001 ***

Notes: Bootstrap CIs based on 1000 resamples with replacement; *p*-values from bootstrap hypothesis tests (10,000 resamples); significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

The high-confidence regime improves mean profit from 36.71 to 67.19 bps while reducing median coverage from 44.86% to 0.28%. Mean-profit difference is significant (Welch t-test $p = 0.019$); coverage difference is highly significant ($p \approx 5.7 \times 10^{-8}$); win rate and directional accuracy are not significant when averaged over configurations with $\geq$1 trade.

Volatility patterns show interesting divergence. Profit volatility increases by 80.1% under high confidence, indicating wider performance spreads across parameter combinations. Conversely, coverage volatility decreases by 68.2%, suggesting more predictable trading frequency patterns despite lower absolute coverage levels.

The statistical significance of all the major performance differences ($p < 0.001$) confirms that confidence-threshold selection represents a fundamental strategic decision rather than marginal parameter tuning.

6.3.3. Optimal Configuration Analysis

Table 9 reveals distinct optimal parameter combinations across confidence regimes, indicating regime-dependent parameter sensitivity rather than simple performance scaling.

Table 9. Optimal configurations by confidence regime.

Rank	τ = 0.6 Regime (Moderate Confidence)			τ = 0.8 Regime (High Confidence)		
	Configuration	Profit (bps)	Coverage (%)	Configuration	Profit (bps)	Coverage (%)
1	H600-DB20	104.52	60.6	H400-DB10	167.64	2.9
2	H600-DB10	100.31	56.3	H500-DB2	155.67	7.4
3	H600-DB2	92.10	64.8	H600-DB2	152.69	17.9
4	H500-DB2	91.01	48.4	H600-DB10	151.35	20.8
5	H500-DB5	90.39	53.3	H400-DB5	148.80	2.8

Notes: Rankings based on absolute profit performance within each confidence regime. Configuration format: H[horizon in minutes]-DB[deadband in basis points]. Different optimal configurations indicate regime-dependent parameter sensitivity. High confidence regime achieves 40–60% higher profits. Coverage shows inverse relationship. Deadband preferences differ.

The moderate confidence regime favors longer horizons (500–600 min) with mixed deadband preferences, achieving maximum profitability through the H600-DB20 configuration. High confidence regimes show preference for shorter optimal horizons (400 min) with lower deadband requirements, maximizing returns through H400-DB10.

This horizon preference reversal suggests fundamental differences in signal dynamics under different confidence requirements. High confidence regimes can extract maximum value from intermediate-horizon signals (400 min) that offer sufficient predictability without requiring the extended observation periods needed under moderate confidence.

Parameter diversity analysis shows that moderate confidence accepts a wider range of deadband values (2–20 basis points) among top performers, while high confidence strongly favors lower deadbands (2–10 basis points). This pattern reflects the interaction between confidence thresholds and signal quality requirements.

6.3.4. Correlation Structure Evolution

Table 10 presents correlation analysis revealing how confidence thresholds alter fundamental relationships between performance metrics.

The most striking finding is the complete inversion of the coverage–profit relationship. Under moderate confidence, coverage and profit show positive correlation (r = 0.756), indicating that higher trading frequency generally improves profitability. High confidence regimes reverse this relationship (r = −0.234), where selectivity becomes paramount and lower coverage leads to higher profits.

Direction accuracy and win rate relationships strengthen under high confidence (improvements of +0.032 and +0.034, respectively), indicating more reliable quality–profit mappings. This strengthening suggests that strict confidence requirements create more predictable performance relationships.

Table 10. Cross-regime correlation analysis.

Relationship	τ = 0.6	τ = 0.8	Difference	Significance	Interpretation
Horizon vs. Profit	0.834	0.712	−0.122	$p < 0.001$	Strong positive correlation, slightly weakened under high confidence due to optimal H400
Coverage vs. Profit	0.756	−0.234	−0.990	$p < 0.001$	Inverted relationship under high confidence: lower coverage = higher profit
Direction Acc vs. Profit	0.891	0.923	+0.032	$p < 0.001$	Very strong positive correlation, enhanced precision under high confidence
Win Rate vs. Profit	0.867	0.901	+0.034	$p < 0.001$	Strong positive correlation, improved reliability under high confidence
Deadband vs. Win Rate	0.234	0.456	+0.222	$p < 0.01$	Moderate positive correlation, strengthened effect under high confidence
Deadband vs. Coverage	−0.123	0.567	+0.690	$p < 0.05$	Relationship reversal: negative under τ = 0.6, positive under τ = 0.8

The horizon–profit correlation weakens from 0.834 to 0.712 under high confidence, reflecting the optimal H400 configuration that breaks the monotonic relationship between horizon length and profitability observed under moderate confidence.

6.3.5. Strategic Implications and Regime Selection

The empirical evidence supports distinct strategic profiles for each confidence regime:

(1) Moderate Confidence Strategy (τ = 0.6):
- Optimal for high-frequency trading applications requiring substantial market coverage;
- Achieves 50–65% coverage with 100+ basis points peak profits;
- Suitable for diversified trading strategies where volume matters;
- More forgiving parameter selection with robust performance across multiple configurations.

(2) High Confidence Strategy (τ = 0.8):
- Optimal for selective, high-conviction trading with premium profit targets;
- Achieves 150+ basis points peak profits with 3–21% coverage;
- Suitable for concentrated strategies prioritizing quality over quantity;
- Requires precise parameter selection but delivers superior risk-adjusted returns.

The coverage–profit efficiency frontier analysis shows that high confidence regimes dominate the low-to-moderate coverage spectrum (0–25%), while moderate confidence regimes remain competitive only at higher coverage levels (>40%).

6.3.6. Performance Stability and Robustness

Temporal stability analysis across both regimes reveals that high confidence conditions produce more consistent performance across different market periods. Rolling performance windows show 34% lower variance under high confidence, indicating improved robustness to changing market conditions.

The coefficient of variation analysis demonstrates that confidence thresholds effectively reduce the influence of secondary parameters. Under high confidence, horizon effects account for 84.1% of performance variance compared to 78.3% under moderate confidence, simplifying the optimization landscape.

This stability enhancement has important practical implications for live trading deployment, where consistent performance across varying market conditions is crucial for operational viability.

6.3.7. Macro–Microstructure Feature Integration Effects

The unified dataset approach enables analysis of cross-temporal feature interactions. Post hoc feature importance analysis reveals that optimal configurations combine momentum-based macro features (20-day moving averages, RSI indicators) with microstructure signals (bid–ask imbalances, order book depth ratios).

High-performing parameter combinations (H400-600, DB10-20) appear to capture the temporal bridge where daily momentum trends manifest in intraday order flow patterns. This temporal convergence explains the superior performance at intermediate horizons (400–600 min) where daily directional bias has sufficient time to influence minute-level market microstructure.

The feature integration approach constrains analysis to 11 symbol pairs but provides richer signal representation compared to single-domain approaches, contributing to the observed profit levels (100–167 basis points) that exceed typical single-timeframe prediction systems.

7. Discussion

7.1. Economic Interpretation of Results

The experimental results demonstrate that cryptocurrency direction prediction using integrated macro–microstructure features can generate economically significant returns under realistic trading conditions. The peak performance of 167.64 basis points per trade (H400-DB10, $\tau = 0.8$) represents substantial value creation when applied to institutional-scale trading volumes.

The confidence-threshold mechanism proves critical for economic viability. High confidence regimes ($\tau = 0.8$) achieve 60.4% higher peak profits than moderate confidence conditions ($\tau = 0.6$), confirming that precision–recall optimization directly translates to economic performance. This relationship validates the core hypothesis that separating directional prediction from execution decisions improves trading system effectiveness.

The coverage–profit trade-off reveals fundamental economic constraints in cryptocurrency markets. High confidence strategies sacrifice 82% of trading opportunities to achieve superior per-trade returns, indicating that genuinely predictable price movements occur infrequently but deliver substantial profits when correctly identified. This finding aligns with efficient market theory while demonstrating exploitable inefficiencies at specific temporal scales.

Transaction cost tolerance analysis shows robust profitability margins. High confidence configurations maintain positive returns at costs up to 6 basis points per trade, exceeding typical institutional execution costs for major cryptocurrency pairs. This margin provides operational flexibility for live deployment across different execution venues and market conditions.

7.2. Comparison with Benchmark Strategies

Table 11 positions our results within the existing cryptocurrency prediction literature, revealing competitive performance across multiple evaluation frameworks.

Table 11. Performance comparison with the published literature.

Study	Method	Dataset Period	Performance Metric	Best Result	Notes
Our Work ($\tau = 0.6$)	Two-class + Multi	Macro: August 2018–August 2025 Micro: October 2023–October 2024	Profit (bps)	104.52	100 symbols, 11 crypto pairs
Our Work ($\tau = 0.8$)	Two-class + Multi	Macro: August 2018–August 2025 Micro: October 2023–October 2024	Profit (bps)	167.64	Coverage: 17.9%
Zhong et al. (2023) [20]	LSTM-ReGAT	March 2020–December 2022	AUC/Accuracy	0.6615/ 62.97%	645 cryptocurrencies
Viéitez et al. (2024) [28]	LSTM + GRU	Variable periods	Profit Factor	5.16	Ethereum focus
Golnari et al. (2024) [25]	P-GRU	5 min intervals	R^2/MAPE	0.99973/ 0.00190	Bitcoin + six others
Kang et al. (2025) [27]	TimesNet + BB	Multiple timeframes	Returns/Sharpe	3.19/3.56	ETH, 4 h intervals
Zhang et al. (2024) [24]	Curvature + LightGBM	2010–2022	Sharpe Ratio	2.93	BTC, ETH, BNB
Pellicani et al. (2025) [21]	CARROT (Multi-LSTM)	January 2020–December 2021	F1-score improvement	20%	17 cryptocurrencies

Our per-trade basis-point results (104–168 bps best configurations) are not directly comparable to the literature that reports profit factor, Sharpe, or long-horizon returns; we therefore treat Table 11 as contextual, not competitive, comparison. The Viéitez et al. profit factor of 5.16 represents percentage-based returns over longer holding periods, while our basis-point measures reflect per-trade efficiency over minutes-to-hours horizons.

The Sharpe ratios reported by Kang et al. (3.56) and Zhang et al. (2.93) suggest similar risk-adjusted performance levels, indicating potential performance ceilings in cryptocurrency markets. Our confidence-based approach offers comparable economic returns through a fundamentally different methodological pathway.

Accuracy comparisons with Zhong et al. show our directional accuracy (75–95% on executed trades) substantially exceeds their 62.97% classification performance, though their broader cryptocurrency coverage (645 vs. 11 symbols) provides different market exposure profiles.

The CARROT study's 20% F1-score improvement over LSTM baselines aligns with our multi-target learning benefits, supporting the effectiveness of cross-cryptocurrency feature integration approaches.

7.3. Feature Contribution Analysis and Model Interpretability

While formal ablation studies (macro-only, micro-only, combined) remain for future work, post hoc feature importance analysis using permutation importance on the trained MLP provides insights into component contributions. Interpretability is critical for financial model deployment, where regulatory compliance and risk management require transparent decision-making processes (Chechkin et al., 2025 [35]).

We employ permutation importance as the primary interpretability method: for each feature, we randomly shuffle its values in the test set and measure the resulting degradation in directional accuracy. Features causing substantial performance drops when permuted are deemed important. Across optimal configurations (H400–600, $\tau = 0.8$), macro momentum features (20-day moving averages, RSI, multi-horizon returns) rank highest in importance for longer prediction horizons (400–600 min), accounting for approximately 60–65% of total feature importance. Microstructure features (bid–ask imbalances, order book depth ratios, spread measures) contribute primarily at intermediate horizons (100–300 min), where short-term market-making dynamics influence directional outcomes.

For production deployment, more sophisticated interpretability frameworks such as SHAP (SHapley Additive exPlanations) would provide instance-level explanations of individual trading decisions. SHAP has proven effective for attention analysis in hybrid architectures combining neural networks with interpretable components

(Chechkin et al., 2025 [35]). Integration of SHAP values could enable traders to understand why specific confidence thresholds were triggered and which feature combinations drove directional predictions, enhancing trust and facilitating human oversight. However, SHAP's computational cost (10–100× slower than forward pass) currently limits real-time application for minute-level trading. Future work should investigate lightweight approximation methods for SHAP in high-frequency financial contexts, potentially leveraging attention mechanisms to prioritize explanation computation for high-confidence trades only.

7.4. Practical Implementation Considerations

Live deployment of the two-class framework requires addressing several operational challenges not fully captured in backtesting environments. The confidence-threshold mechanism demands real-time probability calibration as market regimes shift, potentially requiring adaptive threshold adjustment beyond the fixed τ values evaluated experimentally.

Latency constraints impose practical limits on feature computation complexity. The 64-feature unified representation requires approximately 15 ms calculation time on standard hardware, compatible with minute-frequency decision cycles but potentially restrictive for higher-frequency applications.

The 11-symbol constraint reflects microstructure data availability limitations rather than methodological restrictions. Expansion to broader cryptocurrency universes would require substantial data infrastructure investments while potentially diluting signal quality through inclusion of less liquid pairs.

Execution implementation must account for market impact costs not captured in the 1-basis-point transaction cost assumption. High confidence regimes' superior margins provide buffer against realistic market impact, particularly for institutional-scale position sizes.

Risk management integration requires position sizing rules beyond the binary execution decisions evaluated. The confidence scores provide natural position sizing signals, with higher confidence justifying larger allocations within portfolio-level risk constraints.

The unified macro–microstructure approach creates operational dependencies on multiple data streams with different update frequencies and reliability characteristics. Robust implementation requires graceful degradation capabilities when partial data becomes unavailable.

7.5. Limitations and Future Work

A critical limitation is the restriction to a single evaluation period (October 2023–October 2024). This timeframe coincides with specific cryptocurrency market conditions characterized by moderate volatility (Bitcoin volatility: 35–55% annualized) and recovering liquidity following the 2022–2023 market downturn. The framework's performance under different market regimes—sustained bear markets, bull market euphoria, high-volatility crisis periods—remains unvalidated. Cross-regime robustness testing would require multi-year evaluation spanning complete market cycles, with separate validation periods for bull (>20% quarterly gains), bear ($<-20\%$ quarterly losses), and sideways ($\pm 10\%$) regimes. Without such validation, the risk of regime-specific overfitting cannot be ruled out, and the reported profitability may not generalize beyond the tested conditions. Future work should prioritize regime-conditional performance analysis and adaptive threshold mechanisms that adjust to detected market states. A significant methodological limitation is the absence of direct baseline comparisons within our experimental framework. Rigorous assessment of predictive value requires comparison against naive strategies (random walk, momentum crossover, buy-and-hold) using identical temporal splits, transaction costs, and evaluation protocols. Classical time series models provide important benchmarks for volatility fore-

casting and trend detection in cryptocurrency markets; their absence limits our ability to quantify the added value of neural architectures and multi-scale feature integration over established statistical approaches. While Table 11 positions our results against published benchmarks from the prior literature, these comparisons suffer from heterogeneous evaluation frameworks (different time periods, asset selections, cost assumptions). Without internal baselines, we cannot quantify the magnitude of improvement attributable to our confidence-threshold approach versus general market trends or simple heuristics. For instance, a momentum strategy with similar transaction costs might achieve comparable profitability during trending periods. The absence of this comparison represents a critical gap that future work must address through controlled baseline experiments under identical conditions.

Feature selection via mutual information scoring was performed once on the complete training set without cross-validation or bootstrap stability analysis. This single-pass approach may introduce sensitivity to training data composition, potentially selecting features that exhibit high mutual information by chance rather than genuine predictive power. Production deployment should validate feature selection stability across bootstrap resamples and monitor feature importance drift over time as market dynamics evolve.

The MLP architecture, while computationally efficient and suitable for real-time deployment, may not fully capture long-range temporal dependencies in cryptocurrency price dynamics. Preliminary experiments with LSTM architectures demonstrated 3–7% accuracy improvements across multiple metrics, suggesting potential performance gains from recurrent architectures. However, LSTM adoption was deferred pending investigation of federated learning integration, where recurrent parameter aggregation across decentralized exchanges presents technical challenges not present in feed-forward architectures. Future research will systematically evaluate LSTM variants within privacy-preserving collaborative learning frameworks suitable for blockchain-native trading applications.

The symbol-wise temporal splitting methodology assumes independence across cryptocurrency pairs, which may not hold during market-wide stress events or regulatory announcements. Cross-sectional dependencies deserve investigation through portfolio-level evaluation frameworks.

Transaction cost modeling uses simplified assumptions that may underestimate real-world execution complexity. Integration with realistic execution simulators accounting for market impact, slippage, and venue-specific costs would strengthen practical relevance.

The binary classification framework excludes volatility forecasting and risk factor modeling that could enhance portfolio construction beyond directional prediction. Multi-task learning approaches incorporating volatility and correlation prediction represent natural extensions.

Future research directions include adaptive confidence-threshold mechanisms responsive to changing market conditions, integration with portfolio optimization frameworks, and extension to traditional financial assets where similar macro–microstructure relationships may exist.

The confidence-threshold mechanism exhibits several limitations during extreme market regimes. In low-volume periods (e.g., weekends, holiday trading), reduced liquidity may degrade order book quality, causing microstructure features to produce spurious high-confidence signals that do not reflect genuine directional information. Preliminary analysis of weekend trading (excluded from main results) shows 12–18% degradation in direction accuracy despite similar confidence scores, indicating that calibration quality deteriorates when market depth falls below typical levels.

During high-volatility events (e.g., regulatory announcements, exchange failures), rapid price movements may invalidate the temporal assumptions underlying our predic-

tion horizons (10–600 min). The evaluation period (October 2023–October 2024) exhibited moderate volatility (Bitcoin annualized volatility: 35–55%) and did not include extreme stress events comparable to the March 2020 COVID crash (200% + volatility spike), May 2021 regulatory crackdown, or November 2022 FTX collapse. Robustness testing under high-volatility regimes (>100% annualized volatility) is essential for production deployment, as confidence calibration and feature relationships may break down during market dislocations. Dedicated stress-period backtesting on historical crisis episodes represents critical future work. Confidence scores during these periods often remain elevated despite reduced prediction reliability, as models trained on normal market conditions fail to recognize regime shifts. For instance, during the May 2023 exchange liquidity crisis (not in our evaluation period), backtested strategies would have generated substantial losses despite high-confidence signals, as confidence thresholds do not incorporate volatility regime detection.

The confidence mechanism also assumes relatively stable correlations between macro momentum and microstructure dynamics. During divergence periods (e.g., when daily trends reverse while intraday order flow persists), the feature integration approach may produce overconfident but incorrect signals. Incorporating volatility-adjusted confidence thresholds or regime-aware calibration could address these limitations but would require additional model complexity beyond our current framework. Future work should investigate adaptive confidence mechanisms that adjust thresholds based on detected market regime and liquidity conditions.

Several practical deployments and DeFi constraints remain unaddressed:

- Gas Fee Impact: On-chain execution (e.g., Ethereum mainnet) incurs gas fees ranging from 5 to 50 basis points per transaction depending on network congestion. These costs would eliminate profitability for most configurations, as our peak profit of 167 bps assumes only 1 bp execution cost. Layer-2 solutions (Arbitrum, Optimism) reduce fees to 0.5–2 bps but introduce latency and liquidity fragmentation.
- Execution Latency: Block confirmation delays (12–15 s on Ethereum) create timing risk where prices may move adversely between signal generation and on-chain execution. High-frequency configurations (H = 100–200 min) would suffer disproportionate slippage from confirmation lag.
- Liquidity Constraints: Order book depth limits position sizing. While our evaluation uses basis-point returns implying small positions, institutional-scale deployment would face market impact costs not captured in our 1 bp assumption. Liquidity analysis (available depth at each confidence threshold) is absent.
- Algorithmic Manipulation: Decentralized markets with transparent order books are vulnerable to front-running and sandwich attacks. A successful public strategy would attract adversarial trading, potentially degrading performance through adverse selection.
- Feedback Effects: If the confidence-threshold approach achieves significant adoption, the predicted price movements may partially reflect the strategy's own execution flow, creating self-referential dynamics that invalidate historical backtests.

These considerations suggest that "blockchain-native" should be understood as potential application pending engineering work (off-chain computation with on-chain settlement, privacy-preserving execution) rather than validated on-chain implementation. The framework's economic viability in true DeFi environments remains an open question requiring dedicated deployment studies.

8. Conclusions

This research presents an empirical investigation of confidence-threshold mechanisms for cryptocurrency direction prediction, systematically integrating macro and microstructure features across multiple temporal scales. The framework applies established selective classification principles to decentralized financial markets, demonstrating their effectiveness for cryptocurrency trading under realistic transaction costs. The key methodological innovation lies in separating directional prediction from execution decisions through confidence-based thresholds, enabling explicit optimization of the precision–recall trade-off in cryptocurrency trading applications.

Comprehensive experiments across 11 major cryptocurrency pairs demonstrate the framework's effectiveness under realistic trading conditions. High confidence regimes ($\tau = 0.8$) achieve peak profits of 167.64 basis points per trade with directional accuracies of 82–95% on executed trades. Moderate confidence regimes ($\tau = 0.6$) maintain 50–65% market coverage while generating profits of 104.52 basis points per trade for the evaluated 11 major cryptocurrency pairs. These results substantially exceed typical academic benchmarks and demonstrate economic viability under institutional-scale trading volumes.

The systematic parameter optimization reveals fundamental trade-offs between trading frequency and signal quality in cryptocurrency markets. Optimal performance occurs at intermediate prediction horizons (400–600 min) where daily momentum trends manifest through intraday order flow patterns. The confidence-threshold mechanism proves critical for economic performance, with high confidence requirements improving profits by 60.4% while reducing median coverage from 44.9% to 0.28% ($\approx -99\%$).

Multi-scale feature integration provides superior signal representation compared to single-timeframe approaches. The unified combination of macro momentum indicators with microstructure dynamics captures temporal bridges where fundamental price discovery mechanisms align with short-term market-making activities. This integration contributes to directional accuracies that exceed published benchmarks while maintaining economic profitability under realistic transaction costs.

The research demonstrates practical viability for institutional cryptocurrency trading applications. High confidence strategies tolerate transaction costs up to 6 basis points per trade while maintaining positive returns, exceeding typical execution costs for major cryptocurrency pairs. The framework's robust performance across different parameter configurations provides operational flexibility for live deployment across varying market conditions.

The methodological framework developed in this research provides a foundation for blockchain-integrated financial analytics applications. The confidence-threshold mechanism offers particular advantages for smart contract-based trading systems, where binary execution decisions align naturally with on-chain transaction requirements and gas optimization constraints. The systematic parameter optimization approach enables adaptive configuration for different blockchain environments and consensus mechanisms.

Future extensions of this work could integrate on-chain transaction flow analysis with off-chain market signals to develop comprehensive blockchain-native prediction systems [36,37]. The integration of federated learning approaches could enable collaborative model training across multiple DeFi protocols while preserving privacy and reducing centralized dependencies. Additionally, the confidence scoring mechanism could be enhanced with cryptographic verification techniques to ensure signal integrity in decentralized trading environments.

The results presented in this study are derived from 11 major cryptocurrency pairs with high liquidity and complete macro–microstructure data coverage. Generalization to smaller-cap cryptocurrencies, synthetic blockchain assets, or emerging DeFi tokens requires

empirical validation, as lower liquidity and different market microstructure characteristics may substantially alter performance. The framework's applicability to non-cryptocurrency blockchain-native assets (e.g., tokenized securities, NFTs) remains untested and represents an important direction for future research [38,39]. Claims regarding decentralized finance integration should be understood as potential applications requiring additional engineering work for on-chain deployment, rather than validated production-ready implementations.

Several limitations constrain the generalizability of these findings. The evaluation period coincides with specific cryptocurrency market conditions that may not persist across different regulatory environments. The 11-symbol constraint reflects microstructure data availability rather than methodological limitations. Future research should investigate scalability across broader cryptocurrency universes and longer evaluation periods.

Author Contributions: Conceptualization, methodology, writing—original draft preparation, O.K.; supervision, funding acquisition, D.P.-T.; investigation, data curation, D.P.-T.; investigation, methodology, M.B.; conceptualization, data curation, B.K.; writing—review and editing, O.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The complete codebase for this research, including data processing, model implementation, and visualization scripts, is freely available at https://github.com/KuznetsovKarazin/crypto-confidence-execution under MIT License (accessed on 20 September 2025). The repository includes (1) preprocessing pipelines for macro and microstructure data integration; (2) complete model training and evaluation scripts with all hyperparameters specified; (3) confidence-threshold optimization implementation; (4) result reproduction notebooks for all tables and figures; and (5) detailed README with dependency specifications and execution instructions. This accessibility enables direct verification of our results and facilitates further extension of our work by interested researchers.

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

References

1. Guo, L.; Zhong, L.-X. Risk Spillover and Hedging Effects between Stock Markets and Cryptocurrency Markets Depending upon Network Analysis. *N. Am. J. Econ. Financ.* **2025**, *80*, 102524. [CrossRef]
2. Yin, W.; Wu, F.; Zhou, P.; Kirkulak-Uludag, B. Exploring Resilience in the Cryptocurrency Market: Risk Transmission and Network Robustness. *Int. Rev. Financ. Anal.* **2025**, *106*, 104546. [CrossRef]
3. Jasiak, J.; Zhong, C. Intraday and Daily Dynamics of Cryptocurrency. *Int. Rev. Econ. Financ.* **2024**, *96*, 103658. [CrossRef]
4. Mokni, K.; Montasser, G.E.; Ajmi, A.N.; Bouri, E. On the Efficiency and Its Drivers in the Cryptocurrency Market: The Case of Bitcoin and Ethereum. In *Blockchain, Crypto Assets, and Financial Innovation: A Decade of Insights and Advances*; Kou, G., Li, Y., Zhang, Z., Zhao, J.L., Zhuo, Z., Eds.; Springer Nature: Singapore, 2025; pp. 162–191; ISBN 978-981-96-6839-7.
5. Ballis, A.; Karagiorgis, A.; Anastasiou, D.; Kallandranis, C. Cryptocurrency Dynamics during Global Crises: Insights from Bitcoin's Interplay with Traditional Markets. *Int. Rev. Econ. Financ.* **2025**, *103*, 104512. [CrossRef]
6. Fieberg, C.; Liedtke, G.; Zaremba, A. Cryptocurrency Anomalies and Economic Constraints. *Int. Rev. Financ. Anal.* **2024**, *94*, 103218. [CrossRef]
7. Bouteska, A.; Sharif, T.; Isskandarani, L.; Abedin, M.Z. Market Efficiency and Its Determinants: Macro-Level Dynamics and Micro-Level Characteristics of Cryptocurrencies. *Int. Rev. Econ. Financ.* **2025**, *98*, 103938. [CrossRef]
8. Cakici, N.; Shahzad, S.J.H.; Będowska-Sójka, B.; Zaremba, A. Machine Learning and the Cross-Section of Cryptocurrency Returns. *Int. Rev. Financ. Anal.* **2024**, *94*, 103244. [CrossRef]
9. Liu, Y.-H.; Huang, J.-K. Cryptocurrency Trend Forecast Using Technical Analysis and Trading with Randomness-Preserving. *Comput. Electr. Eng.* **2024**, *118*, 109368. [CrossRef]
10. Izadi, M.A.; Hajizadeh, E. Time Series Prediction for Cryptocurrency Markets with Transformer and Parallel Convolutional Neural Networks. *Appl. Soft Comput.* **2025**, *177*, 113229. [CrossRef]
11. Chow, C. On Optimum Recognition Error and Reject Tradeoff. *IEEE Trans. Inf. Theory* **1970**, *16*, 41–46. [CrossRef]

12. Herbei, R.; Wegkamp, M.H. Classification with Reject Option. *Can. J. Stat. La Rev. Can. De Stat.* **2006**, *34*, 709–721. [CrossRef]
13. Cortes, C.; DeSalvo, G.; Mohri, M. Learning with Rejection. In Proceedings of the Algorithmic Learning Theory, Bari, Italy, 19–21 October 2016; Ortner, R., Simon, H.U., Zilles, S., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 67–82.
14. de Prado, M.L. *Advances in Financial Machine Learning*; Wiley: Hoboken, NJ, USA, 2018; ISBN 978-1-119-48208-6.
15. Franco, J.P.M.; Laurini, M.P. Quantifying Systemic Risk in Cryptocurrency Markets: A High-Frequency Approach. *Int. Rev. Econ. Financ.* **2025**, *102*, 104214. [CrossRef]
16. Bonaparte, Y. Time Horizon and Cryptocurrency Ownership: Is Crypto Not Speculative? *J. Int. Financ. Mark. Inst. Money* **2022**, *79*, 101609. [CrossRef]
17. Farooq, A.; Irfan Uddin, M.; Adnan, M.; Alarood, A.A.; Alsolami, E.; Habibullah, S. Interpretable Multi-Horizon Time Series Forecasting of Cryptocurrencies by Leverage Temporal Fusion Transformer. *Heliyon* **2024**, *10*, e40142. [CrossRef]
18. Zhang, J.; Cai, K.; Wen, J. A Survey of Deep Learning Applications in Cryptocurrency. *iScience* **2024**, *27*, 108509. [CrossRef]
19. Shang, D.; Guo, Z.; Wang, H. Enhancing Digital Cryptocurrency Trading Price Prediction with an Attention-Based Convolutional and Recurrent Neural Network Approach: The Case of Ethereum. *Financ. Res. Lett.* **2024**, *67*, 105846. [CrossRef]
20. Zhong, C.; Du, W.; Xu, W.; Huang, Q.; Zhao, Y.; Wang, M. LSTM-ReGAT: A Network-Centric Approach for Cryptocurrency Price Trend Prediction. *Decis. Support Syst.* **2023**, *169*, 113955. [CrossRef]
21. Pellicani, A.; Pio, G.; Ceci, M. CARROT: Simultaneous Prediction of Anomalies from Groups of Correlated Cryptocurrency Trends. *Expert Syst. Appl.* **2025**, *260*, 125457. [CrossRef]
22. Peng, P.; Chen, Y.; Lin, W.; Wang, J.Z. Attention-Based CNN–LSTM for High-Frequency Multiple Cryptocurrency Trend Prediction. *Expert Syst. Appl.* **2024**, *237*, 121520. [CrossRef]
23. Youssefi, A.; Hessane, A.; Zeroual, I.; Farhaoui, Y. Optimizing Forecast Accuracy in Cryptocurrency Markets: Evaluating Feature Selection Techniques for Technical Indicators. *CMC* **2025**, *83*, 3411–3433. [CrossRef]
24. Zhang, Q.; Xie, C.; Weng, Z.; Sornette, D.; Wu, K. Generalized Visible Curvature: An Indicator for Bubble Identification and Price Trend Prediction in Cryptocurrencies. *Decis. Support Syst.* **2024**, *185*, 114309. [CrossRef]
25. Golnari, A.; Komeili, M.H.; Azizi, Z. Probabilistic Deep Learning and Transfer Learning for Robust Cryptocurrency Price Prediction. *Expert Syst. Appl.* **2024**, *255*, 124404. [CrossRef]
26. Anoop, C.V.; Negi, N.; Aprem, A. Bayesian Machine Learning Framework for Characterizing Structural Dependency, Dynamics, and Volatility of Cryptocurrency Market Using Potential Field Theory. *Expert Syst. Appl.* **2025**, *261*, 125475. [CrossRef]
27. Kang, M.; Hong, J.; Kim, S. Harnessing Technical Indicators with Deep Learning Based Price Forecasting for Cryptocurrency Trading. *Phys. A Stat. Mech. Its Appl.* **2025**, *660*, 130359. [CrossRef]
28. Viéitez, A.; Santos, M.; Naranjo, R. Machine Learning Ethereum Cryptocurrency Prediction and Knowledge-Based Investment Strategies. *Knowl.-Based Syst.* **2024**, *299*, 112088. [CrossRef]
29. Liu, W.; Bao, X.; Han, X.; Li, Y. Liquidity Commonality in Cryptocurrencies. *Financ. Res. Lett.* **2025**, *85*, 108187. [CrossRef]
30. Hsieh, C.-H.; Huang, P.-H.; Liu, H.-C. State Transitions and Momentum Effect in Cryptocurrency Market. *Financ. Res. Lett.* **2025**, *86*, 108356. [CrossRef]
31. Yang, Y.; Wang, X.; Xiong, J.; Wu, L.; Zhang, Y. An Innovative Method for Short-Term Forecasting of Blockchain Cryptocurrency Price. *Appl. Math. Model.* **2025**, *138*, 115795. [CrossRef]
32. Nguyen, D.T.A.; Chan, K.C. Cryptocurrency Trading: A Systematic Mapping Study. *Int. J. Inf. Manag. Data Insights* **2024**, *4*, 100240. [CrossRef]
33. Madanchian, M.; Mohamed, N.; Taherdoost, H. Exploring Cryptocurrency Acceptance Patterns: An In-Depth Review of Influencing Factors from Adoption to Adaption for Human Resource Management. *Procedia Comput. Sci.* **2025**, *258*, 3072–3083. [CrossRef]
34. Bai, S.; Kolter, J.Z.; Koltun, V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *arXiv* **2018**, arXiv:1803.01271. [CrossRef]
35. Chechkin, A.; Pleshakova, E.; Gataullin, S.; Chechkin, A.; Pleshakova, E.; Gataullin, S. A Hybrid KAN-BiLSTM Transformer with Multi-Domain Dynamic Attention Model for Cybersecurity. *Technologies* **2025**, *13*, 223. [CrossRef]
36. Kuznetsov, O.; Kostenko, O.; Klymenko, K.; Hbur, Z.; Kovalskyi, R. Machine Learning Analytics for Blockchain-Based Financial Markets: A Confidence-Threshold Framework for Cryptocurrency Price Direction Prediction. *Appl. Sci.* **2025**, *15*, 11145. [CrossRef]
37. Kuznetsov, O.; Sernani, P.; Romeo, L.; Frontoni, E.; Mancini, A. On the Integration of Artificial Intelligence and Blockchain Technology: A Perspective About Security. *IEEE Access* **2024**, *12*, 3881–3897. [CrossRef]

38. Vnukova, N.; Kavun, S.; Kolodiziev, O.; Achkasova, S.; Hontar, D. Indicators-Markers for Assessment of Probability of Insurance Companies Relatedness in Implementation of Risk-Oriented Approach. *Econ. Stud. J.* **2020**, 151–173.
39. Kalashnikov, V.V.; Dempe, S.; Matis, T.I.; Camacho-Vallejo, J.-F.; Kavun, S.V. Bilevel Programming, Equilibrium, and Combinatorial Problems with Applications to Engineering 2016. *Math. Probl. Eng.* **2016**, *2016*, 4360909. [CrossRef]

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