mathematics

MDPI

Article

Explainability-Guided Transformer Models for Hourly Cryptocurrency Forecasting: A Comparative Study with SHAP-Based Feature Refinement

Zeynep Hilal Kilimci [1,*] and Erçin Dinçer [2]

[1] Department of Information Systems Engineering, Kocaeli University, 41001 Kocaeli, Türkiye
[2] Kocaeli University Technology Park, 41275 Kocaeli, Türkiye; ercin.dincer@kouteknopark.com
* Correspondence: zeynep.kilimci@kocaeli.edu.tr

Abstract

Accurate cryptocurrency price forecasting represents an important yet challenging problem in financial time-series analysis due to the highly volatile, nonlinear, and noise-sensitive nature of digital asset markets. Although transformer-based architectures have recently demonstrated strong capabilities in temporal sequence modeling, their behavior under high-frequency cryptocurrency dynamics and the role of explainability-guided feature refinement remain insufficiently explored. To address this gap, this study presents a comprehensive transformer-based forecasting framework for hourly cryptocurrency price prediction and investigates the impact of explainability-guided feature optimization on forecasting performance, robustness, and interpretability. Five transformer architectures—Vanilla Transformer, Informer, Autoformer, Reformer, and Temporal Fusion Transformer (TFT)—are systematically evaluated across five major cryptocurrency assets: Bitcoin (BTC), Ethereum (ETH), Solana (SOL), Dogecoin (DOGE), and Ripple (XRP). The experimental framework employs Open, High, Low, Close, and Volume (OHLCV) data together with a broad set of engineered technical indicators and evaluates model performance using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). To improve interpretability and reduce feature redundancy, SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) are integrated directly into the forecasting pipeline. Based on the resulting explanations, asset-specific feature subsets are constructed, and all models are subsequently retrained using the refined feature representations. The results show that explainability-guided feature refinement provides compact, model-aware, and interpretable feature subsets with competitive forecasting performance; however, its effect on prediction accuracy is dependent on the cryptocurrency asset, transformer architecture, retained feature subset size, and market conditions. Additional robustness, sensitivity, alternative feature-selection, and statistical significance analyses indicate that the SHAP–LIME Top-15 subset should be interpreted as a conservative dimensionality-reduction strategy rather than a universally optimal feature-selection rule. The findings further reveal that transformer architectures incorporating sparse attention, decomposition mechanisms, or gating structures generally provide stronger performance than the Vanilla Transformer under highly volatile hourly market conditions. Overall, the proposed framework demonstrates that combining transformer-based forecasting with explainability-guided feature refinement can support interpretable and parsimonious high-frequency financial time-series modeling, while highlighting the importance of evaluating robustness, feature-selection sensitivity, and statistical variability alongside average forecasting errors.

Academic Editor: Saejoon Kim

Received: 30 July 2026
Revised: 31 August 2026
Accepted: 11 September 2026
Published: 19 September 2026

Keywords: cryptocurrency; financial time-series forecasting; machine learning; transformer; explainable artificial intelligence; SHAP; LIME; financial prediction

MSC: 68T07; 62M20; 91G70

1. Introduction

Cryptocurrencies have become an established component of the global financial landscape, providing decentralized, transparent, and programmable alternatives to conventional financial systems. Bitcoin (BTC), Ethereum (ETH), Solana (SOL), Dogecoin (DOGE), and Ripple (XRP) are among the most widely traded digital assets, attracting increasing interest from both researchers and investors. Despite this growing relevance, forecasting cryptocurrency prices remains a demanding problem due to their high volatility, nonlinear market behavior, and sensitivity to speculative and macroeconomic factors [1]. Accurate forecasting is therefore of considerable importance for financial market analysis, risk assessment, and data-driven decision making, while remaining one of the most challenging problems in financial time-series prediction.

From a financial modeling perspective, the increasing availability of high-frequency market data has created new opportunities for machine learning methods to complement conventional econometric approaches. Financial markets, and cryptocurrency markets in particular, exhibit complex temporal dependencies, volatility clustering, regime changes, and nonlinear interactions that may be difficult to represent using predefined statistical assumptions. Machine learning provides an alternative data-driven framework in which such relationships can be learned directly from historical observations. This capability is particularly relevant for cryptocurrency markets, where rapidly changing market conditions and heterogeneous asset dynamics require forecasting models that can adapt to complex and nonstationary temporal structures.

Traditional statistical models, such as the Autoregressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) frameworks, have been widely applied to financial forecasting [2,3]. While these methods can capture short-term temporal correlations and volatility clustering, they rely on relatively restrictive assumptions and may struggle to model the complex nonlinear dynamics typical of cryptocurrency markets. As a result, their predictive performance can degrade under rapidly shifting regimes and nonstationary conditions.

Deep learning techniques have been proposed to overcome these shortcomings by learning nonlinear temporal dependencies directly from data. Models based on Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRUs) have demonstrated improved accuracy for sequential data [4–6]. However, these architectures remain constrained by sequential computation, limited capacity for long-range dependency modeling, and sensitivity to vanishing gradient problems [7]. Such limitations become more pronounced in high-frequency cryptocurrency markets, where temporal dependencies can span extended horizons and market behavior may shift abruptly.

In recent years, transformer-based architectures have shown strong potential for sequence modeling by employing attention mechanisms to capture global temporal dependencies efficiently [8]. Unlike recurrent models, transformers process entire input sequences in parallel, allowing them to model both local and long-term patterns more effectively. Their success in diverse domains—including natural language processing, energy demand forecasting, and weather prediction—has encouraged their adaptation to financial time-series analysis [9–12]. For financial applications, transformer architectures are particularly attractive because their attention mechanisms can capture dependencies across extended temporal horizons while accommodating high-dimensional sets of market-derived predictors.

However, several important limitations remain in the existing literature on transformer-based cryptocurrency forecasting. First, many studies focus on a single architecture, asset, or temporal resolution, limiting the extent to which conclusions can be generalized across heterogeneous digital assets. Second, comparative analyses examining how different transformer mechanisms behave under volatile and nonlinear market conditions remain relatively scarce. Third, although predictive accuracy is central to financial forecasting, the increasing complexity of deep learning models raises concerns regarding interpretability and the identification of market variables that drive model predictions. Finally, the integration of explainability-driven feature refinement within transformer-based cryptocurrency forecasting pipelines remains largely unexplored. Addressing these issues is particularly important for financial applications, where predictive performance and transparency are both relevant to the practical interpretation and use of model outputs.

To address these gaps, the present study conducts a comprehensive comparative analysis of transformer-based deep learning frameworks for high-resolution (hourly) cryptocurrency price forecasting. The objective is twofold: first, to examine how advanced temporal architectures perform under volatile and nonlinear market conditions; and second, to establish a unified and reproducible experimental protocol for evaluating forecasting accuracy, robustness, and generalization. The analysis covers five leading cryptocurrencies—BTC, ETH, SOL, DOGE, and XRP—using standardized data preprocessing, feature engineering, and hyperparameter optimization. The models are evaluated using widely accepted forecasting metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and R^2, supplemented by explainability-driven feature analyses using SHAP and LIME. Importantly, explainability is incorporated not only as a post hoc interpretation mechanism but also as an active component of the forecasting pipeline, where the resulting feature-importance information is used to refine the predictor space and retrain the forecasting models.

It should be noted that the experimental scope of this study is intentionally restricted to transformer-based forecasting architectures. Classical econometric models such as ARIMA and GARCH, as well as recurrent deep learning models such as LSTM and GRU, are discussed in the related literature to position the study within the broader financial forecasting landscape and to motivate the transition toward attention-based architectures. However, the primary objective of this work is not to establish a universal benchmark across all forecasting families but to conduct a controlled intra-family comparison of transformer variants and to evaluate the effect of SHAP–LIME-guided feature refinement under a unified experimental protocol.

The contributions of this work are summarized as follows:

- This study proposes an explainability-guided transformer forecasting framework in which SHAP and LIME are integrated directly into the learning pipeline to guide feature refinement before model retraining.
- The proposed framework is systematically evaluated using five transformer architectures (Vanilla Transformer, Informer, Autoformer, Reformer, and TFT) across five major cryptocurrencies under a unified experimental protocol.
- SHAP and LIME analyses are integrated directly into the forecasting pipeline to identify the most influential market-derived predictors and reduce feature redundancy through explainability-guided feature refinement.
- The study systematically compares forecasting performance before and after XAI-driven feature optimization and further evaluates the resulting feature subsets through robustness, sensitivity, alternative feature-selection, and statistical significance analy-

ses. The results show that explainability-based feature selection can provide compact and interpretable feature representations with competitive forecasting performance, while its effect remains model- and asset dependent.
- The findings provide insights into how different transformer architectures respond to heterogeneous temporal and volatility characteristics across cryptocurrency assets, while demonstrating the potential of interpretable machine learning for high-frequency financial time-series analysis and data-driven financial decision-support applications.

The remainder of this paper is organized as follows. Section 2 reviews related studies on cryptocurrency forecasting and transformer-based financial time-series modeling. Section 3 presents the transformer architectures considered in the study. Section 4 describes the proposed forecasting framework, including data collection, preprocessing, feature engineering, and experimental configuration. Section 5 reports the experimental results and explainability-guided feature refinement analysis. Section 6 discusses the findings and their implications for financial time-series modeling. Section 7 presents the limitations and future research directions, and Section 8 concludes the paper.

2. Related Work

Research on cryptocurrency forecasting lies at the intersection of financial econometrics, machine learning, and high-frequency time-series analysis. The existing literature can broadly be organized into three streams: (i) classical econometric modeling of returns and volatility; (ii) neural time-series models that learn nonlinear market dynamics directly from data; and (iii) attention-based transformer architectures designed to capture long-range temporal dependencies and multi-horizon relationships. These approaches differ in their assumptions regarding market structure, their ability to represent regime changes and volatility clustering, and their scalability to high-frequency financial data. The following subsections review these streams with particular emphasis on their relevance to machine learning-based financial forecasting across multiple cryptocurrency assets.

2.1. Traditional Time-Series Approaches

Traditional econometric models remain central to the quantitative analysis of cryptocurrency returns, volatility, and financial risk. Methods from the ARIMA and GARCH families are widely used because of their interpretability and well-established statistical foundations, although their ability to represent nonlinear and nonstationary digital-asset dynamics is often limited.

Wu et al. [13] provide a detailed comparison of GARCH and Heterogeneous Autoregressive (HAR) models for Bitcoin volatility forecasting. Their results indicate that the forecasting performance of each model depends heavily on the chosen evaluation horizon and realized-volatility proxy, and they demonstrate that HAR specifications can sometimes outperform GARCH during highly volatile periods.

Bergsli et al. [14] propose a dynamic GARCH re-estimation framework in which model parameters are frequently updated to reflect rapidly changing market structures. Their experiments show that frequent recalibration reduces forecast degradation during turbulent phases, thereby enhancing robustness without substantial computational burden.

Ünlü and Bayram [15] conduct a cross-sectional study comparing stochastic volatility (SV) and GARCH-family models across nine large-cap cryptocurrencies. They find that Bayesian SV models provide more accurate forecasts during periods of extreme volatility, while GARCH performs competitively under moderate market conditions, suggesting that model choice should depend on the prevailing volatility regime.

Sözen [16] applies a Markov-Switching GARCH (MS-GARCH) framework to Value-at-Risk estimation for multiple cryptocurrencies. The study reveals that accounting for regime changes significantly improves tail-risk predictions and highlights that the inclusion of time-varying transition probabilities enhances sensitivity to market shocks.

Gherghina and Constantinescu [17] explore the effects of rolling-window lengths in MS-GARCH volatility modeling. Their findings show that the choice of window size has less impact on predictive stability than the correct specification of regime transitions, emphasizing the need to capture regime persistence rather than simply extending the estimation horizon.

Zahid et al. [18] integrate jump components into GARCH specifications to evaluate the role of discontinuities in Bitcoin volatility. Their results demonstrate that jump-augmented models achieve higher predictive accuracy than traditional GARCH, confirming that sudden price jumps are a key determinant of volatility in crypto assets.

Guirguis [19] implements an ARIMA model for daily Bitcoin price forecasting, carefully diagnosing residuals and differencing orders. Although the ARIMA approach performs well over short horizons, its accuracy deteriorates when the series exhibits structural breaks or long memory, reaffirming the limitations of linear models in cryptocurrency forecasting.

Teker et al. [20] perform a comparative assessment of several GARCH-family models—including EGARCH, TGARCH, and CGARCH—across major cryptocurrencies such as Bitcoin, Ethereum, and Binance Coin. Their findings show that models accounting for asymmetric shocks provide superior fit during downturns, while component-GARCH formulations capture long-term volatility persistence.

PhungDuy et al. [21] investigate GARCH and EGARCH models within a unified experimental setup that includes both cryptocurrency and equity data. They observe that EGARCH consistently outperforms GARCH during high-volatility phases, underscoring the value of capturing asymmetry and leverage effects in financial risk modeling.

Finally, Ustaoğlu [22] extends a Heterogeneous Autoregressive (HAR) model by incorporating news sentiment data for multiple cryptocurrencies. The study shows that including textual sentiment features enhances the accuracy of short-horizon volatility forecasts, demonstrating how exogenous information can complement conventional econometric signals in high-frequency financial modeling.

Overall, these studies demonstrate the continued value of econometric approaches for modeling volatility, risk, and regime behavior. At the same time, their dependence on predefined functional structures motivates the use of more flexible machine learning methods capable of learning nonlinear and high-dimensional relationships directly from financial market data.

2.2. Deep Learning and Neural Architectures

Deep learning approaches have gained widespread attention in cryptocurrency forecasting because of their ability to model nonlinear relationships, temporal dependencies, and high-dimensional covariates with fewer restrictive assumptions about the underlying data-generating process. Recurrent architectures, convolutional hybrids, and attention-augmented variants have therefore been explored as data-driven alternatives to conventional econometric models for financial time-series prediction.

Bouteska et al. [23] conduct a comprehensive comparative analysis of deep learning and ensemble methods for cryptocurrency price prediction. Their findings show that LSTM-based architectures consistently outperform ensemble machine learning models in capturing temporal correlations, particularly during high-volatility phases. The study highlights that deep architectures are more resilient to structural breaks than traditional econometric baselines.

Qureshi et al. [24] evaluate a suite of neural networks—MLP, LSTM, GRU, and CNN—on diverse financial datasets, including cryptocurrencies. Their results demonstrate that recurrent architectures outperform feed-forward and tree-based baselines, with GRU showing faster convergence and better generalization on small samples. They also emphasize the need for hyperparameter tuning and rolling validation to ensure robustness across time windows.

Kaur et al. [25] develop a comparative deep learning framework using LSTM and GRU models for forecasting Bitcoin, Ethereum, and Litecoin. They report that GRU achieves comparable or superior accuracy with lower training time, confirming its suitability for large-scale, high-frequency financial data. The study also explores architecture depth and optimizer sensitivity, providing practical tuning guidelines.

Rodrigues and Machado [26] focus on high-frequency cryptocurrency data, employing LSTM models with varying input windows and feature sets. Their experiments indicate that short input windows (6–12 h) yield the most stable predictions, while longer horizons amplify noise and error accumulation. The results underline the importance of horizon-specific optimization in high-resolution financial forecasting.

Peng et al. [27] propose an attention-based CNN-LSTM hybrid that jointly learns spatial and temporal representations from multi-cryptocurrency datasets. The model demonstrates significant gains in directional accuracy and error reduction compared with standalone LSTM and CNN baselines. This work illustrates the benefits of combining convolutional feature extraction with temporal modeling for cryptocurrency markets.

Tiwari et al. [28] design an attention-augmented CNN-LSTM model that integrates Twitter-based sentiment and market indicators to predict Bitcoin price movements. The addition of attention mechanisms improves interpretability and enhances performance during news-driven volatility spikes. Their findings confirm that incorporating social sentiment can strengthen the predictive power of deep learning models in real-time financial markets.

Nouira et al. [29] provide a survey and empirical synthesis of NLP-driven deep learning models for cryptocurrency forecasting. They review text-based LSTM, BERT, and hybrid sentiment–price architectures, showing that transformer-augmented language models can complement numerical data for more context-aware predictions. Their work further illustrates the growing convergence between financial econometrics, natural language processing, and machine learning.

Ross et al. [30] combine deep neural networks with user-generated content to improve price prediction across multiple cryptocurrencies. Their model fuses sentiment polarity, engagement metrics, and price features through a hybrid dense–recurrent architecture. Results show that textual features reduce forecast errors by over 10% compared with models using price-only data.

Somayajulu et al. [31] present a pedagogical application of LSTM for Bitcoin price prediction using daily and hourly datasets. The study demonstrates that hyperparameter optimization—particularly learning rate and hidden-unit count—significantly influences forecast stability. The work serves as a practical baseline reference for researchers replicating LSTM-based cryptocurrency forecasters.

Lind and Cai [32] propose a multi-scale CNN–attention LSTM model for time-series prediction that captures hierarchical dependencies through convolutional blocks and attention layers. Although applied broadly to financial sequences, the framework generalizes well to cryptocurrency data, achieving strong performance under volatile regimes and outperforming standard LSTM baselines.

Collectively, these studies demonstrate the ability of deep neural models to capture nonlinear financial dynamics more flexibly than conventional approaches. However, recurrent and hybrid architectures may still face limitations in modeling long-range dependencies efficiently, particularly when high-frequency financial observations extend over long temporal windows. These limitations have motivated increasing interest in transformer-based forecasting architectures.

2.3. Transformer-Based Forecasting Models

Attention-based architectures have recently become prominent in financial time-series forecasting because they can model long-range dependencies without relying on sequential recurrence. By processing entire input sequences in parallel, transformer models can learn temporal relationships over extended horizons and accommodate high-dimensional financial predictors, making them particularly attractive for volatile and high-frequency cryptocurrency markets.

Farooq et al. [33] apply an interpretable Temporal Fusion Transformer (TFT) to multi-horizon cryptocurrency forecasting. Their model integrates variable selection networks and attention weight analysis to identify key temporal features that influence price dynamics. The results demonstrate improved forecasting accuracy and explainability compared with LSTM and GRU baselines, particularly under volatile market conditions.

Lee [34] extends TFT for portfolio-level trading of multiple cryptocurrencies by integrating on-chain and technical indicators. Their architecture provides both short- and long-horizon forecasts, improving Sharpe ratios and risk-adjusted returns relative to traditional models. The study emphasizes that attention-based interpretability enhances transparency in automated financial decision systems.

Díaz Berenguer et al. [35] propose a causality-driven extension of the TFT architecture to integrate external signals such as stock and news data. The model captures both temporal and causal dependencies, improving predictive stability during macroeconomic shocks. Although their dataset includes traditional assets, the methodological contribution generalizes to cryptocurrency forecasting frameworks that combine heterogeneous financial information sources.

Stefaniuk and Ślepaczuk [36] investigate the Informer model for algorithmic investment strategies based on high-frequency Bitcoin data. They demonstrate that the ProbSparse self-attention mechanism efficiently handles long input sequences while maintaining lower computational complexity than vanilla transformers. Their results show that Informer-based strategies outperform recurrent architectures in both computational efficiency and predictive accuracy.

In a subsequent study, the same authors [37] refine their Informer implementation by experimenting with different look-back horizons and optimizer settings. They find that increasing the receptive field improves long-term trend detection, while short-term noise can be mitigated through adaptive windowing. This work provides useful insights into tuning Informer for financial time-series applications.

Younas et al. [38] introduce a hybrid attention–correlation transformer named CryptoForetell, which captures both sequential dependencies and inter-asset correlations. Their model achieves superior accuracy on Bitcoin and Ethereum datasets by combining cross-asset relational learning with attention-based feature extraction. The study underscores the value of explicitly modeling inter-market dynamics in multi-asset forecasting.

Khaniki and Manthouri [39] design an enhanced transformer model incorporating Performer-based linear attention and bidirectional LSTM layers. Their approach reduces memory usage and training time while maintaining strong predictive performance across several cryptocurrencies. The results indicate that hybridizing transformer layers with recurrent blocks can balance computational efficiency and representational capacity in cryptocurrency forecasting.

Singh and Bhat [40] adapt transformer architectures for Ethereum price prediction by integrating cross-currency correlation and sentiment features. Their model outperforms both LSTM and CNN baselines, demonstrating that fusing market correlation and textual sentiment improves generalization during periods of high volatility. The study also highlights the robustness of transformer models when applied to heterogeneous financial feature spaces.

Mahdi et al. [41] propose a hybrid transformer–GRU framework that leverages attention mechanisms for global dependencies and GRU layers for local temporal refinement. Applied to Bitcoin and Ethereum data, the model achieves lower RMSE and MAE compared with standalone transformers or GRUs. The combination effectively captures short-term noise while retaining long-term contextual patterns.

Finally, Izadi and Hajizadeh [42] employ transformer blocks within a deep forecasting pipeline optimized for cryptocurrency markets. Their architecture captures nonlinear dependencies between technical indicators and returns, outperforming traditional deep recurrent networks across multiple volatility regimes. The study demonstrates the potential of transformer-based designs to generalize across assets and temporal scales.

Although the existing literature demonstrates the increasing effectiveness of transformer-based models in financial forecasting, several gaps remain. Many studies concentrate on a single cryptocurrency, a specific transformer architecture, or a particular temporal resolution, making cross-model and cross-asset conclusions difficult. Moreover, explainability is commonly treated as a post hoc reporting mechanism rather than as an integral component capable of informing model refinement. The present study addresses these limitations by establishing a unified benchmark across Vanilla Transformer, Informer, Autoformer, Reformer, and Temporal Fusion Transformer architectures on an *hourly* basis for five major cryptocurrencies (BTC, ETH, SOL, DOGE, and XRP). In addition, SHAP- and LIME-based explanations are used to guide feature-space refinement and subsequent model retraining. This design enables a systematic evaluation of predictive accuracy, robustness, interpretability, and feature relevance within a common machine-learning framework for high-frequency financial time-series analysis.

Accordingly, ARIMA, GARCH, LSTM, and GRU are reviewed as important methodological baselines in the broader cryptocurrency forecasting literature, but they are not included in the empirical comparison because the present study focuses specifically on differences among transformer-based architectures. This design choice allows the analysis to isolate how architectural mechanisms such as sparse attention, decomposition, locality-sensitive hashing, and gated variable selection affect forecasting performance and explainability-guided feature refinement.

3. Models

This section presents the transformer-based architectures adopted in this study: the Vanilla Transformer, Informer, Autoformer, Reformer, and Temporal Fusion Transformer (TFT). These models were selected because they represent distinct strategies for attention-based time-series modeling, ranging from general-purpose self-attention to architectures specifically designed for long-sequence forecasting, temporal decomposition, computational efficiency, and interpretable multi-horizon prediction. Evaluating these architectures under a common experimental setting provides a systematic basis for examining how different attention mechanisms respond to the nonlinear, volatile, and high-frequency characteristics of cryptocurrency markets.

3.1. Vanilla Transformer

The original Transformer architecture introduced a paradigm shift in sequence modeling by eliminating recurrence and convolution and replacing them with self-attention mechanisms that compute pairwise dependencies between all elements in a sequence [8]. In this framework, each input vector attends to every other vector through scaled dot-product attention, producing context-aware representations that capture both local and global dependencies. Multi-head attention enables the model to learn multiple representation subspaces in parallel, while positional encodings preserve sequential order information that is otherwise absent from purely attention-based operations.

In time-series forecasting, the Transformer's encoder–decoder structure can be adapted for autoregressive or multi-horizon prediction. The encoder captures temporal dependencies within the historical look-back window, whereas the decoder generates future values either sequentially or across multiple forecast horizons. Despite its strong representation capacity, the Vanilla Transformer has quadratic computational complexity with respect to sequence length, making it computationally demanding for long temporal sequences such as high-frequency financial data. Nevertheless, it provides an important baseline against which architectures specifically designed to improve long-sequence efficiency and temporal modeling can be evaluated.

3.2. Informer

The Informer architecture addresses the quadratic computational bottleneck of conventional self-attention through its ProbSparse self-attention mechanism, which selects a subset of dominant queries based on attention energy scores [9]. This reduces attention complexity from $O(L^2)$ to approximately $O(L \log L)$, enabling more efficient processing of long temporal sequences. Informer also employs a generative decoder that predicts extended horizons in a single forward pass and introduces a distillation operation that progressively reduces redundant feature representations across layers.

For financial time-series forecasting, these mechanisms allow Informer to capture extended temporal patterns without the memory overhead associated with full self-attention. This property is particularly relevant to high-frequency cryptocurrency data, where long look-back windows may contain information regarding persistent trends and volatility dynamics. However, sparse attention may occasionally underrepresent short-lived local events or abrupt market movements when a limited subset of temporal interactions dominates the attention computation.

3.3. Autoformer

Autoformer enhances long-term time-series forecasting through two principal mechanisms: a series decomposition block and an auto-correlation mechanism [10]. The decomposition block separates the time series into trend and seasonal components, enabling the model to represent underlying temporal structures explicitly. The auto-correlation mechanism identifies and aggregates similar sub-series patterns in the latent space, allowing the architecture to capture periodic dependencies beyond the effective range of conventional self-attention.

In cryptocurrency forecasting, this decomposition mechanism provides a means of distinguishing persistent temporal structure from shorter-term fluctuations. Such a design may be advantageous when financial series exhibit recurring intraday or inter-day patterns. However, its effectiveness may decrease when periodic structure becomes weak or unstable, particularly during abrupt market regime changes or volatility shocks. Autoformer therefore provides a useful contrast to attention-based architectures that do not explicitly decompose the observed time series.

3.4. Reformer

Reformer aims to preserve the representation capability of the Transformer while substantially reducing its memory requirements and computational cost [11]. It employs locality-sensitive hashing (LSH) attention and reversible residual layers. LSH attention groups similar queries and keys into buckets so that attention is computed primarily among related representations rather than across all possible pairs. Reversible layers further reduce memory requirements by allowing intermediate activations to be reconstructed during backpropagation instead of being stored throughout the forward pass.

For high-resolution financial data, these design choices make Reformer attractive when long input sequences must be processed under computational constraints. The architecture can retain broad temporal context while reducing the resource requirements associated with conventional self-attention. Nevertheless, the hashing mechanism may introduce sensitivity to bucket configuration and other hyperparameters, which can affect stability when modeling noisy and highly volatile cryptocurrency series.

3.5. Temporal Fusion Transformer (TFT)

The Temporal Fusion Transformer (TFT) is designed specifically for interpretable multi-horizon time-series forecasting [12]. Unlike conventional Transformer architectures, TFT integrates specialized components for handling heterogeneous temporal inputs, including variable selection networks, gating mechanisms for adaptive feature processing, and attention layers for capturing both short- and long-term dependencies. These components enable TFT to model temporal dynamics and interactions among multiple covariates while simultaneously providing information regarding variable relevance.

These characteristics are particularly relevant to cryptocurrency forecasting, where market observations are accompanied by multiple derived indicators representing trend, momentum, volatility, and trading activity. The TFT variable-selection and gating mechanisms provide a structured means of identifying and weighting informative inputs while suppressing less useful signals. Its architecture also supports multi-horizon forecasting and interpretable temporal representations, although this flexibility comes at the cost of greater architectural complexity and hyperparameter tuning requirements. The combination of predictive modeling and built-in interpretability therefore makes TFT an important benchmark for high-frequency financial time-series forecasting.

4. Proposed Framework

This section presents the proposed framework for hourly cryptocurrency price forecasting using transformer-based architectures. The framework integrates data acquisition, preprocessing, technical feature engineering, temporal sample construction, model training, performance evaluation, and explainability-guided feature refinement within a unified experimental pipeline. As illustrated in Figure 1, SHAP and LIME are incorporated as an internal explainability stage to identify influential financial predictors and guide the construction of a reduced feature space for subsequent model retraining. This design enables explainability to be evaluated not only as a post hoc interpretation mechanism but also as an active component of the forecasting pipeline that may improve robustness by reducing the influence of redundant or weakly informative market indicators.

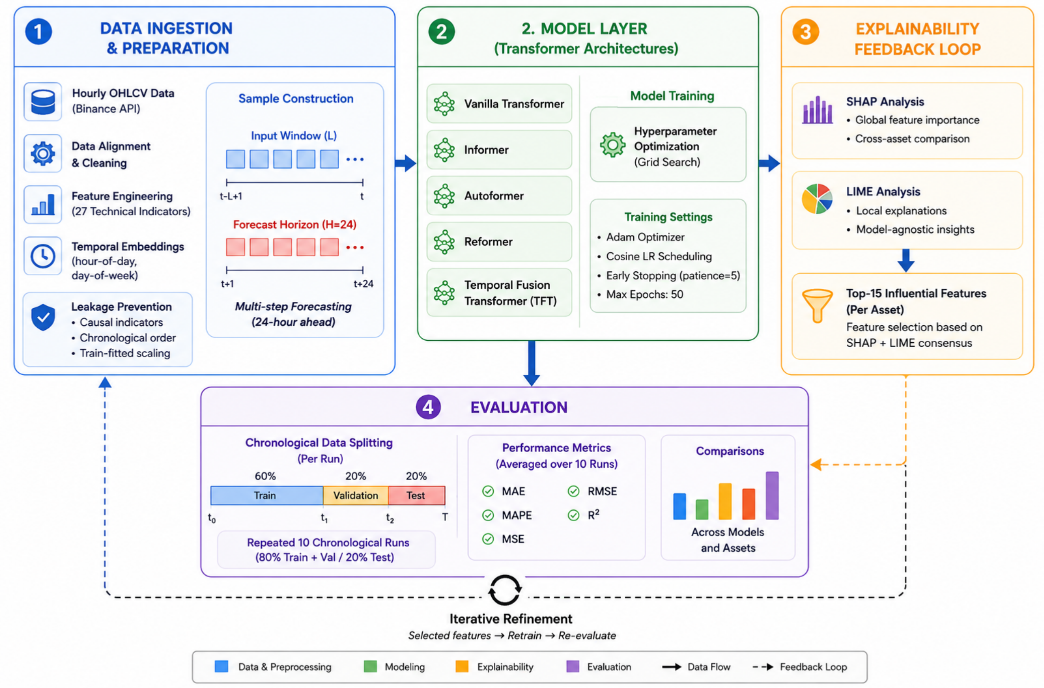

Figure 1. Complete framework for hourly cryptocurrency forecasting, including chronological train–validation–test partitioning, transformer-based forecasting, performance evaluation, and SHAP/LIME-guided feature refinement followed by model retraining.

4.1. Mathematical Formulation of the Forecasting Problem

Let $a \in \mathcal{A}$ denote a cryptocurrency asset, where $\mathcal{A} = \{\mathrm{BTC}, \mathrm{ETH}, \mathrm{SOL}, \mathrm{DOGE}, \mathrm{XRP}\}$. For each asset a, the multivariate input vector at time t is defined as

$$\mathbf{x}_t^{(a)} = [x_{t,1}^{(a)}, x_{t,2}^{(a)}, \ldots, x_{t,d}^{(a)}]^\top \in \mathbb{R}^d,$$

where d denotes the number of market-derived predictors, including OHLCV variables, technical indicators, and temporal features. Given a look-back window of length L, the model input is constructed as

$$\mathbf{X}_t^{(a)} = [\mathbf{x}_{t-L+1}^{(a)}, \mathbf{x}_{t-L+2}^{(a)}, \ldots, \mathbf{x}_t^{(a)}] \in \mathbb{R}^{L \times d}.$$

The forecasting task is to estimate the future closing-price vector over a prediction horizon H, expressed as

$$\hat{\mathbf{y}}_{t+1:t+H}^{(a)} = f_\theta^{(m)}(\mathbf{X}_t^{(a)}),$$

where $f_\theta^{(m)}$ denotes the forecasting function parameterized by θ for model architecture m, with $m \in \{\mathrm{Transformer}, \mathrm{Informer}, \mathrm{Autoformer}, \mathrm{Reformer}, \mathrm{TFT}\}$. In this study, $H = 24$, corresponding to a 24 h forecasting horizon.

4.2. Data Collection and Preprocessing

Hourly OHLCV (*Open*, *High*, *Low*, *Close*, *Volume*) data were collected from the Binance REST API for five liquid cryptocurrencies: Bitcoin (BTC), Ethereum (ETH), Solana (SOL), Dogecoin (DOGE), and Ripple (XRP). The main evaluation period spans 1 May 2025 to 31 October 2025, corresponding to approximately 4416 hourly observations per asset. In addition, historical observations covering 1 May 2021 to 30 April 2025 (approximately 35,000 hourly observations) were used to provide sufficient historical context for rolling feature computation and preprocessing.

All cryptocurrency series were aligned to a strict hourly temporal index, and missing observations were handled within the preprocessing pipeline while preserving the chronological structure of the series. Extreme observations in log-returns and trading volumes were controlled through winsorization at the 1st and 99th percentiles.

Particular attention was given to preventing information leakage throughout the forecasting pipeline. All rolling technical indicators were constructed causally using only observations available up to time t, ensuring that future price information was not incorporated into predictors used to forecast subsequent observations. Numerical scaling was performed using parameters estimated from the training data and subsequently applied to the validation and test sets. Together with strict chronological partitioning, this procedure prevents information from later evaluation periods from being introduced into model fitting or feature construction.

4.3. Feature Engineering

The baseline OHLCV representation was expanded using 27 engineered technical indicators designed to characterize different dimensions of cryptocurrency market behavior, including trend, momentum, volatility, and trading participation. The indicators were calculated using reproducible implementations provided by TA-Lib and pandas-ta, complemented by NumPy and pandas transformations for ratios, slopes, and first-order differences. Table 1 summarizes the resulting predictors and their functional roles within the forecasting framework.

For clarity, the main technical indicators used in this study include Exponential Moving Average (EMA), Simple Moving Average (SMA), Moving Average Convergence Divergence (MACD), Relative Strength Index (RSI), Stochastic Relative Strength Index (StochRSI), Commodity Channel Index (CCI), Money Flow Index (MFI), Bollinger Bands (BB), Average True Range (ATR), Donchian Channel indicators, On-Balance Volume (OBV), Average Directional Index (ADX), Williams %R, volume oscillator, and first-order momentum-change indicators.

Table 1. Engineered technical indicators and their functional roles in the forecasting framework.

Feature	Description
`ema_9`	Short-horizon EMA capturing rapid trend shifts.
`ema_21`	Medium-horizon EMA reflecting intermediate trend behavior.
`ema_50`	Long-horizon EMA representing sustained trend direction.
`sma_20`	Simple moving average smoothing short-term fluctuations.
`macd_line`	EMA(12)–EMA(26), representing short-term momentum.
`macd_signal`	9 h EMA of MACD used to characterize crossover behavior.
`macd_histogram`	Difference between MACD and its signal line.
`rsi_14`	Relative Strength Index representing momentum conditions.
`stoch_rsi`	RSI-derived oscillator capturing rapid momentum changes.
`cci_14`	Commodity Channel Index measuring deviation from recent price behavior.
`mfi_14`	Money Flow Index combining price and volume information.
`bollinger_upper`	Upper Bollinger Band (20, 2).
`bollinger_lower`	Lower Bollinger Band (20, 2).
`bollinger_width`	Bollinger Band width representing volatility magnitude.
`bollinger_percent_b`	Normalized price position within the Bollinger Bands.
`atr_14`	Average True Range representing absolute volatility.
`atr_normalized`	Price-normalized ATR representing relative volatility.
`supertrend_10_3`	ATR-adjusted adaptive trend indicator.
`donchian_upper_20`	Highest high over the preceding 20 h window.
`donchian_lower_20`	Lowest low over the preceding 20 h window.
`obv`	On-Balance Volume representing directional trading participation.
`volume_oscillator`	Relative behavior of short- and long-term volume averages.
`adx_14`	Average Directional Index representing trend strength.
`williams_r_14`	Relative closing-price position within recent price extremes.
`ema_ratio_9_21`	EMA9/EMA21 ratio representing relative trend intensity.
`macd_hist_slope`	First-order change in the MACD histogram.
`delta_rsi`	First difference of RSI representing momentum change.

In addition to market-derived predictors, cyclical temporal information was represented through hour-of-day and day-of-week embeddings. These temporal variables were appended to the feature representation to allow the models to learn recurring intraday and weekly structures that may be relevant to continuously traded cryptocurrency markets.

The complete feature representation was initially retained for baseline model training rather than assuming a priori that all engineered indicators contributed equally to forecasting performance. Feature relevance was subsequently investigated using SHAP and LIME. This separation between the complete baseline feature space and the explainability-guided reduced feature space enables the effect of XAI-driven feature refinement on forecasting performance to be evaluated explicitly.

4.4. Forecasting Architecture and Temporal Evaluation

Following preprocessing and feature construction, the financial time series were transformed into sliding-window samples. Each input window contains L historical hourly observations and is used to forecast the subsequent $H = 24$ h, thereby formulating the task as a high-frequency multi-horizon cryptocurrency price forecasting problem.

The model parameters were learned by minimizing the empirical forecasting loss over the training set:

$$\theta^* = \arg\min_{\theta} \frac{1}{N_{\text{train}}} \sum_{i=1}^{N_{\text{train}}} \mathcal{L}\Big(\mathbf{y}_i, f_\theta^{(m)}(\mathbf{X}_i)\Big),$$

where N_{train} is the number of training samples, $\mathbf{y}_i$ is the observed future target vector, and $\mathcal{L}(\cdot)$ corresponds to the mean squared error loss used during model optimization.

Five transformer-based architectures—Vanilla Transformer, Informer, Autoformer, Reformer, and Temporal Fusion Transformer (TFT)—were evaluated under a common experimental protocol. These architectures represent different mechanisms for modeling complex temporal dependencies, including conventional full self-attention, sparse attention, series decomposition, locality-sensitive hashing, and gated variable selection.

To preserve the temporal structure of the financial series and prevent information leakage, observations were partitioned chronologically into training, validation, and test sets using a 60%/20%/20% split. The earliest 60% of observations were used for model training, the subsequent 20% formed the validation set, and the most recent 20% were retained as a held-out test set. No random shuffling was performed during temporal partitioning.

The validation set was used during model development for hyperparameter selection, convergence monitoring, and early stopping. In contrast, the held-out test set was reserved exclusively for final performance evaluation and did not contribute to model fitting, hyperparameter optimization, or validation-based model selection. This chronological separation is particularly important in financial time-series forecasting, where random partitioning may introduce future market information into earlier prediction periods and lead to overly optimistic estimates of model generalization.

To account for variability arising from stochastic model optimization, the experimental procedure was repeated across ten independent runs while preserving the same chronological train–validation–test structure. Forecasting performance was aggregated across these runs and reported as mean performance. This repeated evaluation strategy provides a more stable assessment of model behavior while maintaining strict temporal separation among training, validation, and test periods.

Forecasting performance was quantified using five complementary metrics: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). These metrics provide complementary perspectives on absolute forecasting deviation, scale-normalized error, sensitivity to larger prediction errors, and overall goodness of fit.

The evaluation metrics were computed as follows:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|,$$

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2,$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2},$$

$$\mathrm{MAPE} = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right|,$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}.$$

Here, y_i, $\hat{y}_i$, $\bar{y}$, and N denote the observed value, predicted value, mean observed value, and number of test observations, respectively.

4.5. Hyperparameter Optimization and Training Configuration

Hyperparameters were optimized using the chronologically separated validation set, which remained distinct from the held-out test set throughout model development. Grid-search optimization was performed for the transformer architectures to account for their different structural characteristics. The resulting configurations are summarized in Table 2.

Table 2. Optimized hyperparameter configurations used for the transformer-based forecasting models.

Model	L	Batch	Heads	Blocks	d_{ff}	LR
Informer	120	64	4	2	256	1×10^{-4}
Autoformer	60	32	4	3	128	1×10^{-5}
Reformer	30	16	4	2	256	1×10^{-4}
Transformer	60	32	4	2	256	1×10^{-4}
TFT	60	32	4	2	256	1×10^{-4}

The input sequence length L determines the amount of historical information available to the forecasting model. The number of attention heads defines parallel attention subspaces, while the number of blocks specifies the depth of the architecture. The feed-forward dimension d_{ff} controls the capacity of the intermediate representation, and the learning rate determines the magnitude of parameter updates during optimization. Batch size was additionally selected to balance convergence stability and computational requirements.

All experiments were implemented in Python 3.10 using PyTorch 2.13.0. Data processing was performed using pandas, NumPy, and scikit-learn, while technical indicators were calculated using TA-Lib and pandas-ta. Model training was conducted on an NVIDIA GPU with at least 24 GB of VRAM using CUDA acceleration and deterministic random seeds.

The Adam optimizer was employed together with cosine-annealed learning-rate scheduling. Training was performed for a maximum of 50 epochs, with early stopping applied using validation loss and a patience of five epochs. Mean Squared Error (MSE) was used as the training objective. The held-out test set was not used for convergence monitoring, hyperparameter selection, or early stopping.

4.6. Explainability-Guided Feature Refinement

Following baseline model training, SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) were incorporated to investigate the contribution of individual financial predictors to the forecasting process. The two approaches provide complementary perspectives on model behavior: SHAP characterizes feature contributions through Shapley-value-based attribution, whereas LIME provides local explanations by approximating model behavior around individual predictions.

Explainability analysis was conducted separately for each cryptocurrency to account for potential differences in feature relevance across heterogeneous market dynamics. The feature-importance information obtained from SHAP and LIME was jointly considered to derive an explainability-guided feature ranking for each asset. Based on the resulting SHAP–LIME consensus, the 15 most influential predictors were retained to construct an asset-specific reduced feature space. This procedure allows complementary explanation information from both methods to contribute to feature refinement rather than relying on a single explainability technique.

For each asset a, let $\phi_j^{\text{SHAP},(a)}$ denote the global SHAP importance of feature j, and let $\phi_j^{\text{LIME},(a)}$ denote the corresponding aggregated LIME-based importance score. A consensus explainability score for each feature was defined as

$$s_j^{(a)} = \alpha\left|\phi_j^{\text{SHAP},(a)}\right| + (1-\alpha)\left|\phi_j^{\text{LIME},(a)}\right|,$$

where $\alpha \in [0,1]$ controls the relative contribution of SHAP and LIME. The asset-specific reduced feature subset was then obtained as

$$\mathcal{S}_{15}^{(a)} = \text{Top}_{15}\left(\{s_j^{(a)}\}_{j=1}^{d}\right).$$

The refined input representation is therefore given by

$$\mathbf{X}_{t,\text{refined}}^{(a)} = \mathbf{X}_t^{(a)}[:, \mathcal{S}_{15}^{(a)}],$$

which retains only the selected Top-15 predictors for asset a.

Following feature selection, all five transformer architectures were retrained using the corresponding reduced Top-15 feature set while preserving the same chronological train–validation–test structure and general evaluation protocol used for the baseline experiments. Forecasting performance before and after explainability-guided feature refinement was then compared using the same five evaluation metrics.

This two-stage experimental design allows SHAP and LIME to serve not only as post hoc interpretation tools but also as components of an iterative model-refinement strategy. Consequently, the framework enables direct assessment of whether explainability-guided feature reduction can provide a compact and interpretable feature representation while maintaining competitive predictive performance and reducing dependence on redundant or weakly informative financial indicators.

5. Experimental Results

This section presents the experimental evaluation of the proposed transformer-based cryptocurrency forecasting framework in four stages. First, baseline forecasting performance is evaluated using the complete feature representation consisting of OHLCV variables, engineered technical indicators, and temporal information. Second, SHAP and LIME are employed to investigate asset-specific feature relevance and identify the predictors that contribute most strongly to the forecasting process. Third, the models are retrained using the explainability-guided Top-15 feature subsets to examine whether the reduced feature representation can provide competitive forecasting performance while improving interpretability and reducing dimensionality. Finally, additional robustness, sensitivity, alternative feature-selection, and statistical significance analyses are conducted to assess the stability and generalizability of the feature-refinement strategy.

All results reported in this section were obtained using the chronological 60%/20%/20% train–validation–test structure described in Section 4. The final 20% of each time series was retained as a held-out test period, while the validation partition was used during model development and early stopping. To account for variability arising from stochastic model optimization, the experiments were repeated over ten independent runs while preserving the same chronological partitioning scheme. The reported MSE, RMSE, MAE, MAPE, and R^2 values represent mean performance across these runs.

5.1. Baseline Performance

Table 3 summarizes the forecasting performance obtained using the complete feature representation before explainability-guided feature refinement. The results are reported as mean $\pm$ standard deviation over ten independent runs, and MAPE values are expressed as percentages. The updated baseline results show that forecasting performance varies across both model architectures and cryptocurrency assets, although the Temporal Fusion Transformer (TFT) provides the strongest baseline performance in terms of MAPE across the evaluated assets.

Table 3. Baseline (pre-XAI) forecasting performance using the full feature representation. Results are reported as mean $\pm$ standard deviation over ten independent runs. MAPE is expressed as a percentage.

Model	Asset	MSE	RMSE	MAE	MAPE (%)	R^2
Informer	BTC	0.0044 ± 0.0001	0.0660 ± 0.0009	0.0487 ± 0.0009	1.293 ± 0.024	0.8696 ± 0.0037
	SOL	0.0074 ± 0.0002	0.0860 ± 0.0009	0.0602 ± 0.0007	2.406 ± 0.026	0.8458 ± 0.0032
	ETH	0.0017 ± 0.0000	0.0412 ± 0.0005	0.0293 ± 0.0005	2.171 ± 0.039	0.7788 ± 0.0057
	DOGE	0.0051 ± 0.0001	0.0712 ± 0.0006	0.0447 ± 0.0004	2.885 ± 0.031	0.8771 ± 0.0020
	XRP	0.0026 ± 0.0003	0.0508 ± 0.0029	0.0349 ± 0.0015	2.349 ± 0.112	0.8622 ± 0.0163
Autoformer	BTC	0.0044 ± 0.0000	0.0660 ± 0.0004	0.0478 ± 0.0003	1.275 ± 0.009	0.8618 ± 0.0016
	SOL	0.0072 ± 0.0001	0.0848 ± 0.0005	0.0587 ± 0.0004	2.337 ± 0.017	0.8394 ± 0.0020
	ETH	0.0017 ± 0.0001	0.0415 ± 0.0007	0.0300 ± 0.0008	2.218 ± 0.057	0.7586 ± 0.0083
	DOGE	0.0048 ± 0.0001	0.0691 ± 0.0004	0.0435 ± 0.0005	2.782 ± 0.029	0.8788 ± 0.0014
	XRP	0.0022 ± 0.0001	0.0471 ± 0.0007	0.0322 ± 0.0005	2.148 ± 0.033	0.8807 ± 0.0035
Reformer	BTC	0.0042 ± 0.0002	0.0646 ± 0.0016	0.0470 ± 0.0012	1.253 ± 0.031	0.8662 ± 0.0065
	SOL	0.0067 ± 0.0001	0.0819 ± 0.0007	0.0572 ± 0.0005	2.275 ± 0.020	0.8448 ± 0.0027
	ETH	0.0015 ± 0.0001	0.0381 ± 0.0012	0.0270 ± 0.0008	2.005 ± 0.063	0.7903 ± 0.0128
	DOGE	0.0046 ± 0.0002	0.0679 ± 0.0012	0.0434 ± 0.0009	2.779 ± 0.056	0.8796 ± 0.0041
	XRP	0.0021 ± 0.0001	0.0454 ± 0.0010	0.0312 ± 0.0006	2.074 ± 0.041	0.8865 ± 0.0051
Transformer	BTC	0.0042 ± 0.0001	0.0644 ± 0.0009	0.0470 ± 0.0007	1.253 ± 0.018	0.8681 ± 0.0035
	SOL	0.0067 ± 0.0002	0.0816 ± 0.0011	0.0573 ± 0.0005	2.277 ± 0.019	0.8512 ± 0.0041
	ETH	0.0015 ± 0.0001	0.0392 ± 0.0009	0.0278 ± 0.0005	2.060 ± 0.039	0.7850 ± 0.0093
	DOGE	0.0047 ± 0.0001	0.0688 ± 0.0008	0.0435 ± 0.0008	2.781 ± 0.051	0.8796 ± 0.0027
	XRP	0.0022 ± 0.0001	0.0472 ± 0.0009	0.0319 ± 0.0006	2.132 ± 0.043	0.8798 ± 0.0046
TFT	BTC	0.0040 ± 0.0000	0.0629 ± 0.0001	0.0442 ± 0.0001	1.177 ± 0.003	0.8743 ± 0.0005
	SOL	0.0069 ± 0.0001	0.0831 ± 0.0005	0.0562 ± 0.0005	2.239 ± 0.018	0.8456 ± 0.0018
	ETH	0.0014 ± 0.0000	0.0372 ± 0.0001	0.0252 ± 0.0000	1.865 ± 0.003	0.8066 ± 0.0009
	DOGE	0.0044 ± 0.0000	0.0667 ± 0.0004	0.0400 ± 0.0000	2.568 ± 0.004	0.8870 ± 0.0012
	XRP	0.0019 ± 0.0000	0.0434 ± 0.0002	0.0284 ± 0.0001	1.895 ± 0.010	0.8987 ± 0.0009

TFT achieves the lowest baseline MAPE for all five cryptocurrencies, with values of 1.177 $\pm$ 0.003% for BTC, 2.239 $\pm$ 0.018% for SOL, 1.865 $\pm$ 0.003% for ETH, 2.568 $\pm$ 0.004% for DOGE, and 1.895 $\pm$ 0.010% for XRP. These results indicate that the TFT gating mechanisms, temporal attention structure, and variable-selection components are effective when the complete multivariate feature representation is used. The relatively small standard deviations further suggest that TFT exhibits stable baseline behavior across repeated runs.

Reformer and the Vanilla Transformer provide competitive intermediate baseline performance. Reformer obtains MAPE values of 1.253 ± 0.031% for BTC, 2.275 ± 0.020% for SOL, 2.005 ± 0.063% for ETH, 2.779 ± 0.056% for DOGE, and 2.074 ± 0.041% for XRP. The Vanilla Transformer obtains very similar baseline results, with MAPE values of 1.253 ± 0.018% for BTC, 2.277 ± 0.019% for SOL, 2.060 ± 0.039% for ETH, 2.781 ± 0.051% for DOGE, and 2.132 ± 0.043% for XRP. These findings suggest that both standard self-attention and locality-sensitive hashing attention can capture relevant temporal patterns in hourly cryptocurrency data, although they do not outperform TFT under the full feature representation.

Informer also achieves competitive baseline performance, particularly for BTC, ETH, and SOL, but it does not provide the lowest MAPE under the updated run-level evaluation. Its MAPE values are 1.293 ± 0.024% for BTC, 2.406 ± 0.026% for SOL, 2.171 ± 0.039% for ETH, 2.885 ± 0.031% for DOGE, and 2.349 ± 0.112% for XRP. These results indicate that sparse attention remains useful for modeling high-frequency cryptocurrency sequences, although its effectiveness varies across assets and does not uniformly exceed that of architectures with stronger variable-selection mechanisms.

Autoformer shows comparable but asset-dependent baseline performance. It obtains MAPE values of 1.275 ± 0.009% for BTC, 2.337 ± 0.017% for SOL, 2.218 ± 0.057% for ETH, 2.782 ± 0.029% for DOGE, and 2.148 ± 0.033% for XRP. The model performs relatively close to the other non-TFT transformer architectures for BTC, SOL, DOGE, and XRP, but its decomposition-based design does not provide the strongest baseline accuracy in this experimental setting. This suggests that explicit temporal decomposition may be beneficial in some financial time-series contexts, but its advantage depends on the structure of the underlying asset dynamics.

Overall, the baseline experiments indicate that TFT provides the strongest performance when the complete feature representation is used, while the remaining transformer architectures produce broadly competitive but asset-dependent results. The observed differences across architectures motivate the subsequent explainability analysis, which investigates whether the original feature space can be reduced through SHAP–LIME-guided feature refinement while maintaining competitive forecasting performance and improving interpretability.

5.2. Explainability Analysis Using SHAP and LIME

To investigate the internal decision mechanisms of the forecasting models and identify the predictors contributing most strongly to their outputs, SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) were applied following baseline model training. The two methods provide complementary perspectives: SHAP characterizes feature contributions from a broader attribution perspective, whereas LIME provides local explanations around individual predictions.

The explainability analysis was performed separately for BTC, ETH, SOL, DOGE, and XRP. This asset-specific analysis is important because the baseline experiments demonstrate that the forecasting behavior of the architectures varies substantially across cryptocurrencies. Consequently, imposing a single universal feature ranking could obscure differences in the information structures associated with individual assets.

For the LIME visualizations in Figure 2, one fixed prediction instance was selected from the held-out test set for each cryptocurrency to provide an illustrative local explanation under the same evaluation setting used for final model assessment. Since LIME provides an explanation for an individual prediction rather than a global summary of model behavior, the selected instance was used as an illustrative local example. The purpose of the LIME panels is therefore to demonstrate how individual technical indicators contribute to a specific prediction, whereas the SHAP analysis provides the corresponding global feature-importance perspective across the test set. Accordingly, the LIME examples should be interpreted as representative local illustrations rather than as complete summaries of the overall dataset.

As illustrated in Figure 2, the SHAP and LIME analyses reveal heterogeneous feature-importance patterns across the five cryptocurrencies. For BTC, SHAP assigns substantial importance to `mfi_14`, `ema_50`, `macd_hist_slope`, and `delta_rsi`, indicating contributions from volume-adjusted market activity, trend information, and momentum changes. The corresponding LIME explanation also emphasizes momentum- and volatility-related predictors, including `stoch_rsi`, `macd_hist_slope`, `delta_rsi`, and `bollinger_width`.

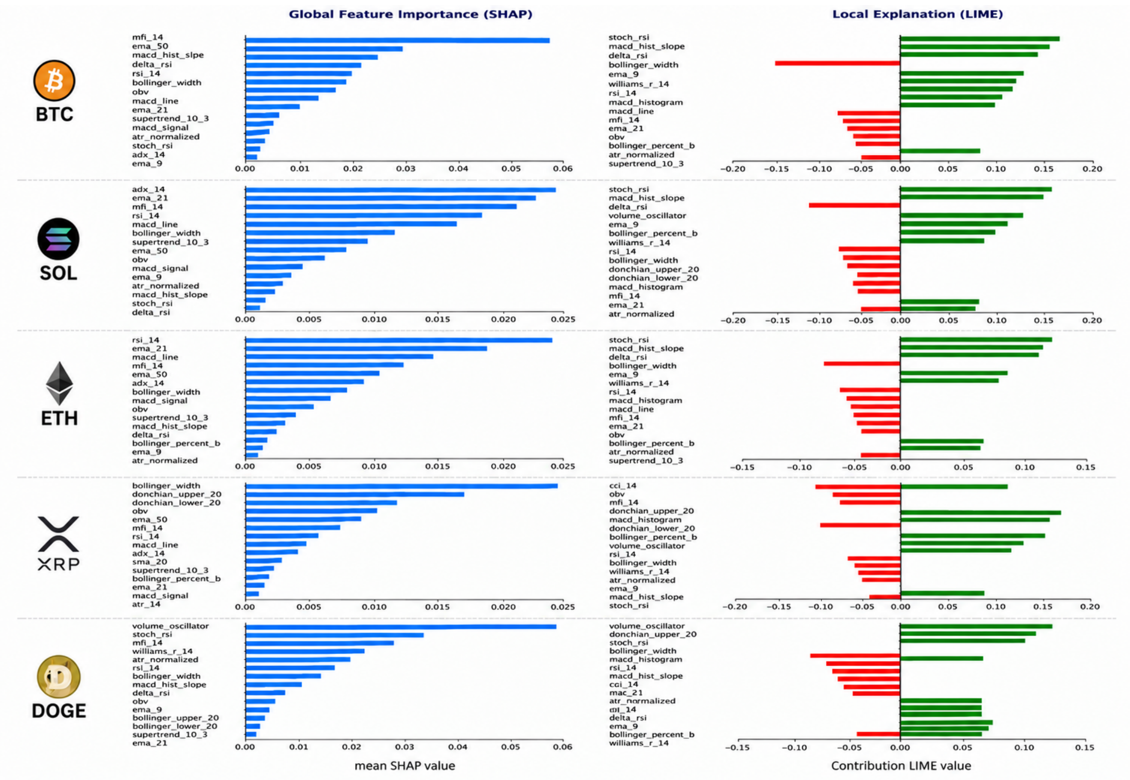

Figure 2. Global and local feature explanations for BTC, SOL, ETH, XRP, and DOGE obtained using SHAP and LIME. In the SHAP panels, blue horizontal bars indicate global feature importance based on mean absolute SHAP values across the test set. In the LIME panels, green bars indicate positive local contributions and red bars indicate negative local contributions for one fixed prediction instance selected from the held-out test set for the corresponding cryptocurrency. Since LIME is inherently local, these examples are intended to provide representative local illustrations rather than complete summaries of the overall dataset. The results demonstrate that feature relevance varies across assets, motivating the construction of asset-specific Top-15 feature subsets for retraining the transformer-based forecasting models.

For ETH, the explanation patterns indicate substantial contributions from trend- and momentum-related variables. SHAP highlights features including `rsi_14`, `ema_21`, `macd_line`, and `mfi_14`, while the corresponding LIME analysis further identifies momentum changes and Bollinger-based volatility information as relevant to local predictions.

For SOL, momentum- and trend-related indicators dominate the explanation patterns. SHAP identifies `rsi_14`, `ema_21`, `macd_line`, `mfi_14`, and `bollinger_width` among the most influential predictors, while the corresponding LIME explanation highlights `stoch_rsi`, `macd_hist_slope`, `delta_rsi`, and `bollinger_width`. These results indicate that both momentum dynamics and short-term trend information play important roles in forecasting SOL.

For DOGE, the explanation results emphasize trading-activity and momentum-related information. Predictors such as `volume_oscillator`, `stoch_rsi`, `mfi_14`, `williams_r_14`, and `rsi_14` contribute prominently to the model explanations, suggesting that trading activity and momentum dynamics provide valuable information for DOGE price forecasting.

For XRP, SHAP primarily highlights volatility-, price-channel-, and volume-related indicators, including `bollinger_width`, `donchian_upper_20`, `donchian_lower_20`, `obv`, and `ema_50`. The corresponding LIME explanation similarly emphasizes volatility- and trend-sensitive variables, indicating that XRP predictions rely on a relatively focused subset of engineered features.

Taken together, the SHAP and LIME results demonstrate that the transformer-based forecasting models do not rely uniformly on the complete set of engineered predictors. Instead, feature relevance varies across assets, with trend, momentum, volatility, and trading-activity indicators contributing to different degrees. Based on the combined SHAP–LIME analysis, an asset-specific Top-15 feature subset was constructed and subsequently used to retrain all five transformer architectures.

5.3. Refined Performance After Explainability-Guided Feature Selection

Following the explainability analysis, the feature space was reduced to the asset-specific Top-15 predictors identified through the combined SHAP and LIME analysis. All five transformer architectures were subsequently retrained using the reduced feature representation while preserving the same chronological train–validation–test structure and evaluation protocol employed in the baseline experiments.

The purpose of this second-stage evaluation is to determine whether explainability-guided feature reduction can provide a compact and interpretable representation while maintaining competitive forecasting performance. Table 4 summarizes the resulting post-XAI forecasting performance using the asset-specific SHAP–LIME Top-15 feature subsets. Results are reported as mean $\pm$ standard deviation over ten independent runs, and MAPE values are expressed as percentages.

Informer shows modest improvements after SHAP–LIME Top-15 feature refinement for several assets. Its MAPE decreases from 1.293 $\pm$ 0.024% to 1.286 $\pm$ 0.016% for BTC, from 2.406 $\pm$ 0.026% to 2.387 $\pm$ 0.033% for SOL, from 2.171 $\pm$ 0.039% to 2.122 $\pm$ 0.027% for ETH, and from 2.885 $\pm$ 0.031% to 2.864 $\pm$ 0.037% for DOGE. However, for XRP, MAPE increases from 2.349 $\pm$ 0.112% to 2.664 $\pm$ 0.103%. These results indicate that sparse attention can benefit from a reduced feature representation in some settings, but the effect is not uniform across all assets.

Autoformer exhibits a slightly less favorable response to the SHAP–LIME Top-15 representation. Its MAPE changes from 1.275 $\pm$ 0.009% to 1.286 $\pm$ 0.028% for BTC, from 2.337 $\pm$ 0.017% to 2.366 $\pm$ 0.022% for SOL, from 2.218 $\pm$ 0.057% to 2.289 $\pm$ 0.049% for ETH, from 2.782 $\pm$ 0.029% to 2.790 $\pm$ 0.028% for DOGE, and from 2.148 $\pm$ 0.033% to 2.177 $\pm$ 0.029% for XRP. Although these changes are relatively small, they suggest that the decomposition-based architecture may rely on complementary information distributed across the broader feature set.

Reformer also shows a mixed response to feature refinement. Its MAPE increases slightly for BTC, SOL, DOGE, and XRP, changing from 1.253 $\pm$ 0.031% to 1.266 $\pm$ 0.041%, from 2.275 $\pm$ 0.020% to 2.281 $\pm$ 0.048%, from 2.779 $\pm$ 0.056% to 2.810 $\pm$ 0.079%, and from 2.074 $\pm$ 0.041% to 2.113 $\pm$ 0.060%, respectively. In contrast, ETH improves from 2.005 $\pm$ 0.063% to 1.982 $\pm$ 0.066%. These results indicate that locality-sensitive hashing attention can maintain broadly competitive forecasting performance under the reduced representation, but it does not consistently benefit from Top-15 feature selection.

Table 4. Refined (post-XAI) forecasting performance using the asset-specific SHAP–LIME Top-15 feature representation. Results are reported as mean ± standard deviation over ten independent runs. MAPE is expressed as a percentage.

Model	Asset	MSE	RMSE	MAE	MAPE (%)	R^2
Informer	BTC	0.0043 ± 0.0001	0.0655 ± 0.0007	0.0484 ± 0.0006	1.286 ± 0.016	0.8715 ± 0.0026
	SOL	0.0072 ± 0.0002	0.0850 ± 0.0010	0.0598 ± 0.0008	2.387 ± 0.033	0.8494 ± 0.0036
	ETH	0.0016 ± 0.0000	0.0403 ± 0.0006	0.0286 ± 0.0004	2.122 ± 0.027	0.7887 ± 0.0060
	DOGE	0.0050 ± 0.0001	0.0704 ± 0.0008	0.0444 ± 0.0005	2.864 ± 0.037	0.8798 ± 0.0027
	XRP	0.0037 ± 0.0005	0.0607 ± 0.0038	0.0392 ± 0.0015	2.664 ± 0.103	0.8038 ± 0.0244
Autoformer	BTC	0.0044 ± 0.0002	0.0661 ± 0.0012	0.0483 ± 0.0011	1.286 ± 0.028	0.8612 ± 0.0051
	SOL	0.0073 ± 0.0002	0.0855 ± 0.0009	0.0595 ± 0.0006	2.366 ± 0.022	0.8367 ± 0.0034
	ETH	0.0018 ± 0.0001	0.0429 ± 0.0008	0.0310 ± 0.0007	2.289 ± 0.049	0.7426 ± 0.0095
	DOGE	0.0048 ± 0.0001	0.0692 ± 0.0004	0.0436 ± 0.0004	2.790 ± 0.028	0.8782 ± 0.0015
	XRP	0.0023 ± 0.0001	0.0480 ± 0.0007	0.0326 ± 0.0004	2.177 ± 0.029	0.8758 ± 0.0037
Reformer	BTC	0.0041 ± 0.0002	0.0643 ± 0.0016	0.0475 ± 0.0015	1.266 ± 0.041	0.8675 ± 0.0064
	SOL	0.0066 ± 0.0002	0.0810 ± 0.0014	0.0573 ± 0.0013	2.281 ± 0.048	0.8478 ± 0.0054
	ETH	0.0014 ± 0.0001	0.0373 ± 0.0014	0.0268 ± 0.0009	1.982 ± 0.066	0.7993 ± 0.0151
	DOGE	0.0046 ± 0.0003	0.0680 ± 0.0019	0.0439 ± 0.0012	2.810 ± 0.079	0.8793 ± 0.0068
	XRP	0.0022 ± 0.0001	0.0472 ± 0.0013	0.0317 ± 0.0009	2.113 ± 0.060	0.8770 ± 0.0069
Transformer	BTC	0.0042 ± 0.0002	0.0646 ± 0.0012	0.0475 ± 0.0009	1.265 ± 0.023	0.8674 ± 0.0048
	SOL	0.0067 ± 0.0002	0.0818 ± 0.0011	0.0578 ± 0.0008	2.292 ± 0.033	0.8506 ± 0.0040
	ETH	0.0015 ± 0.0001	0.0389 ± 0.0009	0.0277 ± 0.0008	2.046 ± 0.058	0.7881 ± 0.0103
	DOGE	0.0048 ± 0.0002	0.0693 ± 0.0013	0.0442 ± 0.0009	2.826 ± 0.060	0.8777 ± 0.0046
	XRP	0.0023 ± 0.0001	0.0477 ± 0.0013	0.0319 ± 0.0009	2.132 ± 0.060	0.8774 ± 0.0064
TFT	BTC	0.0042 ± 0.0003	0.0645 ± 0.0020	0.0457 ± 0.0019	1.220 ± 0.053	0.8677 ± 0.0082
	SOL	0.0069 ± 0.0001	0.0830 ± 0.0004	0.0564 ± 0.0002	2.246 ± 0.009	0.8461 ± 0.0016
	ETH	0.0014 ± 0.0000	0.0371 ± 0.0001	0.0252 ± 0.0000	1.862 ± 0.003	0.8070 ± 0.0009
	DOGE	0.0044 ± 0.0001	0.0662 ± 0.0006	0.0402 ± 0.0002	2.577 ± 0.015	0.8887 ± 0.0020
	XRP	0.0019 ± 0.0000	0.0430 ± 0.0005	0.0280 ± 0.0003	1.871 ± 0.018	0.9003 ± 0.0023

The Vanilla Transformer demonstrates similarly heterogeneous behavior. MAPE increases from 1.253 ± 0.018% to 1.265 ± 0.023% for BTC, from 2.277 ± 0.019% to 2.292 ± 0.033% for SOL, and from 2.781 ± 0.051% to 2.826 ± 0.060% for DOGE. For ETH, MAPE decreases slightly from 2.060 ± 0.039% to 2.046 ± 0.058%, while XRP remains practically unchanged at approximately 2.132%. These findings suggest that conventional self-attention can preserve competitive performance after feature reduction in some cases, but the reduced representation does not provide a consistent improvement over the full feature set.

TFT remains the strongest post-XAI architecture in terms of MAPE across all five cryptocurrencies. After SHAP–LIME Top-15 feature refinement, TFT obtains MAPE values of 1.220 ± 0.053% for BTC, 2.246 ± 0.009% for SOL, 1.862 ± 0.003% for ETH, 2.577 ± 0.015% for DOGE, and 1.871 ± 0.018% for XRP. Relative to the full feature representation, TFT shows slight improvements for ETH and XRP, while the changes for BTC, SOL, and DOGE indicate small degradations. Thus, TFT retains the best overall post-XAI forecasting performance, but the benefit of feature refinement remains asset dependent.

Overall, the updated post-XAI results indicate that the SHAP–LIME Top-15 representation provides a compact, model-aware, and interpretable feature subset with competitive forecasting performance. However, it does not uniformly outperform the full feature representation across all model–asset combinations. The effect of feature refinement depends on the transformer architecture and cryptocurrency asset, with some combinations benefiting from feature reduction and others performing slightly better with the full feature set.

From a financial forecasting perspective, these findings suggest that expanding the predictor space with numerous technical indicators does not always guarantee superior out-of-sample performance, but aggressive feature reduction may also remove useful complementary signals for certain architectures. Therefore, explainability-guided feature refinement should be interpreted as a practical mechanism for constructing parsimonious and interpretable forecasting models, rather than as a universally superior feature-selection rule. This interpretation is further examined through the robustness, sensitivity, alternative feature-selection, and statistical significance analyses presented in Section 5.4.

5.4. Additional Robustness and Sensitivity Analyses

In addition to the main pre-XAI and post-XAI forecasting results, several supplementary analyses were conducted to further evaluate the robustness, stability, and sensitivity of the proposed explainability-guided feature refinement strategy. These analyses examine model behavior under sudden market movements, the effect of varying the number of retained SHAP–LIME features, comparisons with alternative feature-selection baselines, and run-to-run statistical variability.

The purpose of these additional analyses is not to replace the main forecasting evaluation but to provide a more rigorous assessment of whether the SHAP–LIME-based reduced feature representation remains competitive under different experimental and market conditions.

5.4.1. Robustness Under Market Regime Shifts and Sudden Price Jumps

To address the behavior of the forecasting models under non-average market conditions, an additional robustness analysis was conducted on the held-out test period. Since cryptocurrency markets frequently exhibit abrupt price movements, volatility clustering, and short-lived regime changes, evaluating only average forecasting errors may obscure model behavior during unstable market phases. Therefore, the test segment was further examined in terms of volatility regimes and sudden price-jump hours.

Let $P_t^{(a)}$ denote the closing price of asset a at time t. The hourly log-return was computed as

$$r_t^{(a)} = \log(P_t^{(a)}) - \log(P_{t-1}^{(a)}).$$

Sudden price-jump hours were identified using an asset-specific absolute log-return threshold. Specifically, an observation was labeled as a jump hour when the magnitude of the hourly log-return exceeded the corresponding high-percentile threshold:

$$\mathrm{Jump}_t^{(a)} = \begin{cases} 1, & \text{if } |r_t^{(a)}| > \tau_a, \\ 0, & \text{otherwise,} \end{cases}$$

where τ_a denotes the asset-specific jump threshold. In addition, rolling volatility regimes were examined using a 24 h rolling standard deviation of log-returns. High- and low-volatility regimes were identified based on the upper and lower quantiles of the rolling-volatility distribution, respectively.

The held-out test period spans 26 September 2025 to 31 October 2025 and contains 854 hourly observations for each cryptocurrency. As shown in Figure 3, the held-out test segment contains both volatility-regime changes and sudden price-jump observations across all five assets. As summarized in Table 5, all five assets exhibit jump hours during the test segment. BTC has the largest number of jump observations, with 28 jump hours corresponding to 3.28% of the test period. ETH, SOL, DOGE, and XRP also contain sudden-jump observations, with jump ratios ranging from 1.41% to 2.34%. These results confirm that the held-out test set contains non-smooth market behavior and regime-shift-like volatility patterns rather than only stable average-market conditions.

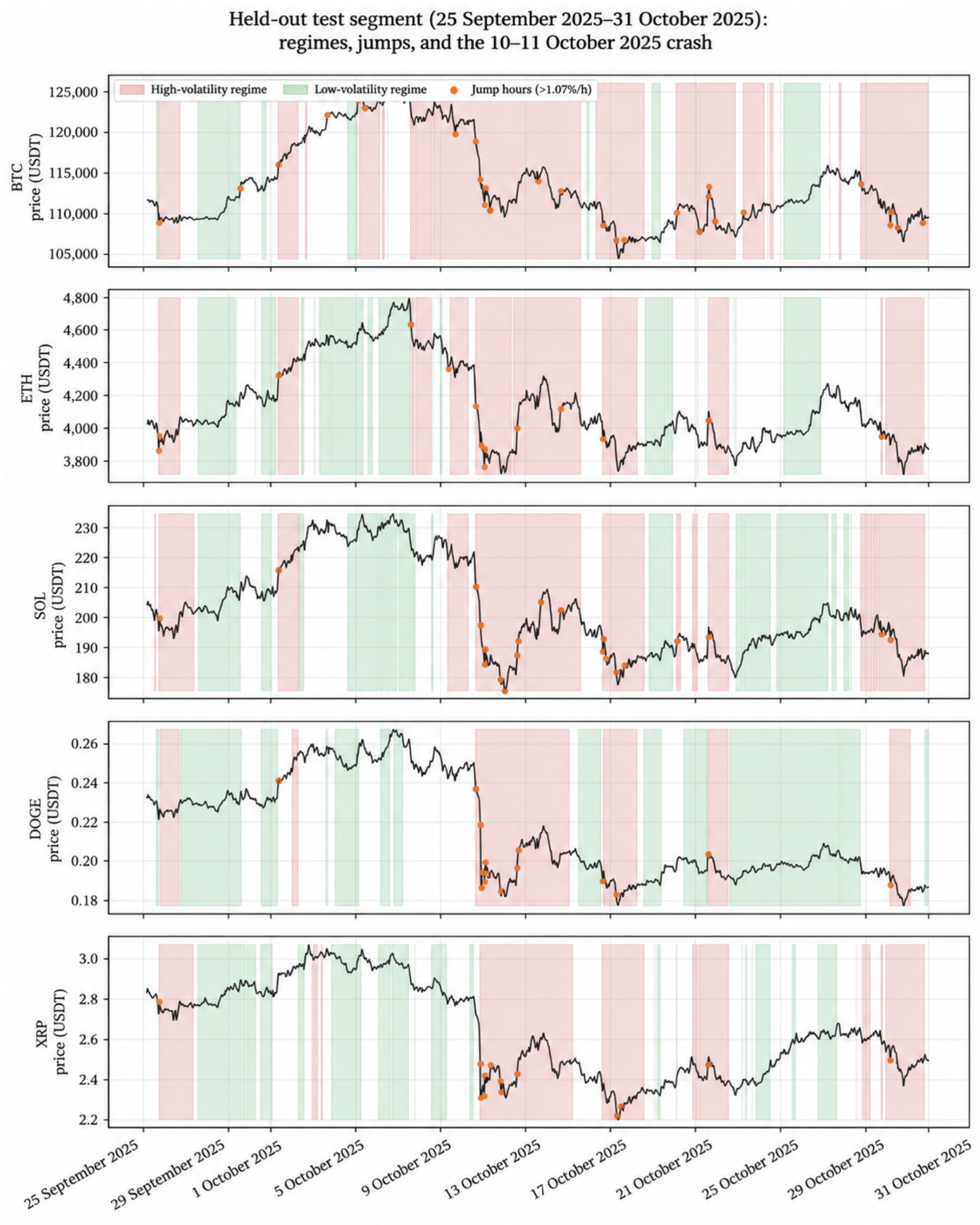

Figure 3. Held-out test segment for BTC, ETH, SOL, DOGE, and XRP, showing price trajectories, high- and low-volatility regimes, and sudden price-jump hours. The shaded regions indicate volatility-regime changes, while the marked points correspond to jump hours identified using asset-specific hourly log-return thresholds. The test period covers 26 September 2025 to 31 October 2025 and includes abrupt market movements, including the 10–11 October 2025 crash period, confirming that the evaluation segment contains regime-shift-like behavior rather than only stable market conditions.

Model robustness was then evaluated by comparing forecasting errors during normal hours and sudden-jump hours. As shown in Figure 4, jump-hour errors are substantially higher than normal-hour errors across the five assets and five transformer architectures. Using the full feature representation, the average normal-hour MAPE is 2.016%, whereas the average jump-hour MAPE is increased to 6.401%, corresponding to an average error-inflation factor of approximately 3.04. With the SHAP–LIME Top-15 representation, the corresponding normal-hour and jump-hour MAPE values are 2.038% and 6.377%, respectively, with a slightly lower average error-inflation factor of approximately 3.00.

Table 5. Summary of volatility regimes and sudden price-jump observations in the held-out test period.

Asset	Test Hours	Jump Threshold (%/h)	Jump Hours	Jump Ratio (%)
BTC	854	1.068	28	3.28
ETH	854	2.172	13	1.52
SOL	854	2.280	20	2.34
DOGE	854	2.846	15	1.76
XRP	854	2.181	12	1.41

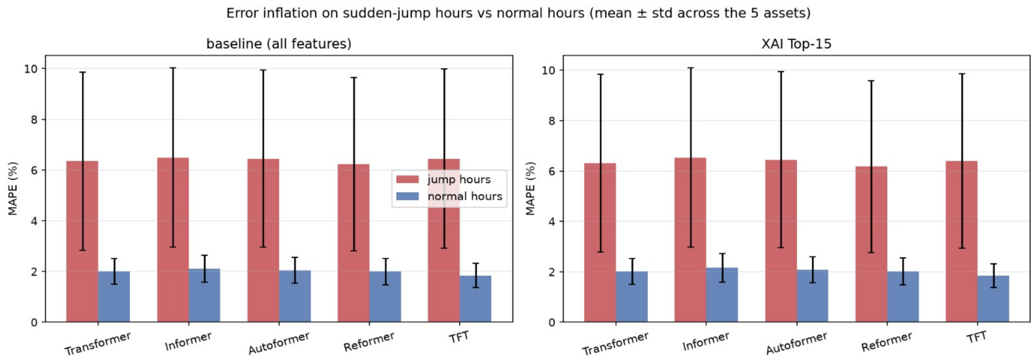

Figure 4. Comparison of normal-hour and sudden-jump-hour MAPE for the full feature representation and the SHAP–LIME Top-15 representation. Bars represent average MAPE values across the five cryptocurrency assets for each transformer architecture. Normal-hour errors correspond to non-jump observations, whereas jump-hour errors correspond to observations exceeding the asset-specific hourly log-return threshold. Jump-hour errors are substantially larger than normal-hour errors for all transformer architectures, indicating that abrupt price movements remain challenging. The SHAP–LIME Top-15 representation slightly reduces the average jump-hour error inflation compared with the full feature set.

These findings indicate that sudden price jumps remain challenging for all evaluated transformer architectures, as expected in highly volatile financial time series. However, the SHAP–LIME Top-15 representation slightly reduces the average jump-hour error and error-inflation factor compared with the full feature set. This suggests that explainability-guided feature refinement can provide a more compact representation that remains competitive under abrupt market movements, although its robustness benefit is modest and model dependent. Therefore, the results should not be interpreted as indicating complete immunity to sudden market shocks, but rather as evidence that the proposed framework maintains stable and interpretable forecasting behavior under both normal and jump-hour conditions.

5.4.2. Sensitivity to the Number of Retained SHAP–LIME Features

To justify the choice of the Top-15 feature subset and to examine the effect of dimensionality reduction more systematically, a sensitivity analysis was conducted with different numbers of retained SHAP–LIME-ranked predictors. Specifically, each transformer architecture was retrained using the Top-5, Top-10, Top-15, and Top-20 features for each cryptocurrency asset while preserving the same chronological train–validation–test structure and ten-run evaluation protocol used in the main experiments.

Figure 5 shows the resulting MAPE values for each model and asset as the number of retained features K varies. The results indicate that forecasting performance is generally stable across different feature subset sizes, but the optimal value of K is not identical for all assets and architectures. In several cases, smaller subsets such as Top-5 provide competitive or slightly better performance, whereas Top-15 and Top-20 tend to preserve a broader set of market indicators with only limited additional error.

Averaged across all assets, models, and runs, the MAPE values for Top-5, Top-10, Top-15, and Top-20 were 2.071%, 2.119%, 2.121%, and 2.128%, respectively. These results show that the forecasting models are not highly sensitive to small changes in the number of retained features, although very compact representations can be competitive. The Top-15 setting was retained in the main experiments as a conservative dimensionality-reduction choice because it substantially reduces the original feature space while preserving a sufficiently broad set of trend, momentum, volatility, and volume-related indicators for interpretation and model retraining.

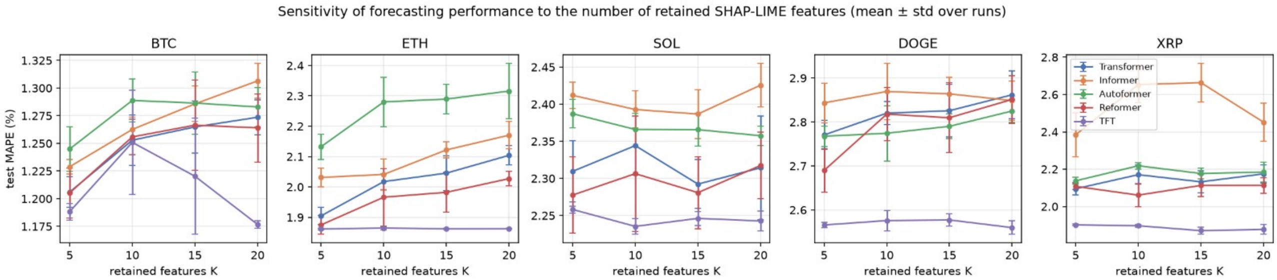

Figure 5. Sensitivity of forecasting performance to the number of retained SHAP–LIME features. Each panel reports the test MAPE values obtained by retraining the five transformer-based architectures with Top-5, Top-10, Top-15, and Top-20 asset-specific features. Results are averaged over ten independent runs. The curves show that forecasting performance is generally stable across different feature subset sizes, while the best number of retained features remains model- and asset dependent.

5.4.3. Comparison with Alternative Feature-Selection Strategies

To further examine whether the SHAP–LIME-based Top-15 subset provides an advantage over simpler feature-selection strategies, an additional comparison was conducted against two baseline selection methods: random Top-15 feature selection and absolute-correlation-based Top-15 feature selection. The correlation-based baseline selected the 15 predictors with the highest absolute Pearson correlation with the target variable using only the training partition, thereby avoiding information leakage from the validation or test periods. The random baseline selected 15 predictors from the full feature space without using either target information or model-attribution information. To reduce dependence on a single random draw, the random baseline was averaged over multiple random subsets.

As shown in Figure 6, the differences among the feature-selection strategies are relatively small and depend on the cryptocurrency asset. Averaged across all assets, models, and runs, the overall MAPE values were 2.100% for the full feature representation, 2.121% for the SHAP–LIME Top-15 subset, 2.077% for the absolute-correlation Top-15 subset, and 2.095% for the random Top-15 baseline. These results indicate that SHAP–LIME-based feature refinement provides competitive forecasting performance while substantially reducing the feature space, although it does not uniformly outperform all alternative selection strategies in terms of average MAPE.

The correlation-based baseline performs strongly in some settings, which suggests that linear associations between technical indicators and future prices remain informative for several cryptocurrency assets. However, unlike random or purely correlation-based selection, the SHAP–LIME Top-15 subset has the advantage of being model-aware and directly interpretable since the selected predictors are derived from global and local explanation scores of the trained forecasting models. Therefore, the role of the SHAP–LIME-based refinement should be interpreted not as a universally optimal feature-selection rule but as an explainability-guided mechanism that provides a compact, interpretable, and competitive feature representation for transformer-based cryptocurrency forecasting. Furthermore, the overall averages are summarized in Table 6.

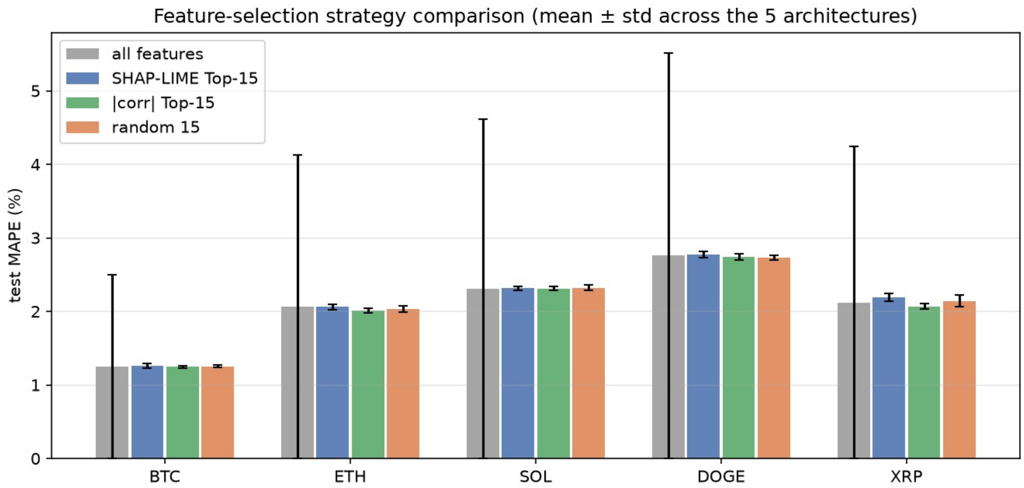

Figure 6. Comparison of feature-selection strategies using the full feature representation, SHAP–LIME Top-15 features, absolute-correlation Top-15 features, and random Top-15 features. Bars represent mean test MAPE values, and error bars indicate standard deviations across the five transformer-based architectures and cryptocurrency assets. The results show that the performance differences among feature-selection strategies are asset dependent and that SHAP–LIME Top-15 provides a compact and interpretable representation with competitive forecasting performance.

Table 6. Average MAPE comparison of feature-selection strategies across all assets, models, and runs.

Feature-Selection Strategy	Average MAPE (%)
Full feature representation	2.100
SHAP–LIME Top-15	2.121
Absolute-correlation Top-15	2.077
Random Top-15	2.095

5.4.4. Run-to-Run Variability and Statistical Significance

To determine whether the observed performance differences between the full feature representation and the SHAP–LIME Top-15 representation are attributable to stochastic training variability, an additional run-to-run variability and statistical significance analysis was conducted. For each model–asset combination, the MAPE values obtained from the ten independent runs were compared between the full feature setting and the SHAP–LIME Top-15 setting.

Because the number of repeated runs is relatively small and normality of the paired differences cannot be assumed, the paired Wilcoxon signed-rank test was used to compare the two feature representations. The null hypothesis was that the median paired difference in MAPE between the full feature representation and the SHAP–LIME Top-15 representation was zero. Statistical significance was assessed at the $p < 0.05$ level.

Figure 7 illustrates the run-to-run MAPE variability for the full feature representation and the SHAP–LIME Top-15 representation across the evaluated assets and transformer architectures. The distributions show that the magnitude and direction of the feature-refinement effect vary across model–asset combinations, indicating that the effect of explainability-guided feature reduction is not uniform.

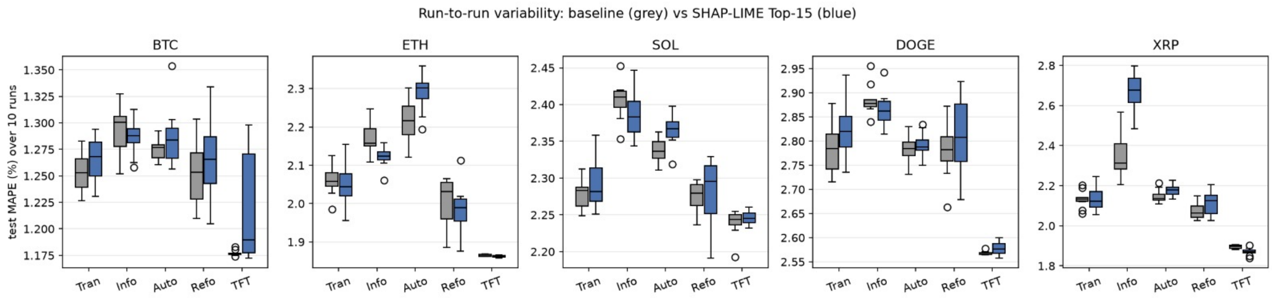

Figure 7. Run-to-run variability of test MAPE for the full feature representation and the SHAP–LIME Top-15 representation across cryptocurrency assets and transformer-based architectures. Each distribution corresponds to repeated independent runs under the same chronological train–validation–test protocol. Different boxplot groups represent different feature representations and model–asset combinations, while individual points indicate run-level MAPE values. The results show that the effect of explainability-guided feature refinement varies across assets and models, motivating the use of statistical testing rather than relying only on average MAPE values.

The paired significance analysis further confirms that the effect of SHAP–LIME-based feature refinement is statistically and practically model-dependent. Statistically significant MAPE reductions were observed for Informer on ETH and TFT on XRP. In contrast, statistically significant increases in MAPE were observed for TFT on BTC, Autoformer on ETH, Autoformer on SOL, and Informer on XRP. The remaining model–asset combinations did not show statistically significant differences at the $p < 0.05$ level. The model–asset combinations for which the paired Wilcoxon signed-rank test indicates statistically significant MAPE differences are summarized in Table 7.

Table 7. Model–asset combinations with statistically significant MAPE differences between the full feature representation and the SHAP–LIME Top-15 representation according to the paired Wilcoxon signed-rank test. Lower SHAP–LIME Top-15 MAPE values indicate improvement after feature refinement, whereas higher values indicate degradation.

Asset	Model	Full MAPE (%)	SHAP–LIME Top-15 MAPE (%)	p-Value
ETH	Informer	2.171	2.122	0.0137
XRP	TFT	1.895	1.871	0.0195
BTC	TFT	1.177	1.220	0.0195
ETH	Autoformer	2.218	2.289	0.0488
SOL	Autoformer	2.337	2.366	0.0195
XRP	Informer	2.349	2.664	0.0020

These results show that the MAPE differences are not purely attributable to random noise in all cases; however, they also demonstrate that SHAP–LIME-based feature refinement should not be interpreted as producing uniformly significant improvements across all architectures and assets. Instead, the statistically supported conclusion is that the Top-15 representation provides a compact and interpretable feature subset with competitive forecasting performance, while the magnitude and direction of performance changes remain dependent on the asset, architecture, and market conditions. Accordingly, the claims in the manuscript were revised to avoid overgeneralizing the improvement effect of explainability-guided feature selection.

6. Discussion

The principal finding of this study is that explainability can be incorporated as an active model-refinement mechanism rather than serving solely as a post hoc interpretation tool. By integrating SHAP and LIME into the forecasting pipeline, the proposed framework identifies asset-specific informative predictors and enables feature-space reduction while maintaining competitive forecasting performance in several settings. Rather than identifying a universally superior architecture or feature-selection strategy, the experiments reveal that model architecture, asset-specific market dynamics, retained feature subset size, and market conditions jointly influence forecasting behavior. The additional robustness, sensitivity, alternative feature-selection, and statistical significance analyses further show that explainability-guided feature refinement should be interpreted as a compact and interpretable modeling strategy whose performance effect is model- and asset dependent.

The updated baseline results indicate that the Temporal Fusion Transformer (TFT) provides the strongest overall performance when the full feature representation is used. This result is consistent with the architectural design of TFT, which combines temporal attention, gating mechanisms, and variable-selection components to process heterogeneous financial predictors. In contrast, Informer, Autoformer, Reformer, and the Vanilla Transformer produce broadly competitive but asset-dependent results. These findings suggest that architectural mechanisms designed for variable selection and temporal filtering can be particularly beneficial in high-frequency cryptocurrency forecasting, although the relative differences among architectures remain dependent on the asset being modeled.

After SHAP–LIME Top-15 feature refinement, TFT remains the strongest post-XAI architecture in terms of MAPE across the evaluated assets. However, the effect of feature refinement is not uniformly beneficial. Informer shows modest improvements for several assets but deteriorates for XRP, Autoformer generally performs slightly better with the full feature representation, and Reformer and the Vanilla Transformer exhibit mixed responses. These results indicate that reducing the feature space can preserve competitive performance in some settings, but may also remove complementary information that is useful for certain architectures. Therefore, explainability-guided feature refinement should not be interpreted as a universal mechanism for improving predictive accuracy, but rather as a structured approach for constructing compact and interpretable feature representations.

One of the principal contributions of this study lies in integrating explainability methods directly into the forecasting pipeline. In many previous studies, SHAP and LIME are primarily employed to interpret trained models after prediction. In contrast, the present framework uses explainability as a data-driven mechanism for feature refinement. The SHAP and LIME analyses identify a relatively compact subset of momentum-, trend-, volatility-, and volume-related indicators as influential predictors for the forecasting models. Retraining the forecasting models using these asset-specific Top-15 feature subsets shows that a reduced, explanation-guided feature space can retain competitive forecasting performance while improving interpretability and reducing dimensionality. However, the additional comparisons with Top-5, Top-10, Top-20, random Top-15, and correlation-based Top-15 subsets indicate that the most favorable feature-selection strategy may vary across assets and architectures.

From a financial forecasting perspective, these findings suggest that expanding the feature space with an increasing number of technical indicators does not necessarily improve predictive performance. At the same time, aggressive feature reduction may remove useful complementary signals, especially for architectures that benefit from broader multivariate representations. The results therefore support a balanced interpretation: explainability-guided feature refinement is valuable for improving model transparency and reducing

https://doi.org/10.3390/math14183402

dimensionality, but its predictive effect should be evaluated empirically for each asset, architecture, and market regime.

The additional market-regime and sudden-jump analyses further show that average forecasting errors should not be interpreted in isolation. Sudden price-jump hours substantially increase prediction errors across all evaluated architectures, confirming that abrupt market movements remain challenging for transformer-based forecasting models. The SHAP–LIME Top-15 representation slightly reduces the average jump-hour error inflation compared with the full feature representation, but this improvement is modest and model-dependent. These findings highlight the importance of evaluating forecasting robustness under non-average market conditions, particularly in highly volatile cryptocurrency markets.

The sensitivity and feature-selection baseline analyses provide further evidence that the Top-15 subset is best interpreted as a conservative dimensionality-reduction strategy rather than a universally optimal feature subset. Top-5, Top-10, Top-15, and Top-20 representations yield broadly comparable results, and the correlation-based Top-15 baseline performs strongly in some settings. Nevertheless, the SHAP–LIME-based subset has the advantage of being model-aware and directly interpretable, since the selected predictors are derived from global and local explanation scores. This makes the proposed approach especially useful when interpretability and feature transparency are important alongside forecasting accuracy.

Overall, the results demonstrate that explainability should not be viewed solely as a post hoc interpretation technique. Instead, integrating SHAP and LIME into the model development process enables the construction of more compact and interpretable forecasting models with competitive predictive performance. The proposed framework therefore contributes not only to transformer-based cryptocurrency forecasting but also to the broader development of explainable machine-learning methodologies for financial time-series analysis.

7. Limitations and Future Work

Despite the encouraging results obtained by the proposed explainability-guided forecasting framework, several limitations should be acknowledged.

First, the study relies exclusively on historical market information represented by OHLCV variables and engineered technical indicators. Although these features capture important aspects of market behavior, they do not encompass the full spectrum of information influencing cryptocurrency prices. External information sources, including macroeconomic announcements, regulatory events, on-chain blockchain metrics, order-book dynamics, financial news, and investor sentiment extracted from social media, were intentionally excluded to isolate the contribution of technical market indicators. Integrating these heterogeneous information sources may further enhance forecasting performance, particularly during periods of elevated market uncertainty.

Second, the proposed framework was evaluated using hourly market data. Hourly forecasting provides a realistic setting for many short-term trading and decision-support applications; however, it is also characterized by considerable market noise and rapidly changing temporal dynamics. Consequently, the conclusions reported in this study should not be interpreted as directly transferable to lower-frequency forecasting problems, such as daily or weekly prediction, where market behavior may exhibit different statistical characteristics.

Third, the experimental evaluation was conducted on five major cryptocurrencies (BTC, ETH, SOL, XRP, and DOGE). Although these assets represent different levels of market capitalization, liquidity, and volatility, the generalizability of the proposed framework to other financial domains remains an open research question. Future studies should investigate its applicability to equities, foreign exchange markets, commodities, exchange-traded funds, and multi-asset portfolios to evaluate the robustness of the explainability-guided forecasting strategy under diverse financial conditions.

Another limitation concerns the explainability techniques employed in this work. The proposed framework utilizes SHAP and LIME because they provide complementary global and local interpretations while remaining model-agnostic. Nevertheless, other explainability approaches, including Integrated Gradients, DeepSHAP, attention-based attribution methods, and counterfactual explanations, may provide additional insights into transformer decision mechanisms. Comparing multiple explainability paradigms within a unified forecasting framework represents an interesting direction for future research.

Furthermore, the present study focuses on deterministic point forecasting. While point predictions are widely adopted in financial forecasting, they do not explicitly quantify predictive uncertainty. Future research may therefore incorporate probabilistic forecasting methods, prediction intervals, or uncertainty-aware deep learning techniques to better support risk-sensitive financial decision making.

The exclusion of non-transformer baselines, including ARIMA, GARCH, LSTM, and GRU, is another limitation of the present study. These models were discussed as part of the methodological background, but the empirical analysis was deliberately restricted to transformer-based architectures in order to provide a controlled comparison within the transformer family. Future work should extend the benchmark by incorporating econometric, recurrent, hybrid, and ensemble baselines to further assess the relative advantages of explainability-guided transformer forecasting.

Finally, recent advances in foundation models and large language models create new opportunities for financial forecasting. Future extensions of the proposed framework may integrate transformer-based time-series models with multimodal financial information, retrieval-augmented knowledge, or large language models capable of incorporating textual financial information alongside numerical market data. Such hybrid architectures may provide more comprehensive representations of financial markets while preserving the explainability and feature refinement strategy proposed in this study.

8. Conclusions

This study demonstrated that explainability can be integrated into transformer-based financial forecasting not only for model interpretation but also as an effective mechanism for explainability-guided feature refinement. Unlike conventional forecasting studies that employ explainability solely for post hoc model interpretation, the proposed framework incorporates explainability directly into the model development process to identify informative predictors and construct compact, asset-specific feature representations. The experimental evaluation was conducted on five major cryptocurrencies (BTC, ETH, SOL, XRP, and DOGE) using a chronological train–validation–test protocol and multiple transformer architectures, including Informer, Autoformer, Reformer, Vanilla Transformer, and Temporal Fusion Transformer.

The experimental results demonstrate that transformer architectures provide competitive forecasting capability for hourly cryptocurrency markets, although their performance varies across assets and feature representations. Under the full feature setting, TFT achieved the strongest overall performance in terms of MAPE across the evaluated assets. After SHAP–LIME Top-15 feature refinement, TFT also retained the strongest post-XAI performance, while the remaining architectures exhibited model- and asset-dependent responses to feature reduction. These findings indicate that architectural design, feature representation, and market characteristics jointly influence forecasting performance.

The explainability-guided feature refinement analysis shows that SHAP–LIME-based feature selection can produce compact and interpretable feature subsets while maintaining competitive forecasting performance in several settings. However, the additional sensitivity, alternative feature-selection, and statistical significance analyses demonstrate that the Top-15 subset should be interpreted as a conservative dimensionality-reduction strategy rather than a universally optimal feature-selection rule. In particular, the effect of feature refinement remains dependent on the cryptocurrency asset, transformer architecture, retained feature subset size, and market conditions.

From a methodological perspective, the principal contribution of this study is the integration of explainability into the forecasting pipeline as an active feature-refinement mechanism rather than a purely interpretive tool. The results indicate that momentum-, trend-, volatility-, and volume-related indicators often play important roles in hourly cryptocurrency forecasting. At the same time, the comparison with random and correlation-based feature-selection baselines shows that explainability-guided feature reduction should be assessed together with simpler selection strategies and statistical variability.

The proposed framework contributes to the growing body of research on explainable artificial intelligence for financial time-series forecasting by demonstrating that model transparency can be enhanced while maintaining competitive predictive performance. Beyond cryptocurrency forecasting, the methodology can be extended to other financial forecasting problems where feature redundancy, model interpretability, and predictive robustness are equally important. It is therefore expected that the integration of explainability-driven feature refinement with advanced deep learning architectures will play an increasingly important role in the development of trustworthy and interpretable intelligent decision-support systems for financial markets.

Author Contributions: Conceptualization, Z.H.K. and E.D.; methodology, Z.H.K. and E.D.; software, E.D.; validation, Z.H.K. and E.D.; formal analysis, Z.H.K.; investigation, Z.H.K. and E.D.; resources, Z.H.K.; data curation, E.D.; writing—original draft preparation, Z.H.K.; writing—review and editing, Z.H.K. and E.D.; visualization, E.D.; supervision, Z.H.K.; project administration, Z.H.K.; funding acquisition, E.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Kocaeli University Technopark (KOÜ Teknopark), Türkiye. The funding was provided as institutional support, and no specific grant number is applicable.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets analyzed during this study were obtained from the publicly available Binance API. The source code and experimental implementation will be made publicly available upon publication of the article. During the peer-review process, they are available from the corresponding author upon reasonable request.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT to assist with language editing, improving academic writing, and refining the organization of the manuscript. The authors reviewed and edited all generated content and take full responsibility for the content of this publication.

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

Use of Artificial Intelligence: Generative AI was used solely to improve the language, grammar, and readability of the manuscript. All scientific content, methodology, analyses, results, interpretations, and final decisions remain the responsibility of the authors.

References

1. Ciaian, P.; Rajcaniova, M.; Kancs, D. The Economics of BitCoin Price Formation. *Appl. Econ.* **2016**, *48*, 1799–1815. [CrossRef]
2. Poongodi, M.; Vijayakumar, V.; Chilamkurti, N. Bitcoin Price Prediction Using ARIMA Model. *Int. J. Internet Technol. Secur. Trans.* **2020**, *10*, 396–406. [CrossRef]

3. Katsiampa, P. Volatility Estimation for Bitcoin: A Comparison of GARCH Models. *Econom. Lett.* **2017**, *158*, 3–6. [CrossRef]
4. Seabe, P.L.; Moutsinga, C.R.B.; Pindza, E. Forecasting Cryptocurrency Prices Using LSTM, GRU, and Bi-Directional LSTM: A Deep Learning Approach. *Fractal Fract.* **2023**, *7*, 203. [CrossRef]
5. McNally, S.; Roche, J.; Caton, S. Predicting the Price of Bitcoin Using Machine Learning. In *Proceedings of the 26th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP), Cambridge, UK, 21–23 March 2018*; IEEE: New York, NY, USA, 2018; pp. 339–343.
6. Lahmiri, S.; Bekiros, S. Chaos, Randomness and Multi-Fractality in Bitcoin Market. *Chaos Soliton. Fract.* **2018**, *106*, 28–34. [CrossRef]
7. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
8. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. In *Advances in Neural Information Processing Systems*; Guyon, I., von Luxburg, U., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S.V.N., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30, pp. 5998–6008.
9. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 11106–11115. [CrossRef]
10. Wu, H.; Xu, J.; Wang, J.; Long, M. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. In *Advances in Neural Information Processing Systems*; Ranzato, M., Beygelzimer, A., Dauphin, Y.N., Liang, P., Vaughan, J.W., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2021; Volume 34, pp. 22419–22430.
11. Kitaev, N.; Kaiser, Ł.; Levskaya, A. Reformer: The Efficient Transformer. In Proceedings of the International Conference on Learning Representations (ICLR), Addis Ababa, Ethiopia, 26–30 April 2020.
12. Lim, B.; Arık, S.Ö.; Loeff, N.; Pfister, T. Temporal Fusion Transformers for Interpretable Multi-Horizon Time Series Forecasting. *Int. J. Forecast.* **2021**, *37*, 1748–1764. [CrossRef]
13. Wu, X.; Yin, X.; Umar, Z.; Iqbal, N. Volatility Forecasting in the Bitcoin Market: A New Proposed Measure Based on the VS-ACARR Approach. *N. Am. J. Econ. Finance* **2023**, *67*, 101948. [CrossRef]
14. Bergsli, L.Ø.; Lind, A.F.; Molnár, P.; Polasik, M. Forecasting Volatility of Bitcoin. *Res. Int. Bus. Finance* **2022**, *59*, 101540. [CrossRef]
15. Ünlü, U.; Bayram, V. Analysis of Bitcoin Volatility during the COVID-19 Pandemic: An Examination Using ARCH and GARCH Models. *J. Res. Econ. Politics Finance* **2024**, *9*, 812–831. [CrossRef]
16. Sözen, Ç. Volatility Dynamics of Cryptocurrencies: A Comparative Analysis Using GARCH-Family Models. *Future Bus. J.* **2025**, *11*, 166. [CrossRef]
17. Gherghina, Ș.-C.; Constantinescu, C.-A. Towards Examining the Volatility of Top Market-Cap Cryptocurrencies Throughout the COVID-19 Outbreak and the Russia–Ukraine War: Empirical Evidence from GARCH-Type Models. *Risks* **2025**, *13*, 57. [CrossRef]
18. Zahid, M.; Iqbal, F.; Koutmos, D. Forecasting Bitcoin Volatility Using Hybrid GARCH Models with Machine Learning. *Risks* **2022**, *10*, 237. [CrossRef]
19. Guirguis, M.Z. Application of a TGARCH, EGARCH, and PARCH Models to Test the Volatility Clusters of Shiba, Bitcoin and Ethereum Digital Cryptocurrencies. *SSRN Electron. J.* **2024**. [CrossRef]
20. Teker, D.L.; Teker, S.; Gümüştepe, E.D. Determinants of Bitcoin Price Movements. *Press. Procedia* **2024**, *19*, 75–78. [CrossRef]
21. Phung Duy, Q.; Nguyen Thi, O.; Le Thi, P.H.; Pham Hoang, H.D.; Luong, K.L.; Nguyen Thi, K.N. Estimating and Forecasting Bitcoin Daily Prices Using ARIMA-GARCH Models. *Bus. Anal. J.* **2024**, *45*, 11–23. [CrossRef]
22. Ustaoğlu, E. Return and Volatility Spillover between Cryptocurrency and Stock Markets: Evidence from Turkey. *J. Account. Finance* **2022**, *93*, 117–126. [CrossRef]
23. Bouteska, A.; Abedin, M.Z.; Hajek, P.; Yuan, K. Cryptocurrency Price Forecasting: A Comparative Analysis of Ensemble Learning and Deep Learning Methods. *Int. Rev. Financ. Anal.* **2024**, *92*, 103055. [CrossRef]
24. Qureshi, S.M.; Saeed, A.; Farooq, A.; Khattak, A.R.; Almotiri, S.H.; Al Ghamdi, M.A.; Rukh, M.S. Evaluating Machine Learning Models for Predictive Accuracy in Cryptocurrency Price Forecasting. *PeerJ Comput. Sci.* **2025**, *11*, e2626. [CrossRef]
25. Kaur, R.; Uppal, M.; Gupta, D.; Juneja, S.; Arafat, S.Y.; Rashid, J.; Kim, J.; Alroobaea, R. Development of a Cryptocurrency Price Prediction Model: Leveraging GRU and LSTM for Bitcoin, Litecoin and Ethereum. *PeerJ Comput. Sci.* **2025**, *11*, e2675. [CrossRef]
26. Rodrigues, F.; Machado, M. High-Frequency Cryptocurrency Price Forecasting Using Machine Learning Models: A Comparative Study. *Information* **2025**, *16*, 300. [CrossRef]
27. 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]
28. Tiwari, D.; Bhati, B.S.; Nagpal, B.; Alturki, N.; Bayisenge, L. Attention-Augmented Hybrid CNN–LSTM Model for Social Media Sentiment Analysis in Cryptocurrency Investment Decision-Making. *Sci. Rep.* **2025**, *15*, 33201. [CrossRef]
29. Nouira, A.Y.; Bouchakwa, M.; Amara, M. Role of Social Networks and Machine Learning Techniques in Cryptocurrency Price Prediction: A Survey. *Soc. Netw. Anal. Min.* **2024**, *14*, 152. [CrossRef]
30. Ross, M.; Parker, L.; Cooper, E.; Reed, J.; Watson, R.; Howard, N.; Emmanuel, M. Predicting Market Trends Using Sentiment Analysis on User-Generated Content. *SSRN* **2025**, 5524218.

31. Somayajulu, M.V.N.S.S.R.K.S.; Ahmed, M.; Kotaiah, B. Bitcoin Price Prediction Using LSTM and CNN. In Proceedings of the International Conference on Innovative Computing & Communication (ICICC 2024), New Delhi, India, 16–17 February 2024.
32. Lin, Y.; Cai, C. CNN–Attention–LSTM with Bayesian Optimization for Multi-Level Sump Well Anomaly Early Warning. *Mathematics* **2026**, *14*, 1528. [CrossRef]
33. Farooq, A.; Uddin, M.I.; 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]
34. Lee, M.C. Temporal Fusion Transformer-Based Trading Strategy for Multi-Crypto Assets Using On-Chain and Technical Indicators. *Systems* **2025**, *13*, 474. [CrossRef]
35. Díaz Berenguer, A.; Da, Y.; Bossa, M.N.; Oveneke, M.C.; Sahli, H. Causality-Driven Multivariate Stock Movement Forecasting. *PLoS ONE* **2024**, *19*, e0302197. [CrossRef]
36. Stefaniuk, F.; Ślepaczuk, R. A Neural Network Informer in Algorithmic Investment Strategies on High-Frequency Bitcoin Data. *Expert Syst. Appl.* **2026**, *317*, 131599. [CrossRef]
37. Niu, Y. Forecasting Logarithmic Returns on Bitcoin Based on the Informer Model. In *Proceedings of the 2024 International Conference on Digital Economy and Computer Science (DECS '24)*; Association for Computing Machinery: New York, NY, USA, 2024; pp. 255–260.
38. Younas, R.; Rehman, H.M.R.U.; Choi, G.S. Crypto Foretell: A Novel Hybrid Attention-Correlation Based Forecasting Approach for Cryptocurrency. *J. Big Data* **2025**, *12*, 229. [CrossRef]
39. Khaniki, M.A.L.; Manthouri, M. Enhancing Price Prediction in Cryptocurrency Using Transformer Neural Network and Technical Indicators. *arXiv* **2024**, arXiv:2403.03606.
40. Singh, S.; Bhat, M. Transformer-Based Approach for Ethereum Price Prediction Using Crosscurrency Correlation and Sentiment Analysis. *arXiv* **2024**, arXiv:2401.08077.
41. Mahdi, E.; Martin-Barreiro, C.; Cabezas, X. A Novel Hybrid Approach Using an Attention-Based Transformer + GRU Model for Predicting Cryptocurrency Prices. *Mathematics* **2025**, *13*, 1484. [CrossRef]
42. 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]