Received: 18 January 2026 | Revised: 13 April 2026 | Accepted: 23 July 2026 | Published online: 26 August 2026

Journal of Computational and Cognitive Engineering
2026, Vol. 00(00) 1–21
DOI: 10.47852/bonviewJCCE62029131

RESEARCH ARTICLE

VMD-RDIC-DL: A Composite Relevance-Driven Hybrid Decomposition and Deep Learning Framework for Cryptocurrency Forecasting

Maryam Maatallah[1,*], Mourad Fariss[2], Hakima Asaidi[1] and Mohamed Bellouki[1]

[1] *Laboratory of Applied Mathematics and Information Systems (LMASI), Mohammed First University, Morocco*

[2] *Chemistry, Computer Science and Artificial Intelligence Research Team (ERCI2A), Abdelmalek Essaâdi University, Morocco*

Abstract: This study proposes a framework combining Variational Mode Decomposition (VMD) with a relevance-driven selection process to reduce noise and redundancy in financial time-series forecasting. The original time series is decomposed by VMD into intrinsic mode functions (IMFs), which are then evaluated using three relevance metrics: relative energy contribution, mutual information, and Spearman's rank correlation coefficient. These metrics identify the IMFs most strongly associated with future price movements. As opposed to conventional VMD-based approaches that treat all IMFs equally, the proposed relevance-driven selection process adapts IMF selection to the statistical properties of the analyzed market, thereby improving model generalization across different volatility conditions and forecasting horizons. This study makes three main contributions: (i) developing a relevance-driven IMF selection strategy to overcome limitations of traditional VMD methods, (ii) designing a hybrid framework that integrates multiscale decomposition with nonlinear information filtering, and (iii) conducting a comprehensive empirical evaluation of the proposed models. Experiments on hourly Bitcoin (BTC)/USD data from 2018 to 2025 show that the VMD-RDIC-deep learning models achieves strong forecasting performance. The results show that the proposed relevance-driven decomposition framework improves prediction accuracy and robustness compared with traditional statistical models, including Autoregressive Integrated Moving Average (ARIMA), as well as machine learning and deep learning approaches, highlighting its suitability for complex and volatile financial markets.

Keywords: Bitcoin forecasting, variational mode decomposition, deep learning, financial time series, cryptocurrency markets

1. Introduction

Traditional financial markets have slower-moving prices and behaviors than Bitcoin, which has quick price swings and quick movements in the way it behaves in the marketplace. Due to the nature of the sudden movement in the price and behavior of Bitcoin, the raw price data cannot be analyzed using the data as is. This results in a very nonstationary time-series signal with both long-term and short-term cycles and significant amounts of noise, making it extremely challenging for conventional methods to model or forecast. Linear approaches such as ARIMA [1] or GARCH [2] often fail to capture the underlying nonlinear structure, and many classical machine learning methods remain sensitive to noise and break down during unexpected market reversals [3].

Deep learning (DL) models have shown clear advantages for modeling nonlinear and long-range dependencies in financial time series [4, 5]. However, their predictive success depends critically on the quality of the input representation. Raw Bitcoin prices contain overlapping frequency components and embedded noise that degrade training stability, blur temporal relations, and increase the risk of overfitting [6]. This has motivated the development of hybrid forecasting frameworks that combine signal decomposition with neural networks to extract more interpretable and better-structured inputs [7, 8].

Among the many multiscale decomposition methods available, variational mode decomposition (VMD) has become one of the top-performing methods for decomposing a complex nonstationary signal into intrinsic mode functions (IMF) with specific frequency characteristics [9, 10]. VMD has several advantages over previous methods like Empirical Mode Decomposition (EMD) [11] including reduced mode mixing and improved stability in the generated subseries. However, a major limitation exists in that the importance of each IMF is typically assumed equal, while many of the IMFs will likely include redundant information or noise-dominated components. When these less important IMFs are used in neural models, it can lead to poor performance,

***Corresponding author:** Maryam Maatallah, Laboratory of Applied Mathematics and Information Systems (LMASI), Mohammed First University, Morocco. Email: maryam.maatallah.d23@ump.ac.ma

increased computation time, and decreased ability to generalize [12, 13].

Recently, researchers have employed information-theoretic metrics to determine the level of nonlinear dependence between each IMF and the target variable. Unlike other methods of forecasting based on VMD, which are typically based on a single measure of relevance (i.e., correlation, energy, or mutual information), the proposed relevance-driven IMF combination (RDIC) strategy uses a unified scoring system for three complementary measures of relevance to identify informative IMFs. Each measure captures one specific aspect of the relationship between each decomposed component and the target signal: mutual information (nonlinear dependency), Spearman rank correlation (monotonic association), and relative energy (structural contribution to the variance in the signal). The combination of these relevance measures can provide a more accurate selection of informative IMFs and reduce the likelihood of selecting noisy or weakly predictive IMFs. The use of multiple criteria to evaluate relevance is a significant methodological contribution of this study. While these efforts have been successful, there are still limitations to their application. Most studies have focused on the use of a single measure of relevance or employed these measures solely for the purpose of descriptive analysis; therefore, limiting the ability to use them as part of an integrated selection mechanism as part of a broader forecasting pipeline. Furthermore, most of the prior research employing VMD as part of a hybrid model has utilized datasets from stationary or energy-related systems, thus creating a clear methodological void when it comes to developing methodologies for highly volatile systems, such as the Bitcoin system [14].

Given this background, the current study proposes a unified forecasting framework that incorporates VMD, along with a novel selection mechanism based on the relevance of the IMFs, and employs various state-of-the-art DL architectures. The RDIC mechanism evaluates three separate measures of relevance: mutual information (MI), Spearman rank correlation coefficient, and the relative energy content of each IMF to select those that have the highest predictive power. This enables the generation of a more accurate and clean representation of the Bitcoin price series and, ultimately, improves the learning stability of the models and minimizes the transmission of noise from the input to the output. Once these informative IMFs are identified, they are inputted into several types of Recurrent Neural Models (Autoregressive Neural Networks [ARNNs], Long Short-Term Memory [LSTMs], gated recurrent units [GRUs]), allowing for a thorough examination of the hybrid pipeline under a variety of architectures.

The effectiveness of the VMD-RDIC-DL framework was evaluated utilizing hourly BTC/USD historical data spanning 2018–2025, which included multiple bull and bear cycles, periods of high volatility, and periods of consolidation. The results of the experiment demonstrated significant improvements in terms of Root Mean Square Error (RMSE), Mean Absolute Error (MAE), R^2, and directional accuracy (DA) when compared to both traditional machine learning algorithms and baseline deep models, demonstrating the value added by integrating multiscale decomposition with information-based feature selection.

The remainder of this article is organized as follows. Section 2 reviews related work on decomposition-based forecasting and DL models for financial time series. Section 3 presents the proposed VMD-RDIC-DL methodology. Section 4 describes the experimental setup and implementation details. Section 5 discusses empirical results and comparative analysis. Section 6 concludes the study and outlines future directions.

2. Literature Review

The results from recent studies in this area, such as those reported in Fakhfekh and Jeribi [15] as well as Bouteska et al. [16], confirm that traditional modeling techniques for time-series data such as ARIMA and for the estimation of volatility (parametric volatility models such as Autoregressive Conditional Heteroskedasticity (ARCH) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH)) are unable to capture abrupt regime shifts and short-term price movements characteristic of the behavior of cryptocurrency markets [1]. The studies report the failure of traditional methods to model the highly nonlinear and irregular dynamics of cryptocurrency prices, which is a direct consequence of the frequent and extreme price swings observed in these markets [17, 18]. Other researchers have shown that machine learning models produce inconsistent forecasting performance across different cryptocurrencies and market conditions and that no single model consistently outperforms others across all cryptocurrency markets and forecasting horizons [16, 19]. Therefore, these findings highlight the structural complexity, nonstationarity, and extreme heterogeneity inherent in cryptocurrency time-series data.

2.1. Hybrid models combining signal decomposition and deep learning

Statistical models have several drawbacks; for example, they don't learn the temporal relationships that exist between values in time-series data, and their performance can be limited by the quality of the input data. As a result, researchers have started using deep learning architectures including ARNNs, LSTMs, and GRUs [20], which can model temporal relationships in the data and therefore capture both linear and nonlinear patterns in sequences of data. Several recent studies have demonstrated that hybrid forecasting frameworks combining VMD with advanced learning architectures significantly enhance predictive performance when dealing with complex and highly volatile time series. Lv et al. [9] proposed a hybrid forecasting model that uses VMD to decompose the load signal and then uses an LSTM network to forecast each of these components. They were able to produce a better and more consistent short-term load forecast than classical statistical and DL methods by choosing a suitable number of modes through permutation entropy and applying seasonal corrections. Niu et al. [21] developed a hybrid forecasting model combining VMD with an attention-based GRU network for stock price index prediction. By decomposing the original time series into multiscale components prior to prediction, their model effectively captures complex nonlinear temporal patterns and significantly improves forecasting accuracy and robustness under highly volatile conditions.

However, DL models remain limited by the quality of raw input data, as cryptocurrency price series inherently contain multiple overlapping frequency components, including short-term volatility and long-term trends. This multi-frequency entanglement produces noisy training signals, making learning and prediction more difficult and limiting generalization capability. Several signal processing techniques have been developed to decompose financial time series into their constituent parts so that they may be preprocessed prior to use in a model. Two of these techniques are EMD and VMD; each has demonstrated considerable success in identifying stable frequency components within nonstationary time series. Specifically, VMD was first introduced by Dragomiretskiy and Zosso [22] and decomposes a time series into a finite set of IMFs, each corresponding to a specific frequency band. This formulation enables an effective separation of high-frequency noise from low-frequency trend components within the same time series. Consequently, VMD provides a denoised and frequency-resolved representation of market dynamics, which can significantly enhance the quality of inputs used for forecasting models. Zhang et al. [23] proposed a CNN-BiLSTM-Attention model that extracts local patterns with CNN, learns bidirectional temporal dependencies with BiLSTM, and highlights key features through an attention layer. This hybrid design consistently achieves higher forecasting accuracy than LSTM, CNN-LSTM, and CNN-LSTM-Attention across 12 Chinese and international stock indices. Wu and Cai [24] proposed a short-term load forecasting model combining VMD, LSTM, and Improved Sparrow Search Algorithm (ISSA)-based hyperparameter optimization, achieving significant error reductions on Australian and Tetouan datasets. Despite these advances, most VMD-based hybrid models apply decomposition without selective filtering of IMFs, leading to the inclusion of weakly informative or noise-dominated components. Furthermore, decomposition parameters are often predefined or empirically selected, limiting adaptability to different market regimes. This work has initiated a new generation of hybrid VMD-DL models. Soomro et al. [25] applied a VMD-LSTM architecture to electric load forecasting, incorporating seasonal corrections and error compensation, which resulted in a significant reduction in RMSE as compared to the classic EMD-LSTM approach. Kazanci and Shukuya [26] developed VMD-BiLSTM models where each IMF was forecasted independently before being combined into a single output, which improved both convergence speed and temporal stability.

In the financial domain, Sun et al. [27] demonstrated the effectiveness of a VMD-GRU hybrid framework for modeling Bitcoin market volatility. Their empirical results show that the proposed architecture achieves forecasting accuracy improvements exceeding 16% compared with both classical machine learning and conventional DL models. However, these approaches still lack a unified relevance-driven filtering mechanism, and their performance remains sensitive to noisy or redundant decomposed components, particularly in highly volatile cryptocurrency environments.

2.2. Information-theoretic and adaptive approaches

To counteract the limitations discussed above, recent research has increasingly employed information-theoretical methods to identify the most relevant sub-components obtained through signal decomposition. In particular, Shan et al. [28] employed MIC-based feature selection combined with a two-stage decomposition strategy using CEEMDAN and VMD, followed by a CNN-BiLSTM predictor, to significantly improve wind power forecasting accuracy. Similarly, Zhang and Duan [29] extended informer-based architectures through the use of a combination of VMD, MIC, and feature engineering to develop the enhanced LFTSformer, a framework designed specifically for long-horizon financial forecasting. These results confirmed that the integration of multiscale decomposition techniques, with information-guided feature extraction methods, can improve the capture of long-term temporal relationships, while reducing the amount of residual noise present in the data.

More recently, Yuan et al. [30] developed a VMD-Informer-BiLSTM (Efficient Additive Attention (EAA)) model that incorporates enhanced attention mechanisms and adaptive aggregation strategies to capture the complex structure of multiscale signals. Their results demonstrate that the proposed model achieves a high level of long-term forecasting stability. However, the authors also point out that no selective filtering mechanism is applied prior to the learning stage, which limits the model's effectiveness when dealing with highly noisy data.

2.3. Beyond deep learning: Alternative ML paradigms for nonlinear forecasting

In addition to recurrent DL architectures, recent research has explored a broader range of machine learning paradigms for modeling nonlinear and regime-dependent time series. Conventional neural network models, including feedforward and multilayer structures, have been widely applied in forecasting tasks across several domains. Previous studies have demonstrated their ability to capture complex nonlinear relationships in applications such as agricultural production forecasting, commodity price prediction, and financial market analysis [31].

Beyond deterministic neural models, probabilistic learning approaches have also gained increasing attention. In particular, Gaussian Process Regression (GPR) provides a flexible Bayesian framework capable of modeling nonlinear relationships while simultaneously quantifying predictive uncertainty [32]. This property is especially valuable in financial forecasting, where uncertainty estimation plays an important role in risk management and decision-making processes. Recent studies have shown that GPR-based models can effectively capture volatility dynamics and nonlinear patterns in financial and economic time series [33, 34].

Another emerging research direction involves graphical and structural modeling techniques aimed at identifying dependencies among interacting variables. Graph-based representations enable the analysis of complex relationships within economic and financial systems and facilitate the understanding of information propagation across interconnected variables. Such approaches can improve model interpretability and provide complementary insights into the structure of time-series forecasting problems.

In addition, ensemble forecasting strategies have demonstrated strong performance in noisy environments by combining multiple predictive models. Techniques such as boosting, stacking, and hybrid ensemble learning have been widely applied in economic forecasting and financial prediction tasks, often improving robustness by reducing both prediction variance and model bias [35, 36].

Although these approaches provide complementary perspectives for modeling complex temporal dynamics, the present study focuses on a decomposition-guided learning framework for cryptocurrency forecasting. The proposed relevance-driven IMF selection mechanism is designed to remain model-agnostic and could potentially be integrated with probabilistic, ensemble, or graph-based predictors. This property allows the framework to remain compatible with a wide range of machine learning paradigms while maintaining its focus on improving signal decomposition and feature relevance.

2.4. Comparative analysis of existing studies

For an understanding of the development of hybrid forecasting methods that incorporate signal decomposition, information-theoretic model selection, and DL, a comparison of the most

Table 1
Comparative synthesis of hybrid VMD-enhanced methods: methodology, performance, and limitations

Reference	Methodology	Main objective	Key results	Identified limitations
[37]	VMD-LSTM	Short-term electric load forecasting	≈20% RMSE reduction compared to EMD-LSTM; better stability	No IMF relevance evaluation, performance sensitive to VMD parameters
[27]	VMD-GRU	Bitcoin volatility forecasting	R^2 improved from 0.97 to 0.985; 16% error reduction	IMFs treated equally, limited noise suppression
[25]	VMD-LSTM	Electric load forecasting with seasonal correction	Significant RMSE reduction and better handling of seasonal effects	No nonlinear feature relevance screening
[38]	VMD-LSTM-Seasonal correction	Short-term load forecasting	Higher accuracy and stability compared with statistical and deep models	Fixed mode number; no adaptive IMF selection
[24]	VMD-LSTM-ISSA optimization	Short-term power load forecasting	Lower error metrics on Australian and Tetouan datasets	Hyperparameter search dependent on meta-heuristic stability
[39]	EMD/VMD-RGARCH	Bitcoin volatility modeling	VMD-RGARCH with jump-robust estimators yields best volatility forecasts	High computational cost and model sensitivity
[23]	CNN-BiLSTM-Attention	Stock index forecasting	Outperforms LSTM, CNN-LSTM, CNN-LSTM-Attention on 12 indices	No decomposition stage; sensitive to high-frequency noise
[40]	VMD-MIC-BiLSTM	Nonlinear financial forecasting	Lower RMSE/MAE and faster convergence; selective IMF retention	Effective mainly for moderately volatile series
[29]	VMD-MIC-Informer-FE	Long-horizon stock forecasting	R^2 = 0.994; strong multiscale noise control	High computational cost; Transformer tuning required
[30]	VMD-Informer-BiLSTM (EAA)	Handling complex multiscale temporal patterns	Better long-horizon stability and accuracy	No filtering of irrelevant IMFs before training
[14]	VMD-TMFG-LSTM	Multi-feature stock forecasting	Noticeable error reduction and better feature structure	Not tested on cryptocurrency markets

recently published studies can be found in Table 1. As evident in the table, there has been a significant growth in the number of hybrid models developed using these techniques during the last couple of years with respect to both methodologies used and forecast accuracy achieved. The table also highlights the limitations of previously developed models, specifically their lack of selective filtering and the need for predetermined decomposition parameters that motivated the development of the proposed VMD-RDIC-DL model. Using multiple DL models (ARNN, GRU, and LSTM) and a unifying preprocessing methodology, the proposed model builds upon the strengths of previously developed models and mitigates their structural constraints.

3. Research Methodology

This study proposes a unified hybrid forecasting framework designed to address the nonstationary and nonlinear characteristics of Bitcoin price dynamics. The methodology is structured as a sequential pipeline that integrates signal decomposition, relevance-driven component selection, and DL-based prediction. First, the original price series is decomposed into multiple IMFs using VMD, enabling a multiscale representation of underlying temporal patterns. Next, an RDIC strategy is employed to identify and retain only the most informative components by jointly considering nonlinear dependency, monotonic association, and energetic contribution. The selected IMFs are then reconstructed into a filtered signal that serves as input to forecasting models. Finally, three complementary DL architectures–ARNN, LSTM, and GRU—are trained and evaluated on the reconstructed series. This structured approach aims to enhance predictive stability, reduce noise-induced degradation, and improve forecasting accuracy in highly volatile financial time series.

3.1. Multiscale signal decomposition with VMD

First, normalize the Bitcoin price data with min–max normalization for numerical stability in order to analyze it. They use a technique called VMD that breaks down the normalized signal into IMFs. Each IMF captures oscillations at a unique frequency range; therefore, the VMD decomposition gives them the following representation of $x(t)$:

$$x(t) = \sum_{k=1}^{K} u_k(t) \tag{1}$$

where $u_k(t)$ denotes the k^{th} IMF and K is the total number of decomposed modes.

The core principle of VMD is to determine each $u_k(t)$ and its corresponding ω_k such that all modes are compactly distributed around their center frequencies in the spectral domain. This is formulated as the following constrained variational problem in Equation (2), where $\delta(t)$ refers to the Dirac delta function, "*" denotes the convolution operator, ∂_t represents the temporal derivative, and $x(t)$ corresponds to the financial data sequence.

$$\begin{cases} min \sum_{k=1}^{k} \left\| \partial_t \left[\left(\delta(t) + \frac{i}{\pi t} \right) \times u_k(t) \right] e^{\{-i\,\omega_k t\}} \right\|_2^2 \\ s.t.\ \sum_{k=1}^{K} u_k(t) = x(t) \end{cases} \tag{2}$$

The decomposition isolates low-frequency structures (trend dynamics), medium-frequency oscillations (cyclical behavior), and high-frequency variations (short-term volatility spikes). This step reduces nonstationarity and produces a cleaner representation for subsequent analysis.

3.2. Relevance-driven IMF combination (RDIC)

Although VMD produces an effective multiresolution representation for the Bitcoin price signal, it also decomposes into many IMFs which contribute differently to the predictive performance of the forecast. In some cases, the IMFs may have redundant information; they may show little dependency on the future price of the asset or be comprised of random fluctuations. If these additional IMFs are included without discrimination, they will likely degrade the performance of the model and limit its ability to generalize. Therefore, the objective of this research is to develop a relevance-based IMF selection strategy.

In this study, the RDIC strategy has been developed to assess the relative contribution of each IMF by use of complementary statistical and information-theoretic measures that evaluate the nonlinear dependence, the monotonic relationship, and the energy content of each $\mathrm{IMF}_k(t)$ to the original normalized price series $x(t)$.

3.2.1. Mutual information (MI)

MI is employed to quantify the nonlinear dependency between $\mathrm{IMF}_k(t)$ and $x(t)$. MI is estimated using a nonparametric regression-based approach, allowing the detection of complex relationships that are not captured by linear correlation measures. A higher MI value indicates stronger informational relevance for prediction, expressed as:

$$MI(IMF_k; x) = \sum_i \sum_j p(i,j) \log\left(\frac{p(i,j)}{p(i)p(j)} \right) \tag{3}$$

MI is estimated using a nonparametric regression-based estimator. Since MI is unbounded, it is normalized across all IMFs:

$$MI_k^* = \frac{MI_k - \min_j MI_j}{\max_j MI_j - \min_j MI_j}. \tag{4}$$

3.2.2. Spearman rank correlation

The Spearman rank correlation is a statistical measure used to evaluate the strength of a monotonic relationship between two variables. Spearman relies on the ranking of values. By working with ranks rather than with raw numerical values, it becomes less sensitive to extreme observations and can capture relationships that increase or decrease in a consistent way, even when the shape of the relationship is not linear. To compute the coefficient, each value is replaced by its rank in the corresponding series. If X_i and Y_i denote the two series and $R(X_i)$ and $R(Y_i)$ their ranks, the coefficient is given by:

$$\rho_s(k) = 1 - \frac{6 \sum_{t=1}^{T} \left[R(IMF_k(t)) - R(x(t)) \right]^2}{T(T^2 - 1)} \tag{5}$$

Its absolute value is normalized:

$$\rho_k^* = \frac{|\rho_s(k)| - \min_j |\rho_s(j)|}{\max_j |\rho_s(j)| - \min_j |\rho_s(j)|}. \tag{6}$$

The result varies between -1 and 1. A value close to 1 indicates that the two series increase together in a regular way. A value close to -1 shows that one increases while the other decreases. A value near zero means that no clear monotonic pattern appears. Because it does not require linearity and is robust to irregular fluctuations, Spearman rank correlation is often used in the study of time-dependent signals and financial data and in the selection of relevant features after decomposing a signal into several components.

3.2.3. Relative energy

Relative energy quantifies the contribution of an IMF to the overall variance of the decomposed signal. For a given IMF $u_k(t)$, its energy is defined as the sum of its squared amplitudes:

$$E_k = \sum_{t=1}^{T} u_k(t)^2 \tag{7}$$

Since raw energy values may differ significantly across IMFs, a min–max normalization is applied to obtain a comparable measure:

$$RE_k^* = \frac{E_k - \min(E)}{\max(E) - \min(E)} \tag{8}$$

This normalized relative energy reflects how strongly each IMF contributes to the global oscillatory structure of the signal while preventing high-amplitude modes from dominating the selection process. IMFs with very low relative energy typically correspond to weak or noise-dominated fluctuations, whereas IMFs with higher energy capture more substantial dynamic components.

3.2.4. Combined relevance score

To identify the most informative IMFs, a composite relevance indicator was constructed by integrating three complementary criteria: nonlinear dependence, monotonic association, and relative energy contribution. For each IMF k, the combined relevance score is defined as:

$$RDICScore_k = w_{MI}\, MI_k^* + w_{Sp} |\rho_k^*| + w_{En}\, RE_k^* \tag{9}$$

MI is given the largest weight since it will measure all types of dependency between the various IMFs and the target series, especially the nonlinear ones. Nonlinear dependency is also the type of dependency that is most likely to exist in cryptocurrency markets where price action has been shown to have many different forms of nonlinearity; that is, there are many different forms of price action in cryptocurrency markets that cannot be described using either linear or monotonic models of price action. Spearman's rank correlation is used as another criterion that measures whether there are any monotonic relationships and/or if the relationship is one directional in the two time series. While Spearman's rank correlation is less effective at detecting nonlinear relationships between the two time series than MI, it can detect whether the individual IMF follows the same direction of change as the original time series. The relative energy component is used to determine how much each of the individual IMFs contributes to the total variance of the original time series. Energy is a good indicator of the size of the fluctuation of each IMF. However, having high energy does not mean that the IMF is highly relevant for prediction purposes; this is true since some of the most energetic IMFs could be capturing large-scale movements or noise in the original time series. Therefore, we use the relative energy with a smaller weight than the other two criteria so that we do not select IMFs that capture primarily noise in the original time series but still capture significant portions of the variance in the original time series.

The weighting configuration $(0.5, 0.3, 0.2)$ was determined during a preliminary validation phase in which several alternative combinations were tested. The selected configuration provided the most favorable compromise between predictive performance and stability of the IMF ranking across the dataset covering the period 2018–2025.

To ensure that the weighting scheme was not arbitrarily selected, a sensitivity analysis was conducted on a validation subset obtained from the training data using a chronological training-validation split. Several weight combinations satisfying the constraint.

$$w_{MI} + w_{SP} + w_{EN} = 1 \tag{10}$$

Figure 1 illustrates the sensitivity analysis of the RDIC weighting parameters. The heatmap reports the validation RMSE obtained for different combinations of (w_{MI}, w_{SP}, w_{EN}), while the color intensity reflects the forecasting performance. The results indicate that the prediction error remains relatively stable across a wide range of weight combinations, suggesting that the RDIC mechanism is robust to moderate variations of the weighting parameters. The configuration $(0.5, 0.3, 0.2)$, which emphasizes MI while maintaining contributions from rank correlation and signal energy, consistently yields competitive performance within the lowest RMSE region and lies within a region of stable performance.

3.2.5. IMF ranking and selection

Therefore, these values provide a preference for nonlinear dependencies as well as monotonic associations, which are both relevant to the forecasting tasks, while using energy as an indicator that is secondary to ranking and to remove noisy components from data without allowing variance to be the primary factor for determining the scores. Therefore, all IMFs will be scored based on the combination of their composite score in descending order:

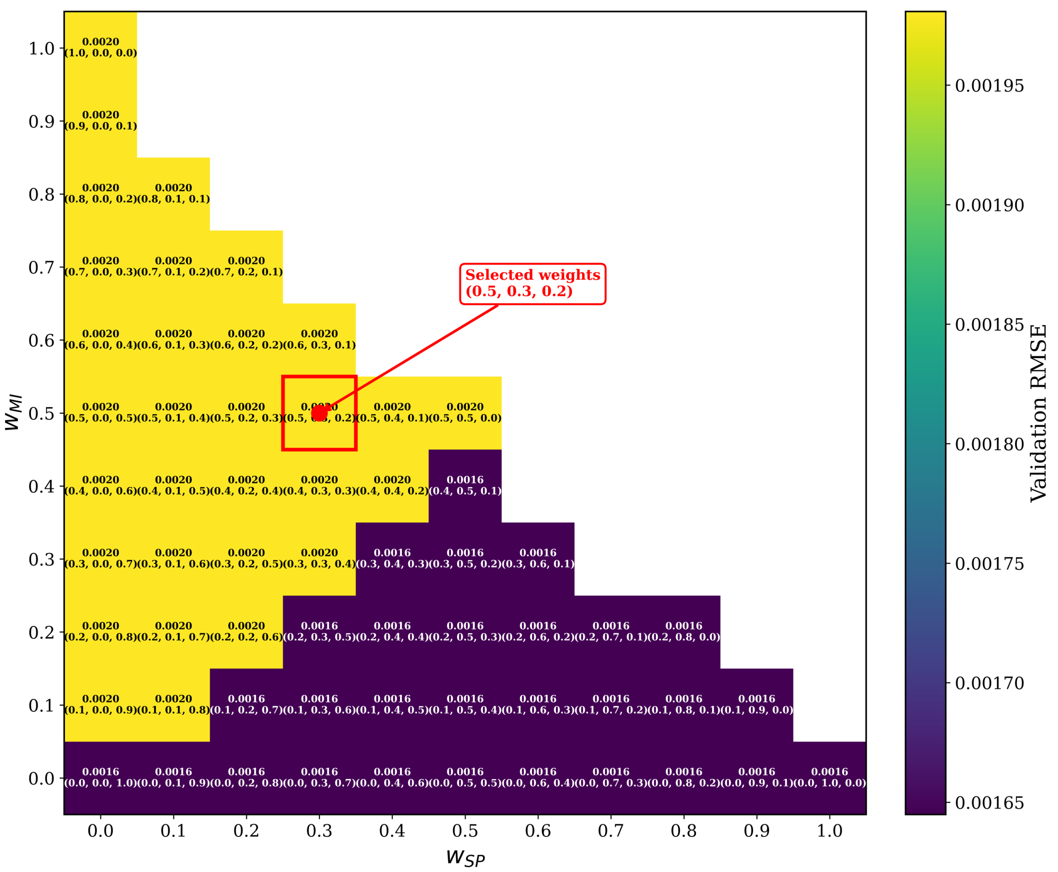
$$\text{Ranked IMFs} = \operatorname{argsort}\left(RDICScore_k\right)\big|_{\text{descending}} \tag{11}$$

The top three IMFs with the highest scores are selected for reconstruction: $\mathcal{S} = \{k_1, k_2, \ldots, k_n\}$, and the filtered signal is obtained by summing only the selected IMFs:

$$x_{\text{filtered}}(t) = \sum_{k \in \mathcal{S}} u_k(t) \tag{12}$$

The processing stage is used to select the most meaningful (long term) as well as the appropriate frequency content

Figure 1
RDIC weights sensitivity analysis with weight triplets



(intermediate as well as high frequency), from a broad spectrum of information, and to eliminate the “noise” or less important information which could potentially degrade the performance of the predictive model. The experimental results have shown that this processing stage has improved the predictive model’s ability to provide stable predictions and to reduce the effect of unwanted frequency components.

3.3. Forecasting model deep learning modeling

The selected IMFs are reconstructed into a single time series (filtered) for use as the model’s input. As opposed to training separate networks for each IMF, a single ARNN, LSTM, and GRU model is trained on this single time series to capture all the individual selected modes in an efficient manner.

3.3.1. ARNN model

The ARNN [41] (Figure 2) is a lightweight nonlinear forecasting model that builds on the idea of using past observations to predict future values, but without relying on recurrent connections. Instead of storing an internal memory state, the ARNN directly feeds a sequence of lagged values into a small neural structure, which allows it to model nonlinear variations that classical autoregressive models cannot capture on their own.

In practice, the model receives the last p observations of the series as input. These lagged values are transformed by one or several hidden units that extract nonlinear patterns. The output neuron then generates the next predicted value. A simplified representation of this mechanism can be expressed as:

$$h = \phi(W_x + b) \tag{13}$$

$$\hat{y}_t = w^\top h + c \tag{14}$$

where $x_t = [y_{t-1}, \ldots, y_{t-p}]$ regroups the lagged inputs, $\phi(\cdot)$ is a nonlinear activation function, and W, w, b, c are the trainable parameters of the model.

Because ARNN does not include gates or recurrent loops, its training procedure is lighter and does not face the same stability issues as recurrent neural networks (RNNs). This simplicity can be an advantage when the objective is to learn short-term irregularities or local nonlinearities rather than long-range temporal dependencies. However, the absence of an internal memory also means that ARNN is less suited to representing long-term structure compared with LSTM or GRU.

3.3.2. LSTM model

Because it can learn long-term relationships and temporal relationships in sequence data, the Long Short-Term Memory (LSTM) model is one of the most commonly used types of RNN models for time-series prediction. In general, Figure 3 illustrates the common structure of an LSTM cell that uses memory units

Figure 2
Diagram of an Autoregressive Neural Network (ARNN)

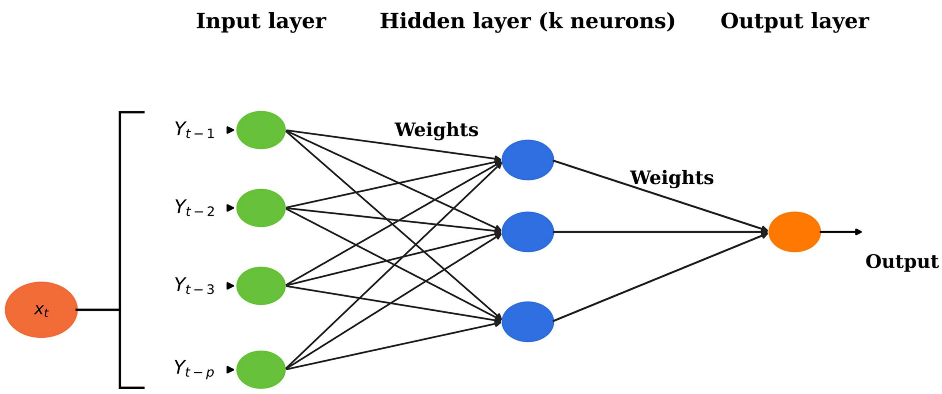
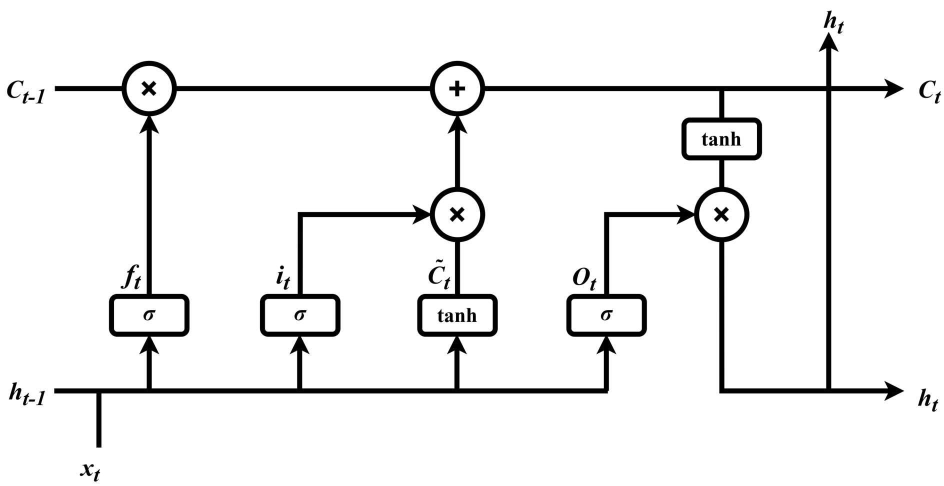


Figure 3
Internal structure of an LSTM cell illustrating the forget, input, and output gates



and gates to manage the amount of information flowing through the model at each time step. The typical process for making predictions using an LSTM generally consists of the following three processes [42]:

The forgetting process: All information from previous time steps that is deemed irrelevant or unnecessary is removed based on the output of the Forget Gate.

The selective memory process: All relevant information is stored/updated in the hidden memory units of the LSTM based on the output of the Input Gate.

Output stage: The last hidden state of the LSTM is determined and used to generate a prediction by the model. This process has been mathematically defined in detail in the equations presented above and was discussed previously by Zhang et al. [14].

$$f^{(t)} = \sigma\left(w_f h^{(t-1)} + w_f x^{(t)} + b_f\right) \tag{15}$$

$$i^{(t)} = \sigma\left(w_i h^{(t-1)} + w_i x^{(t)} + b_i\right) \tag{16}$$

$$c^{(t)} = \tanh\left(w_c h^{(t-1)} + w_c x^{(t)} + b_c\right) \tag{17}$$

$$C^{(t)} = C^{(t-1)}.f^{(t)}.\,c^{(t)} \tag{18}$$

$$o^{(t)} = \sigma\left(w_o h^{(t-1)} + w_o x^{(t)} + b_o\right) \tag{19}$$

$$h^{(t)} = o^{(t)}.\,tanh(C^{(t)}) \tag{20}$$

$$\tilde{y}^{(t)} = \sigma\left(Vh^{(t)} + c\right) \tag{21}$$

In the LSTM formulation, the input vector x_t corresponds to the RDIC-filtered signal defined in Equation (15). σ denotes the sigmoid activation function, while b_f and w_f represent, respectively, the bias and weight associated with the forget gate. Similarly, w_i and w_c correspond to the weights, and b_i and b_c to the bias terms of the input gate and the cell state update mechanism. The variable h_t refers to the output vector of the hidden layer at time step t. Furthermore, $V$and c designate the weight matrix and bias vector connecting the hidden layer to the output layer, ensuring the proper transmission of learned representations through the network.

3.3.3. GRU model

The GRU is a reduced recurrent architecture developed by Cho et al. [43] that reduces the structural complexity of traditional LSTM networks. The GRU is different from LSTM in that it maintains no cell state and regulates its memory via two gates only, namely, the update gate and the reset gate. With fewer trainable parameters than LSTM and with an increased speed of model convergence, the GRU has the capability to track long-term temporal dependencies in nonstationary financial data series. The internal computation of the GRU can be represented as follows (see Figure 4):

$$\begin{pmatrix} r \\ z \end{pmatrix} = \begin{pmatrix} sigm \\ sigm \end{pmatrix} w^l \begin{pmatrix} h_t^{l-1} \\ h_{t-1}^l \end{pmatrix} \tag{22}$$

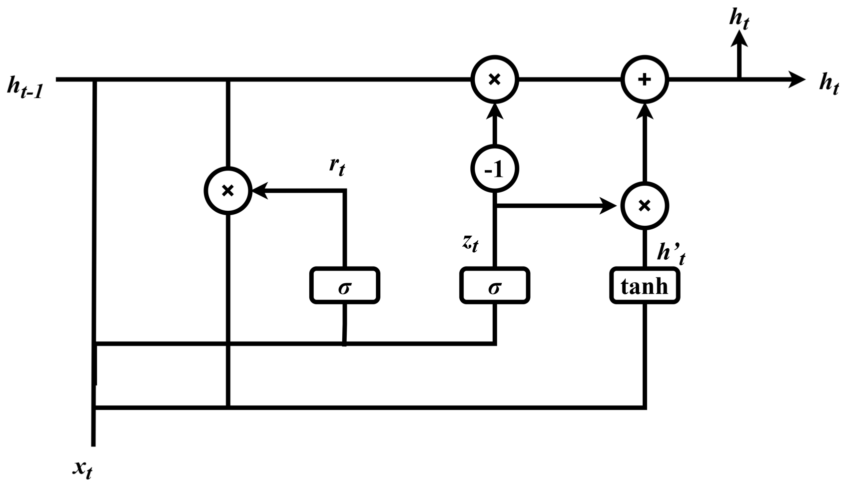
$$\tilde{h}_t^l = \tanh(w_x^l h_t^{l-1} + w_g^l(r_t \odot h_{t-1}^l)) \tag{23}$$

$$h_t^l = (1 - z_t) \odot h_{t-1}^l + z_t \odot \tilde{h}_t^l \tag{24}$$

In this formulation, x_t refers to the input vector at time t, while h_t denotes the current hidden state. The update gate z_t controls the proportion of previous information that should be preserved, whereas the reset gate r_t regulates the contribution of past memory to the current candidate update. The term $\tilde{h}_t$ represents the candidate activation, $\sigma(\cdot)$ corresponds to the sigmoid activation function, and $\odot$ indicates element-wise multiplication. This structure ensures an efficient equilibrium between expressive capacity and computational cost, which makes the GRU particularly well suited for forecasting highly volatile time series such as cryptocurrency prices [42].

The GRU model relies on two gates, the reset gate and the update gate, whereas the LSTM uses three gates (input, output, and forget). The reset gate determines how new information is combined with the past state, while the update gate controls the proportion of previous memory that must be retained. In this sense, the update mechanism fulfills a similar functional role to the combination of the input and forget gates in the LSTM model. Unlike the LSTM, the GRU does not maintain a separate cell state, which reduces the structural complexity and the number of trainable parameters, making it a competitive alternative when evaluating performance on sequential forecasting tasks.

Figure 4
Diagram of a single gated recurrent unit (GRU)



3.4. Workflow summary

The overall workflow of the proposed VMD-RDIC-DL (ARNN, LSTM, GRU) models is presented in Figure 5 and summarized in Algorithm 1. The process follows a bottom-up hybrid strategy that combines signal decomposition, information-theoretic feature optimization, and deep sequential learning.

As depicted in Figure 5, we developed an integrated hybrid model based on three stages: multiscale time-series analysis by means of VMD for the extraction of the relevant IMFs, feature extraction through the use of RDIC for the determination of the most informative components, and sequential prediction using DL. The entire process can be visualized in the flow chart.

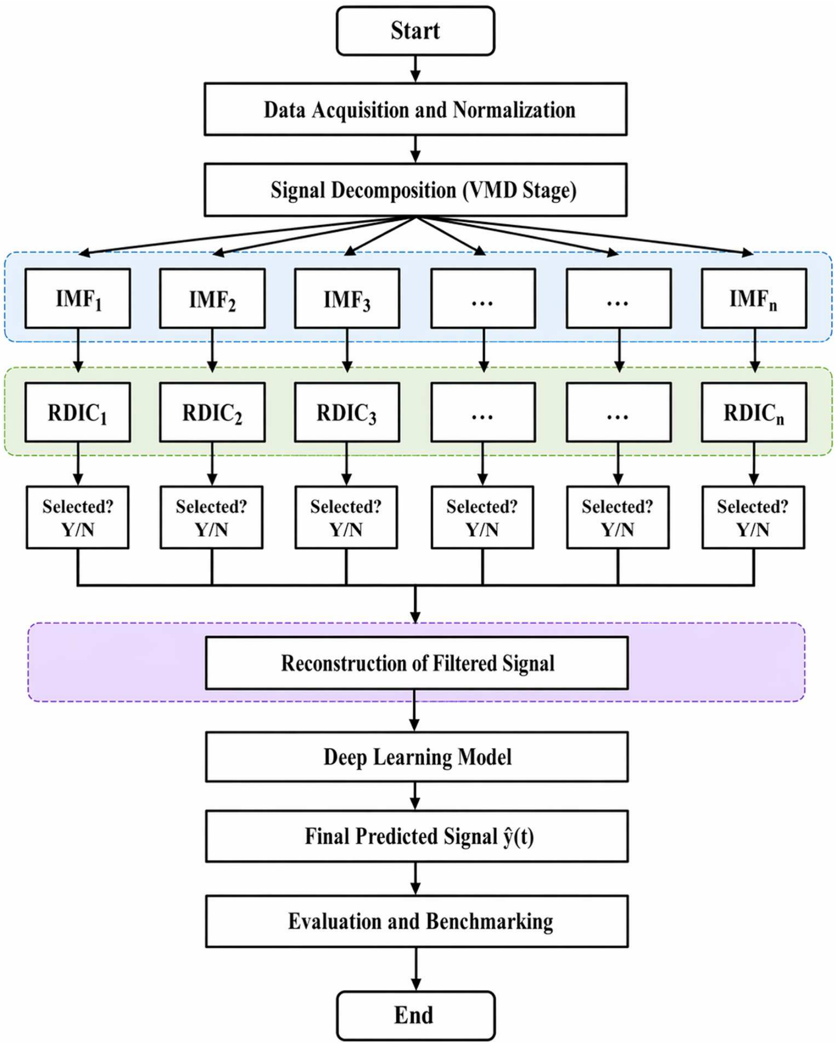
First, all available financial time series—BTC/USD—have to be preprocessed (data acquisition and normalization) so that they can be treated numerically with the same units, that is, scaled properly for the training of different machine learning models.

In the subsequent VMD stage, the normalized series is decomposed into a limited number of IMFs. Each IMF captures oscillatory behavior within a specific frequency band. Consequently, the VMD stage identifies high-frequency volatility in the financial time series from lower-frequency (mid-term and long-term) variability of the time series, allowing for the development of a clearer view of the time-series dynamics.

The RDIC Evaluation Phase is followed by determining which IMF(s) are most relevant to predict future values of the time series using an integrated measure of MI, Spearman rank correlation, and normalized energy. In doing so, the methodology allows for the isolation of the IMF(s) with the greatest amount of information regarding the future value(s) of the time series while removing redundancy and/or "noise" from the IMFs, thus minimizing the potential detrimental influence upon the DL model.

Rather than developing an independent DL model for each IMF that has been identified, the selected IMF modes can be used to construct a filtered version of the input signal based on the IMF modes that have been selected. In turn, this filtered input sequence will be processed by a single DL model (LSTM, GRU, ARNN) that is designed to capture short- and long-term patterns in the filtered data. The output generated by the DL model is the final predicted values of the $\hat{y}(t)$. Once these final predicted values of $\hat{y}(t)$ have been produced, they are quantitatively assessed with respect to their accuracy relative to the actual values of $\hat{y}(t)$ using standard performance metrics (MAE, RMSE, Relative Root Mean Square Error (RRMSE), R^2, and DA).

Figure 5
Workflow of the proposed VMD-RDIC-DL hybrid forecasting framework



The VMD-RDIC-DL workflow balances the interpretability provided by decomposing the input signal at multiple scales with the DL models' ability to provide accurate predictions of complex time-series signals. The complete operational workflow of the VMD-RDIC-DL framework is outlined below in detail.

Algorithm 1: VMD-RDIC-DL

Input: Normalized Bitcoin price series $x(t)$
Output: Final predicted series $\hat{y}(t)$
Step 1: Multiscale Signal Decomposition (VMD)
-Apply VMD to $x(t)$.
-Decompose $x(t)$ into K IMFs: $IMF_1(\mathrm{t}), IMF_2(\mathrm{t}), \ldots, IMF_K(\mathrm{t})$,
-Each IMF represents a specific oscillatory component of the price signal.
Step 2: Relevance-Driven IMF Combination (RDIC)
For each IMF_k, compute the following relevance indicators with respect to the original signal $x(t)$:

- Mutual information $\mathrm{MI_k}$, measuring nonlinear dependency
- Spearman's rank correlation coefficient $\mathrm{SP_k}$
- Relative energy contribution $\mathrm{EN_k}$

-Normalize all relevance scores to the interval $[0, 1]$.
-Compute the global RDIC relevance score:

$$RDICScore_k = w_{MI}\, MI_k^* + w_{Sp}|\rho_k^*| + w_{En}\, RE_k^*$$

Where: $w_{MI} = 0.5$, $w_{SP} = 0.3$, $w_{EN} = 0.2$
Step 3: IMF Ranking and Selection
-Rank IMFs in descending order according to $RDICScore_k$.
- Select the **top** M IMFs with the highest scores.
Step 4: Signal Reconstruction:

$$x_{\text{filtered}}(t) = \sum_{k \in S} u_k(t)$$

-where S denotes the selected IMF set.
Step 5: Deep Predictive Modeling
-Train a deep learning predictor DL (ARNN, LSTM, or GRU) using $x_{filtered}(t)$
-Generate the final predicted series:

$$\hat{y}(t) = \mathrm{DL}(x_{\text{filtered}}(t))$$

Step 6: Evaluation and Benchmarking
-Compare the predicted series $\hat{y}(t)$ with the real Bitcoin prices.
- Assess performance.

4. Experimental Setup

4.1. Materials and methods

All experiments were conducted on a workstation running Windows 10 (64-bit), equipped with an Intel Core i7 CPU (2.8 GHz), 16 GB RAM, and TensorFlow 2.19 using the Keras backend with Python 3.10.

4.2. Dataset description

In order to assess the feasibility of the proposed hybrid framework, we employed the 1-h interval sample of the historical Bitcoin price dataset from 01/2018 to 04/2025, whose original price signal is shown in Figure 6. This dataset was acquired from a publicly accessible source on Kaggle and is comprised of five standard attributes: Open, High, Low, Close, and Volume.

Data selection: We identified the Close price as our main target variable to be forecasted using the other four attributes as input variables. The correlation matrix depicted in Figure 7 reveals an almost perfect statistical dependence among Bitcoin price-related variables, namely, the opening, high, low, and closing prices, with correlation coefficients approaching unity. This strong coherence reflects a stable intraday structure in which different price components evolve synchronously and maintain relatively narrow spreads within each trading period. In contrast, trading volume exhibits a weak negative correlation with all price variables, indicating that periods of intensified market activity are not systematically associated with higher price levels. This partial decoupling between volume and price dynamics highlights the presence of more intricate market mechanisms, potentially driven by liquidity conditions, investor behavior, or external events, and supports the need for nonlinear modeling frameworks to adequately capture such subtle interactions.

The scatter plots in Figure 8 indicate a near-perfect linear relationship between the opening, high, and low prices and the Bitcoin closing price, confirming strong intraday coherence among price components and significant informational redundancy within the price vector. In contrast, trading volume exhibits a weak and highly dispersed association with the closing price, suggesting that market activity does not directly determine price levels and that their interaction is governed by complex, nonlinear dynamics influenced by multiple external factors.

Data segmentation: The dataset was divided chronologically into three parts: 80% for training, 10% for validation, and 10% for testing. This temporal split preserves the order of the observations and allows the model to be evaluated on future data that were not used during training.

4.3. Selection of the number of retained IMFs

An empirical study was carried out to find how many of the retained IMFs are best represented by evaluating the forecasting performance and the cumulative explained variance obtained using different values of k (from 1 through 8). In Figure 9(a), it is illustrated how the forecasting error (RMSE) varies as the number of IMFs being selected varies. The results show clearly that there is a decrease in the forecasted error as k increases from 1 through 3, and then the error increases as k continues to grow, showing that adding more than three IMFs will add little to no useful information but will add more noise to the forecasts. The cumulative explained variance of the reconstructed data is shown in Figure 9(b). It is seen that three IMFs already explain over 99.95% of the variance of the original data. Therefore, this shows that a compact form can capture all of the necessary information of the original time-series signal. These findings illustrate that selecting three IMFs provides a suitable compromise for retaining the most important dynamic features of the time series while excluding the more rapidly changing high-frequency elements of the signal.

4.4. Model configuration

The LSTM, ARNN, GRU, and VMD-RDIC DL models were trained for 60 epochs with a batch size of 256. A batch size of 256 was selected; a sliding window of length 256 was selected;

Figure 6
Original Bitcoin price signal

Figure 7
Correlation matrix

Figure 8
Scatter plots

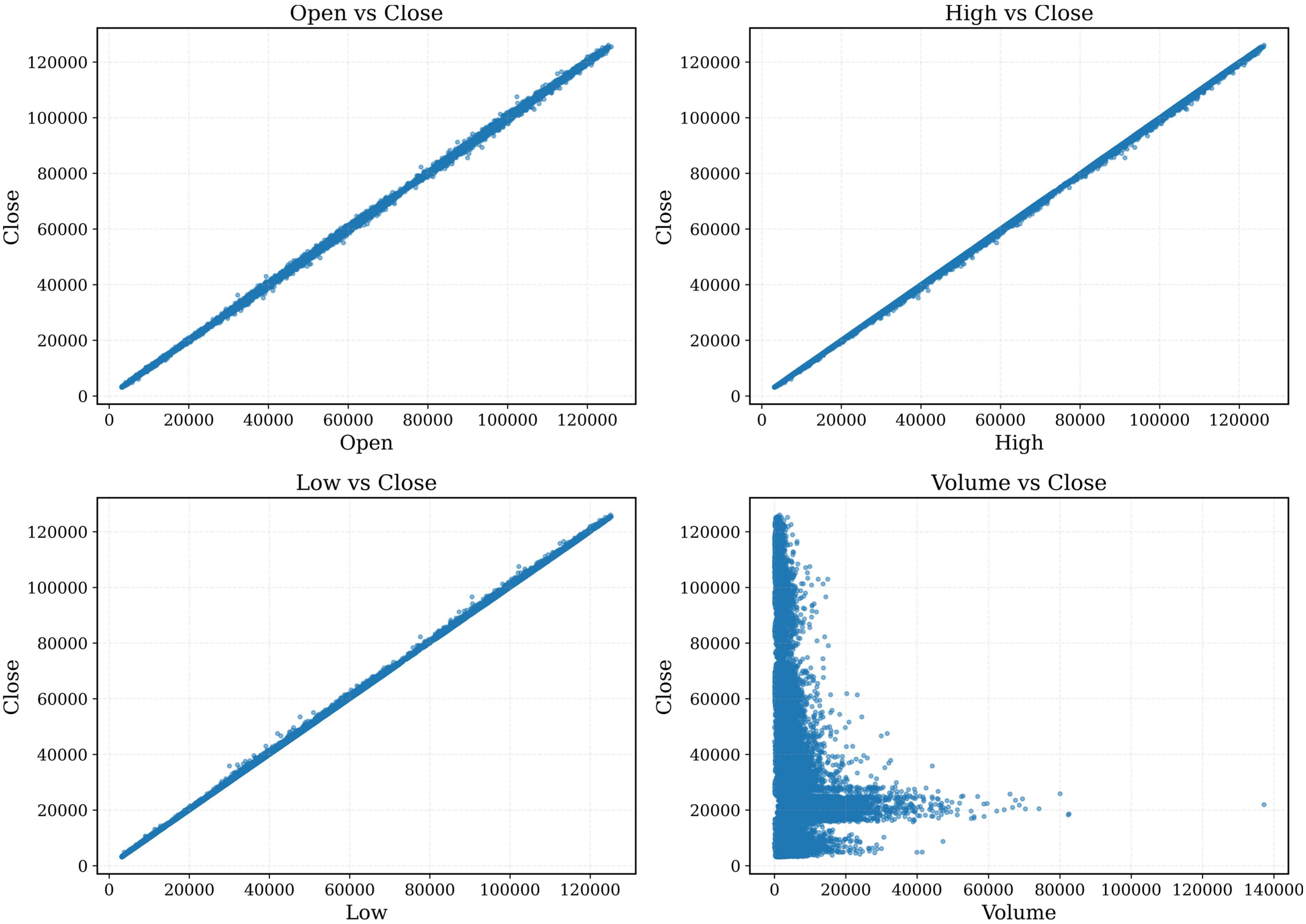
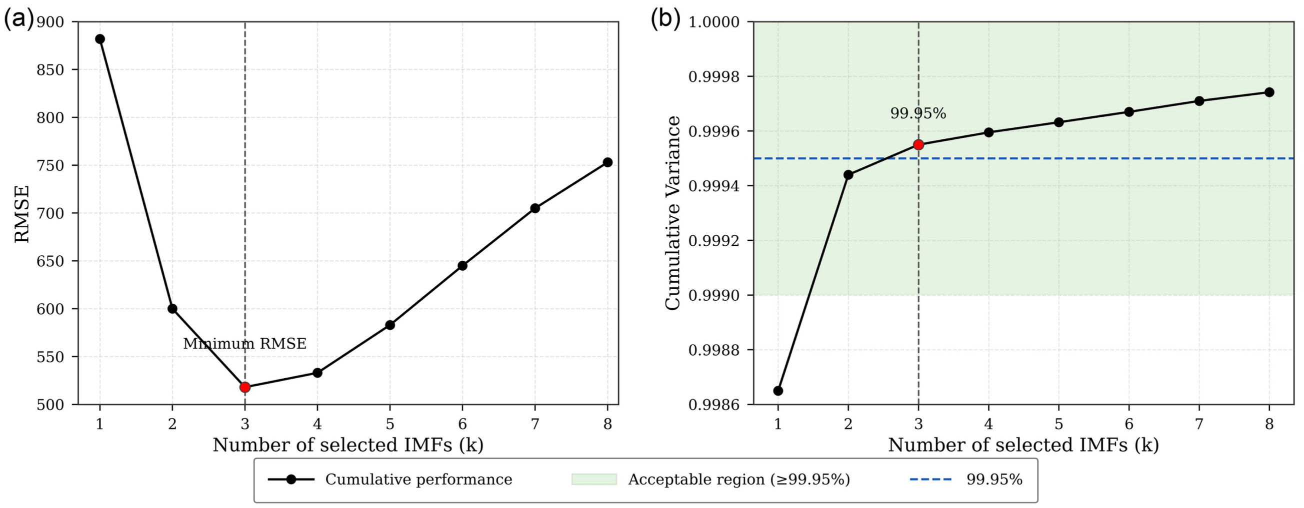


Figure 9
Experimental justification for selecting the number of retained IMFs: (a) forecasting error (RMSE) obtained with different numbers of IMFs and (b) cumulative explained variance of the reconstructed signal



an initial learning rate of 0.001 was selected (the Adam optimizer); and the prediction horizon was 24-time steps. Table 2 summarizes the main hyperparameters used in the experiments.

Table 2
Model parameter configuration

Parameter	Value
Window size (sequence length)	256
Forecasting horizon (steps)	24
Number of hidden layers	2
Units per layer	128
Activation function	ReLU
Optimizer	Adam
Learning rate	0.001
Batch size	256
Epochs	60
Dropout rate	0.2

4.5. Evaluation metrics

To evaluate model accuracy and robustness, five complementary metrics were employed:

$$\text{MAE} = \frac{1}{N}\sum |\, y_t - \hat{y}_t \,| \tag{25}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum (y_t - \hat{y}_t)^2} \tag{26}$$

$$\text{RRMSE} = \frac{\sqrt{\frac{1}{N}\sum (y_t - \hat{y}_t)^2}}{\bar{y}} \tag{27}$$

$$R^2 = 1 - \frac{\sum (y_t - \hat{y}_t)^2}{\sum (y_t - \bar{y})^2} \tag{28}$$

$$\text{DA} = \frac{1}{N}\sum_{t=1}^{N} 1(sign(y_t - y_{t-1}) = sign(\hat{y}_t - \hat{y}_{t-1})) \tag{29}$$

where y_t denotes the actual value at time t, $\hat{y}_t$ is the predicted value, N is the total number of observations, and y is the mean of the actual values $y = \frac{1}{N}\sum y_t$. These indicators jointly capture absolute error magnitude, scale-adjusted accuracy, and predictive consistency.

4.6. Statistical significance analysis

1) Binomial test for directional accuracy

To verify whether the DA is statistically significant, a binomial test is applied. This test evaluates whether the proportion of correctly predicted directions is greater than what could be expected by random guessing (50%). If $\boldsymbol{k}$ denotes the number of correct predictions and $\boldsymbol{n}$ the total number of predictions, the probability is computed as:

$$\boldsymbol{p - value} = \sum_{i=k}^{n} \binom{\boldsymbol{n}}{\boldsymbol{i}} (\mathbf{0.5})^{\boldsymbol{i}} (\mathbf{0.5})^{\boldsymbol{i-1}} \tag{30}$$

A p-value lower than 0.05 indicates that the model predicts the direction of the time series significantly better than chance.

2) Diebold–Mariano test for predictive accuracy

The Diebold–Mariano (DM) test is used to compare the predictive accuracy of two forecasting models. Let $\boldsymbol{e}_{1,t}$ and $\boldsymbol{e}_{2,t}$ denote the forecast errors produced by the two models at time t. The loss differential is defined as:

$$\text{dt} = \text{L}(\boldsymbol{e}_{1,t}) - \text{L}(\boldsymbol{e}_{2,t}) \tag{31}$$

The DM statistics are then calculated as:

$$DM = \frac{\bar{d}}{\sqrt{\frac{\hat{\sigma}_d^2}{T}}} \tag{32}$$

where $\bar{d}$ is the mean loss differential, $\hat{\sigma}_d^2$ its variance, and T the number of observations. A significant result indicates that the difference in forecasting performance between the models is statistically meaningful.

5. Results and Discussion

This section presents and interprets the empirical findings obtained from the forecasting experiments conducted on the hourly BTC/USD dataset covering the period from 2018 to 2025. The proposed VMD-RDIC-DL models were evaluated in comparison with several benchmark approaches, including machine learning algorithms (Support Vector Regression (SVR), Random Forest, and XGBOOST).

First, the time series is decomposed into a set of eight IMFs using VMD, as shown in Figure 10, which allows the signal to be separated into distinct frequency components. This multi-scale representation helps isolate long-term trends, medium-range fluctuations, and short-term volatility.

Table 3 presents the RDIC-based evaluation of the IMFs obtained from the VMD decomposition. The results show that IMF_1 exhibits the highest relevance (MI = 0.964, Spearman = 0.938, Energy = 0.912), indicating a strong dependence on the original signal and capturing the dominant long-term trend of the Bitcoin market. IMF_2 also demonstrates a meaningful level of relevance (MI = 0.712, Spearman = 0.421, Energy = 0.146), corresponding to medium-frequency oscillations related to cyclical market behavior. IMF_6 preserves informative high-frequency dynamics (MI = 0.436, Spearman = 0.208, Energy = 0.082), which reflect short-term price fluctuations. In contrast, the remaining components (IMF_3, IMF_4, IMF_5, IMF_7, and IMF_8) display substantially lower values across the three relevance indicators and mainly correspond to noise-dominated variations. Based on the combined evaluation of MI, Spearman rank correlation, and relative energy, IMF_1, IMF_2, and IMF_6 were selected to reconstruct the filtered signal used as input for the DL forecasting models.

As illustrated in Figure 11, IMF_1 is related to the long-term trend in the market, IMF_2 is related to the middle-frequency cycle, and IMF_6 is related to the high-frequency details and short-term trend of the prices. All other IMFs were removed because they contained little information or mostly represented noise.

After applying the RDIC-based filtering to each decomposed time series, the predictive performance of the proposed VMD-RDIC-ARNN, VMD-RDIC-LSTM, and VMD-RDIC-GRU models is systematically evaluated.

Figures 12, 13, and 14 present a comparative visual assessment of the forecasting performance of baseline recurrent models

Figure 10
Clustering results of IMF with VMD and $K = 8$

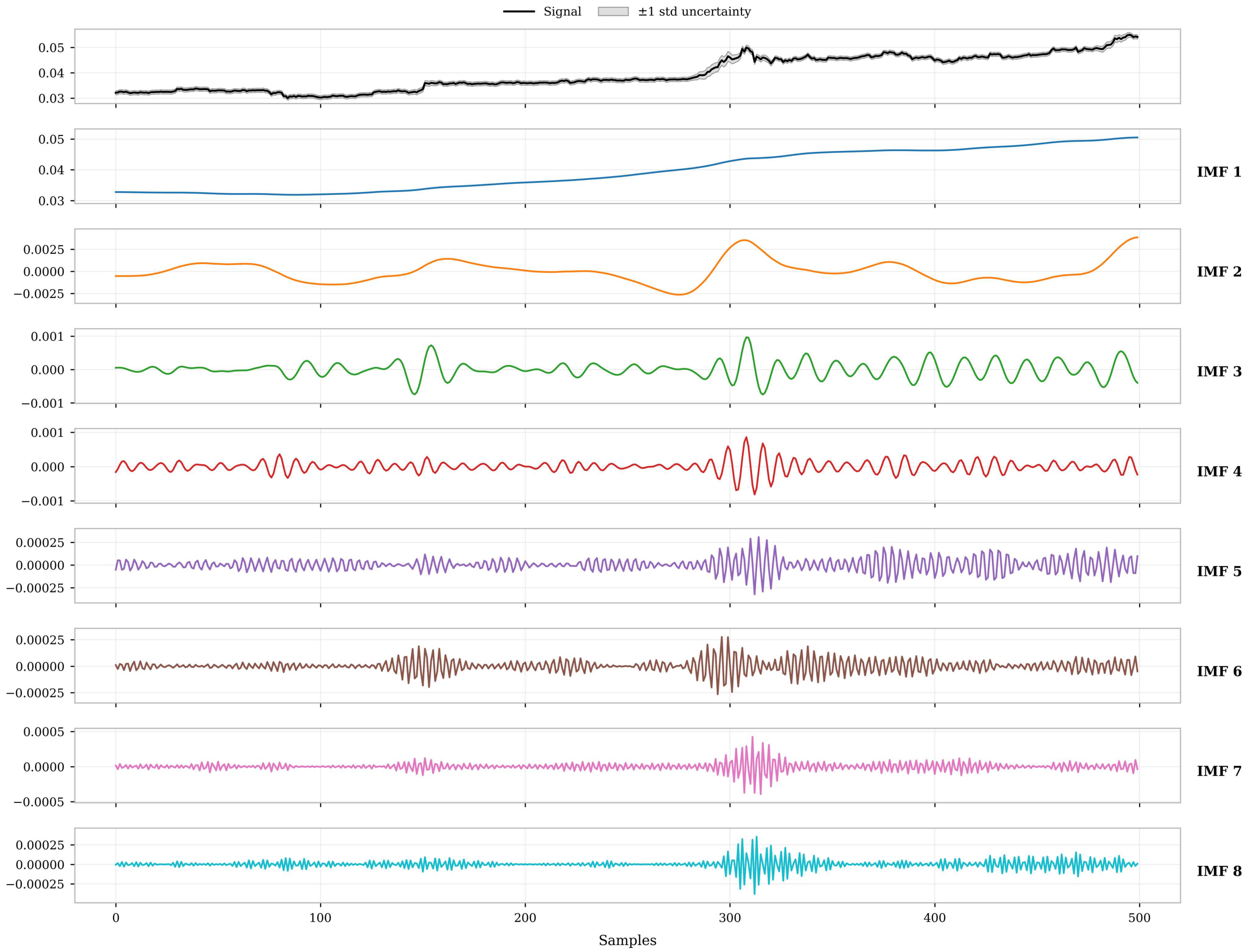
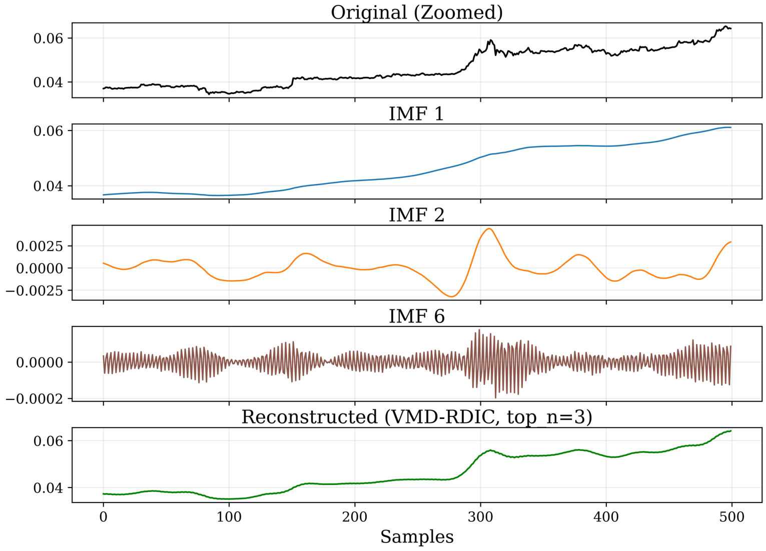


Table 3
Evaluation of RDIC scores for VMD-decomposed IMFs

IMF	MI	Spearman	Energy	Selected	Interpretation
IMF_1	0.964	0.938	0.912	✔	Dominant low-frequency component capturing the main trend
IMF_2	0.712	0.421	0.146	✔	Medium-frequency oscillatory component reflecting market cycles
IMF_3	0.214	0.118	0.073	–	Weak oscillatory component with limited predictive relevance
IMF_4	0.158	0.093	0.052	–	High-frequency component dominated by noise
IMF_5	0.134	0.071	0.041	–	Low-energy IMF with negligible predictive contribution
IMF_6	0.436	0.208	0.082	✔	Informative high-frequency component capturing short-term volatility
IMF_7	0.097	0.056	0.029	–	Noise-like IMF with weak statistical association
IMF_8	0.062	0.044	0.018	–	Residual component dominated by noise

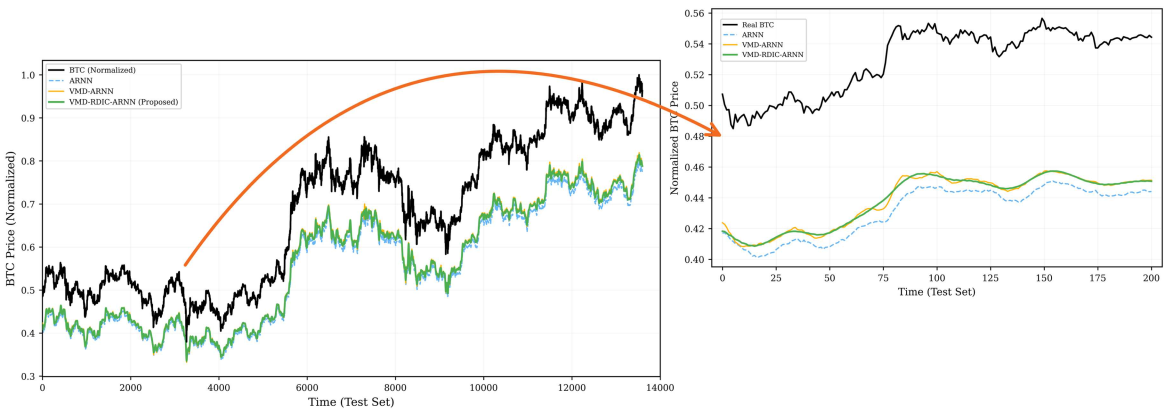
(ARNN, LSTM, and GRU) and their enhanced variants incorporating VMD and the proposed relevance-driven filtering mechanism (VMD-RDIC). For each architecture, the global prediction trajectories over the test horizon are shown together with a zoomed-in comparison of BTC forecasting results over the interval [zoomStart–zoomEnd] = [839−919].

Figure 11
Visualization of the original Bitcoin price signal and its reconstruction from the three selected IMFs (IMF_1, IMF_2, IMF_6), identified through RDIC evaluation within the VMD framework



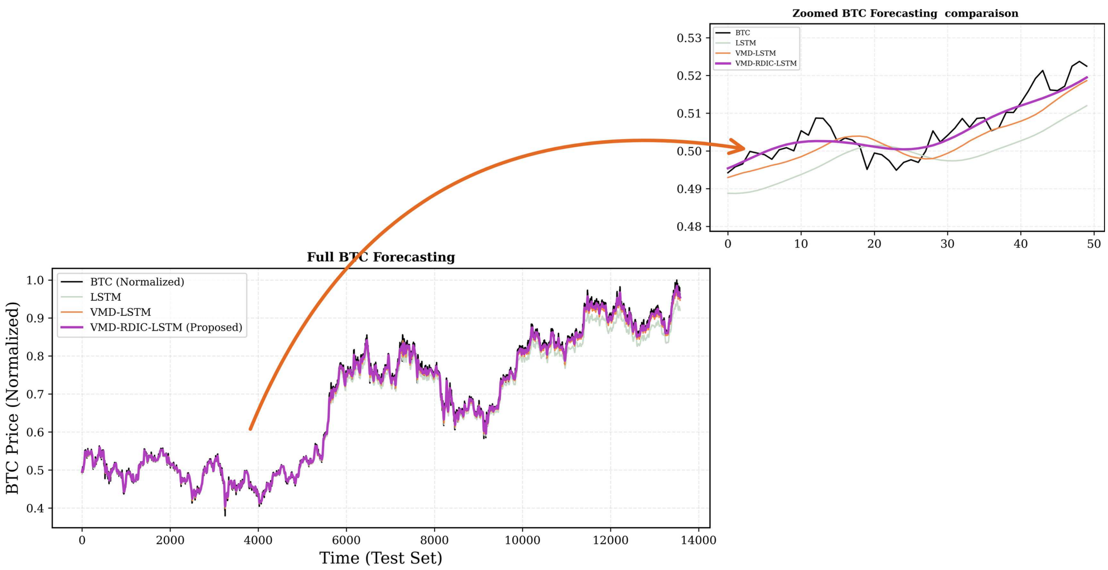
1) ARNN, VMD-ARNN, and VMD-RDIC-ARNN

Figure 12
Comparison between ARNN, VMD-ARNN, and VMD-RDIC-ARNN



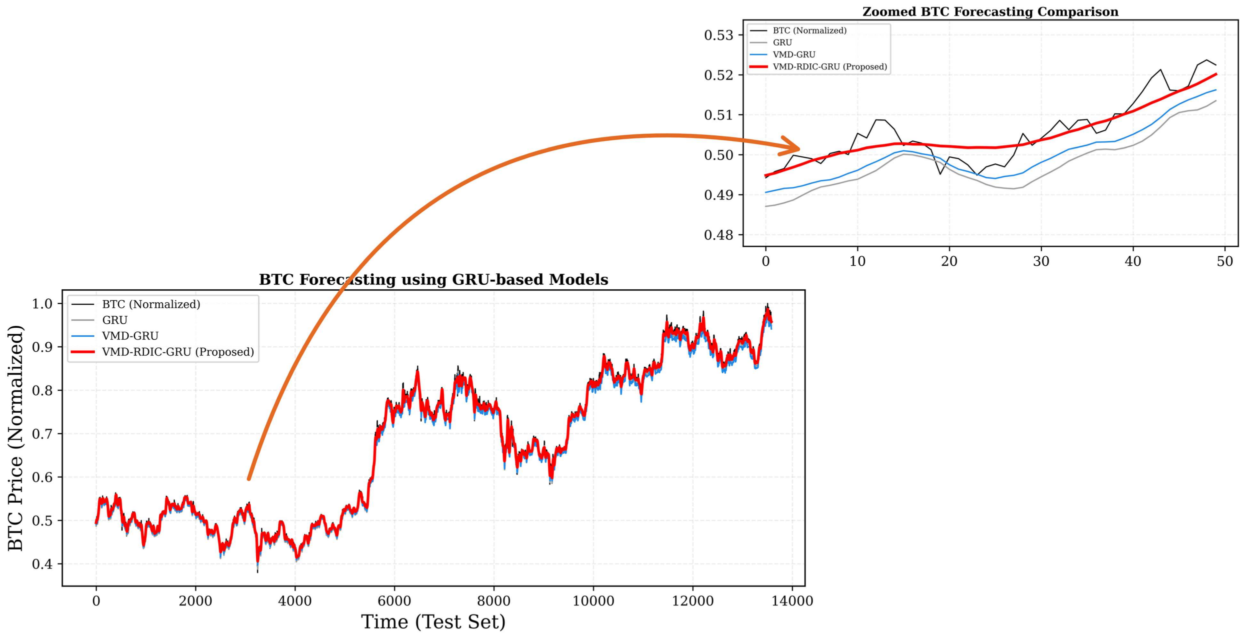
2) LSTM, VMD-LSTM, and VMD-RDIC-LSTM

Figure 13
Comparison of LSTM, VMD-LSTM, and VMD-RDIC-LSTM models



3) GRU, VMD-GRU, and VMD-RDIC-GRU

Figure 14
Overall comparison of GRU, VMD-GRU, and VMD-RDIC-GRU forecasting results



The GRU structure also enjoys this upgrade to a similar degree, as illustrated in Figure 15. The GRU's ability to predict and actualize values is virtually in sync, as the values are rarely out of phase or at least far out of phase when the values are at turning points and at times of rapid market reversals. These overall results highlight that the hybridization of VMD and RDIC is both robust and generally applicable to all types of RNNs, as the hybridization has consistently shown an improvement in stability for models, reduction of sensitivity to noise, and increased accuracy for predictive purposes.

Figure 16 illustrates the absolute prediction error over the test period for the VMD-RDIC-ARNN, VMD-RDIC-LSTM, and VMD-RDIC-GRU models. It highlights the ability of each architecture to track Bitcoin price movements and to limit forecasting deviations, particularly during volatile market phases. This figure confirms that combining VMD with RDIC substantially improves forecasting stability and that GRU provides superior temporal modeling capabilities compared to LSTM and ARNN in the context of highly volatile cryptocurrency markets.

Among the evaluated models, VMD-RDIC-GRU emerges as the most reliable and accurate configuration, achieving the smallest and most stable prediction errors over time.

Figure 17 presents a localized (zoomed-in) comparison of Bitcoin forecasting performance obtained using ARNN, LSTM, and GRU models, along with their variants incorporating VMD and the proposed RDIC.

ARNN-based models (a–c). The standard ARNN model (a) exhibits a pronounced underestimation of price levels and a limited ability to reproduce local variations, reflecting excessive smoothing and weak sensitivity to multiscale dynamics. Incorporating VMD (b) slightly improves the predicted trajectory through noise reduction and frequency separation; however, noticeable discrepancies remain during transition phases. The proposed VMD-RDIC-ARNN model (c) achieves the closest agreement among ARNN variants, producing a trajectory more consistent with the true signal and better capturing local trends, thereby confirming the benefit of selective IMF filtering.

Figure 15
Individual comparison of GRU, VMD-GRU, and VMD-RDIC-GRU predictions against the original BTC series: (a) GRU, (b) VMD-GRU, and (c) VMD-RDIC-GRU

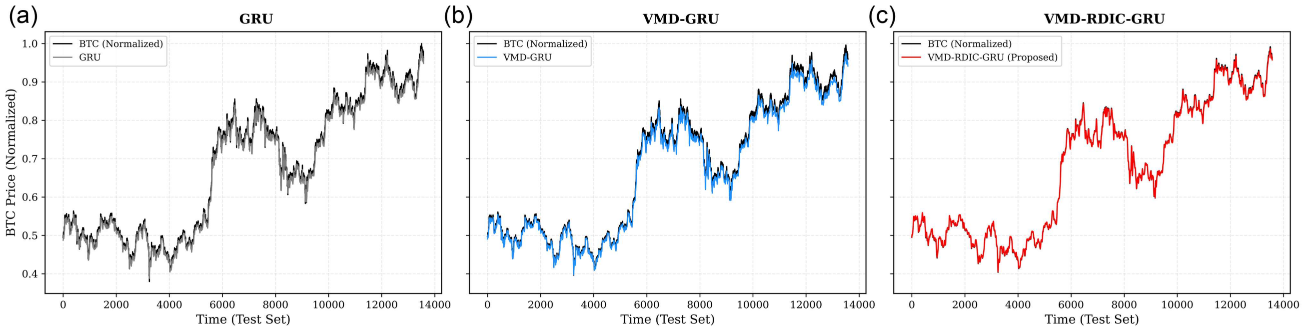


Figure 16
Temporal evolution of absolute forecasting error for VMD-RDIC hybrid models

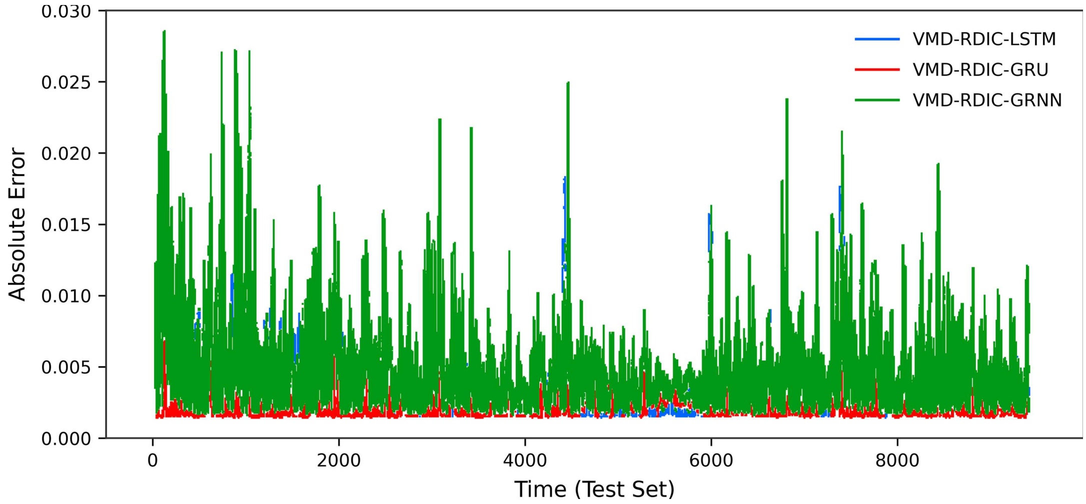
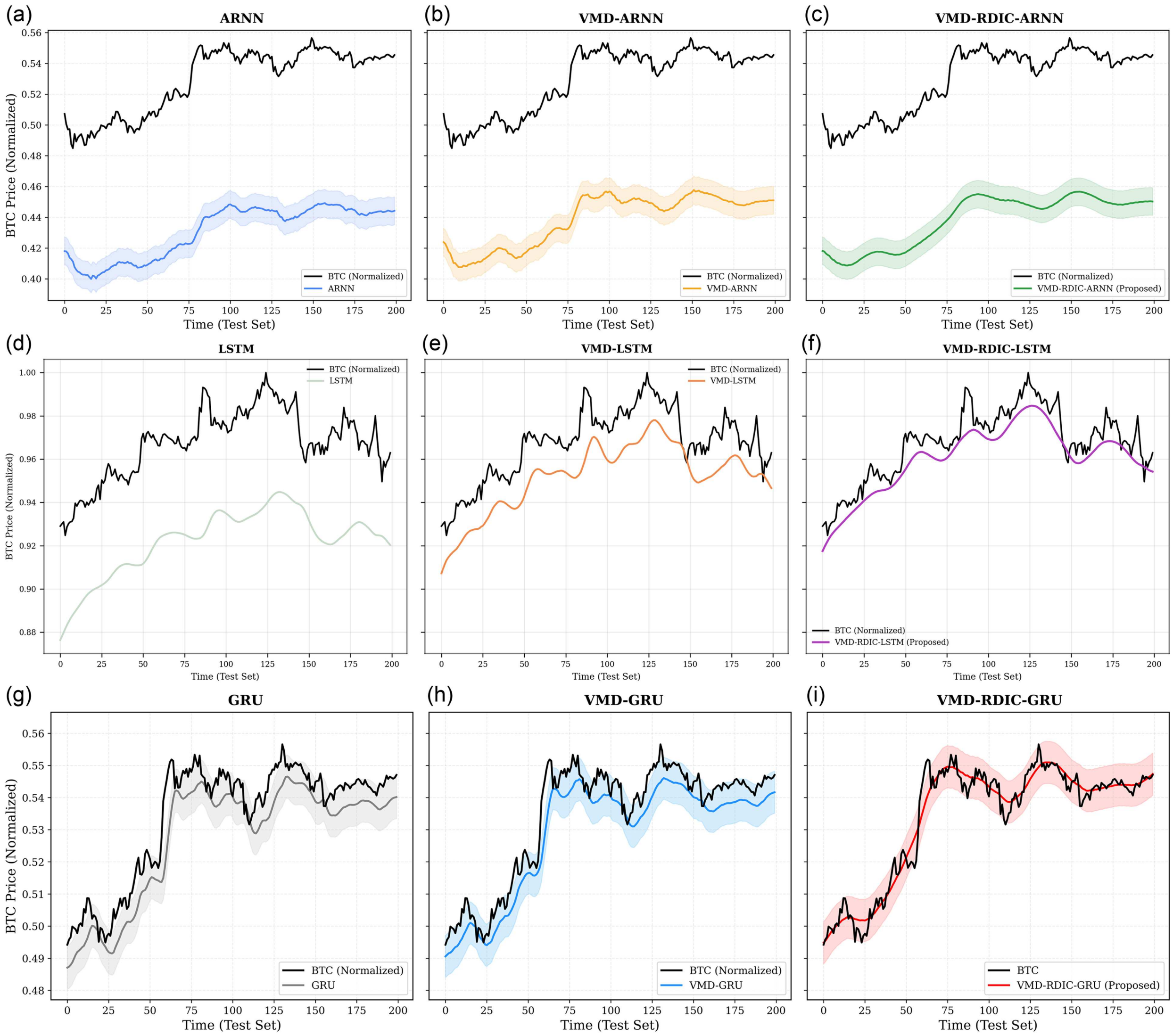


Figure 17
Forecasting behavior of ARNN, LSTM, and GRU models integrated into the VMD-RDIC framework (zoomed-in analysis of BTC price forecasting results)



LSTM-based models (d–f). The baseline LSTM (d) captures the overall trend adequately but exhibits temporal lag and amplitude attenuation during rapid fluctuations. The VMD-LSTM version (e) reduces these deviations by improving the structure of the input signal and yet remains limited in accurately tracking turning points. In contrast, the VMD-RDIC-LSTM model (f) closely follows the true trajectory in both amplitude and phase, indicating an enhanced capability to jointly represent long-term components and short-term variations.

GRU-based models (g–i). The standard GRU model (g) demonstrates greater stability than ARNN but remains affected by excessive smoothing. The introduction of VMD (h) increases responsiveness and reduces local error. The proposed VMD-RDIC-GRU model (i) delivers the best overall performance, achieving near-perfect overlap with the ground-truth series over most of the zoomed interval, which highlights its strong ability to track rapid fluctuations and directional changes.

The performance of three traditional machine learning algorithms is compared to the proposed VMD-RDIC-DL model by comparing them to each other and to each other for the same test interval as shown in Figure 18. The performance indicators RMSE, MAE, RRMSE, R^2, and DA are summarized in Table 4, which provide a comparison of the performance of the classical approaches (XGBoost, SVR, Random Forest) to the proposed approach. The classical approaches have significantly less accurate predictions and therefore show that they are unable to capture the complexity of the nonlinear relationships as well as the rapid changes in regimes contained within the Bitcoin price time series. In addition, these models tend to oversimplify the temporal dependencies, which leads to unstable forecast results due to the high degree of dynamicity present in various market phases.

The global comparison of the evaluated models is presented in Table 4. A normalized scale is used to facilitate the comparison between the traditional statistical model (ARIMA), classical machine learning approaches, DL models, and the proposed hybrid architectures.

The results presented in Table 4 show that the hybrid VMD-RDIC architecture consistently outperforms the baseline machine learning and DL models across most evaluation metrics. In particular, the proposed VMD-RDIC-GRU model achieves the best overall performance, obtaining the lowest RMSE (0.0049) and MAE (0.0038) values, while also providing the smallest RRMSE

Figure 18
Comparative normalized BTC/USD forecasts of ML baselines and hybrid VMD-RDIC-deep learning models

Table 4
Global comparison of machine learning and deep hybrid models

Model	RMSE	RRMSE	MAE	R^2
ARIMA	0.1450	0.2235	0.1240	0.5693
SVR	0.0606	0.0932	0.0442	0.8695
Random Forest	0.1936	0.2978	0.1388	0.3314
XGBoost	0.2053	0.3158	0.1499	0.5281
ARNN	0.1419	0.2098	0.1370	0.8531
VMD-ARNN	0.1324	0.1957	0.1280	0.8675
VMD-RDIC-ARNN	0.1308	0.1933	0.1266	0.8987
LSTM	0.0149	0.0220	0.0120	0.9412
VMD-LSTM	0.0093	0.0137	0.0071	0.9568
VMD-RDIC-LSTM	0.0075	0.0111	0.0059	0.9684
GRU	0.0123	0.0188	0.0103	0.9486
VMD-GRU	0.0098	0.0115	0.0081	0.9627
VMD-RDIC-GRU	0.0049	0.0068	0.0038	0.9789

(0.0068) and the highest coefficient of determination value (R^2 = 0.9789).

These results demonstrate that combining VMD with the RDIC-based IMF selection mechanism significantly improves forecasting stability and noise filtering in highly volatile cryptocurrency markets. Overall, the comparative analysis confirms that hybrid architectures integrating VMD-RDIC with recurrent models such as ARNN, LSTM, and GRU provide more reliable short-term BTC/USD forecasting, with VMD-RDIC-GRU emerging as the most effective configuration among the evaluated approaches.

1) Binomial test for directional accuracy (> 50%)

To evaluate whether the DA is statistically better than random guessing (50%), a binomial test was conducted. The results shown in Table 5, the baseline models (ARNN, LSTM, and GRU) do not achieve statistically significant directional prediction. In contrast, the proposed VMD-RDIC models significantly improve directional forecasting performance, with DA values of 55.19%, 55.09%, and 60.73% for ARNN, LSTM, and GRU, respectively ($p < 0.001$).

2) Diebold–Mariano test

The results in Table 6 demonstrate that the VMD-RDIC models have statistically significant improvements in terms of the quality of the forecasts compared with their respective baselines (VMD-RDIC-ARNN, VMD-RDIC-LSTM, and VMD-RDIC-GRU). All of the calculated p-values were less than .001, demonstrating statistical significance of the differences in forecasting quality for each comparison. This confirms that the improvement from VMD-RDIC-ARNN, VMD-RDIC-LSTM, and VMD-RDIC-GRU is real and not merely an artifact of random chance. Additionally, the consistent results between squared error loss and absolute error loss provide evidence that the method is also robust regarding different performance metrics.

Table 5
Binomial test for directional accuracy

Model	DA (%)	Binomial	Significant
ARNN	47.15	1.000	No
VMD-RDIC-ARNN	55.19	<0.001	Yes
LSTM	48.38	0.9992	No
VMD-RDIC-LSTM	55.09	<0.001	Yes
GRU	47.73	1.000	No
VMD-RDIC-GRU	60.73	<0.001	Yes

6. Conclusion

As an alternative to the difficulties presented in the context of the Bitcoin price forecasting process with respect to the presence of strong nonstationarity, the presence of noise in the data, and complex nonlinearities (dynamics) among the different possible ways of handling these issues, we propose a novel framework for the integration of VMD, relevance-driven component selection, and deep sequential learning models. The key motivation of our proposal is to generate an organized and information-rich representation of the price series, which can be learned more easily than the raw price series; that is, we want to create a structure within the data that allows us to learn the temporal relationships.

First, the price signal of Bitcoin is decomposed into IMFs using VMD, allowing the possibility to separate clearly and unequivocally the long-term trend of the price, the oscillations with intermediate frequency, and the short-term fluctuations. Compared to empirical methods of decomposition, the use of VMD provides a better separation of the mode mixing, and therefore, it produces sub-components that are more stable and easier to interpret. Then, each of the decomposed signals is selected according to relevance-driven criteria based on three features: nonlinearity, monotonicity, and energy contribution. The goal of this step is to retain only the most relevant modes and eliminate the redundancy and noise components that have been previously retained. Therefore, the signal has a greater signal-to-noise ratio and a more coherent structural representation.

Then, the filtered time series resulting from the previous steps is provided as the input for the RNNs, such as ARNNs, LSTMs, and GRUs. The purpose of this integration is to allow the RNNs to capture the complexity of the temporal patterns contained in the time series, while at the same time to reduce the training instability problems often associated with noisy and poorly structured raw financial data. An extensive set of experiments was performed on hourly time series of BTC/USD prices from 2018 to 2025 to demonstrate the validity of our proposal. We demonstrated that the models proposed in this research, specifically those that are based on VMD-RDIC, provide better results compared to traditional machine learning algorithms (such as SVR, Random Forest, and XGBoost), and also compared to the results obtained from DL models (LSTM, GRU, etc.) that are trained directly on the original price series. Specifically, the best configuration was the one that used VMD-RDIC and GRU, since it demonstrated the lowest errors in predictions, the highest degree of goodness of fit, and the highest degree of DA, especially in periods of high volatility in the market.

Table 6
Diebold–Mariano test

Comparison	Loss function	DM statistic	p-value	Better model
VMD-RDIC-ARNN vs ARNN	Squared error loss	7.82	<0.001	VMD-RDIC-ARNN
	Absolute error loss	8.64	<0.001	VMD-RDIC-ARNN
VMD-RDIC-LSTM vs LSTM	Squared error loss	6.15	<0.001	VMD-RDIC-LSTM
	Absolute error loss	6.98	<0.001	VMD-RDIC-LSTM
VMD-RDIC-GRU vs GRU	Squared error loss	5.74	<0.001	VMD-RDIC-GRU
	Absolute error loss	6.21	<0.001	VMD-RDIC-GRU

Overall, the proposed framework presents a good and viable solution for the problem of short-term forecasting of the price of Bitcoin in dynamic and uncertain environments. The relevance-guided multiscale processing step is fundamental to improve the robustness of the method, to decrease the sensitivity to noise, and to increase the learning efficiency of the RNNs.

From a practical viewpoint, forecasting models can be used in real trading environments and must consider additional operational factors that are not explicitly modeled by this study. These additional operational factors include transaction costs, market slippage, and liquidity constraints that may impact the profitability of strategies based on short-term forecasting. Thus, although the proposed framework has demonstrated very good forecasting performance, its use in practice should be evaluated as part of a complete trading system that includes these market frictions.

Furthermore, the proposed framework is not restricted to Bitcoin data and can be extended with ease to other cryptocurrencies or financial asset classes such as stock indices, commodities, or foreign exchange markets.

Future research may extend this approach to other cryptocurrency markets to test the ability of our proposal to generalize and to transfer knowledge. Also, additional future extensions of our proposal include the incorporation of macroeconomic indicators and/or microstructure market variables, the development of online learning strategies to tackle regime shifts and concept drift, and finally, the evaluation of transformer-based architectures (such as Informer and/or FEDformer). A final area of potential extension includes modeling several correlated assets simultaneously in order to capture the interdependencies between the different markets and to improve the predictive performance of the models.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available at https://www.kaggle.com/datasets/novandraanugrah/bitcoin-historical-datasets-2018-2024.

Author Contribution Statement

Maryam Maatallah: Conceptualization, Methodology, Software, Data curation, Writing – original draft, Visualization. **Mourad Fariss:** Software, Formal analysis, Writing – original draft. **Hakima Asaidi:** Validation, Investigation, Writing – review & editing. **Mohamed Bellouki:** Resources, Writing – review & editing, Supervision, Project administration.

References

[1] Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., & Lachhab, A. (2018). Forecasting of demand using ARIMA model. *International Journal of Engineering Business Management*, *10*, 1847979018808673. https://doi.org/10.1177/1847979018808673

[2] Franses, P. H., & van Dijk, D. (1996). Forecasting stock market volatility using (non-linear) GARCH models. *Journal of Forecasting*, *15*(3), 229–235. https://doi.org/10.1002/(SICI)1099-131X(199604)15:3<229::AID-FOR620>3.0.CO;2-3

[3] Saxena, H., Aponte, O., & McConky, K. T. (2019). A hybrid machine learning model for forecasting a billing period's peak electric load days. *International Journal of Forecasting*, *35*(4), 1288–1303. https://doi.org/10.1016/j.ijforecast.2019.03.025

[4] Peng, P., Chen, Y., Lin, W., & Wang, J. Z. (2024). Attention-based CNN–LSTM for high-frequency multiple cryptocurrency trend prediction. *Expert Systems with Applications*, *237*, 121520. https://doi.org/10.1016/j.eswa.2023.121520

[5] Zhu, M., Qi, H., Ni, S., & Liu, Y. (2025). MSGformer: A hybrid multi-scale graph–transformer architecture for unified short- and long-term financial time series forecasting. *Electronics*, *14*(12), 2457. https://doi.org/10.3390/electronics14122457

[6] Fariss, M., Maatallah, M., Bay, B. B. A. Y., Asaidi, H., & Bellouki, M. (2024). Enhancing Forex trading predictions with machine learning: Cloud and local performance evaluation. In *2024 Sixth International Conference on Intelligent Computing in Data Sciences*, 1–8. https://doi.org/10.1109/ICDS62089.2024.10756412

[7] Arvanitidis, A. I., Bargiotas, D., Kontogiannis, D., Fevgas, A., & Alamaniotis, M. (2022). Optimized data-driven models for short-term electricity price forecasting based on signal decomposition and clustering techniques. *Energies*, *15*(21), 7929. https://doi.org/10.3390/en15217929

[8] Jin, C., & Li, Y. (2023). Cryptocurrency price prediction using frequency decomposition and deep learning. *Fractal and Fractional*, *7*(10), 708. https://doi.org/10.3390/fractalfract7100708

[9] Lv, L., Wu, Z., Zhang, J., Zhang, L., Tan, Z., & Tian, Z. (2022). A VMD and LSTM based hybrid model of load forecasting for power grid security. *IEEE Transactions on Industrial Informatics*, *18*(9), 6474–6482. https://doi.org/10.1109/TII.2021.3130237

[10] Zhang, S., Liu, G., Sun, S., & Cai, J. (2025). Deep learning method based on multivariate variational mode decomposition for classification of epileptic signals. *Brain Sciences*, *15*(9), 933. https://doi.org/10.3390/brainsci15090933

[11] Yu, H., Baek, S., Lee, J., Sohn, I., Hwang, B., & Park, C. (2024). Deep neural network-based empirical mode decomposition for motor imagery EEG classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *32*, 3647–3656. https://doi.org/10.1109/TNSRE.2024.3432102

[12] Lotfipoor, A., Patidar, S., & Jenkins, D. P. (2024). Deep neural network with empirical mode decomposition and Bayesian optimisation for residential load forecasting. *Expert Systems with Applications*, *237*, 121355. https://doi.org/10.1016/j.eswa.2023.121355

[13] Fang, M., Zhang, F., Zhu, D., Tao, R., & Xiao, R. (2025). The influence of variational mode decomposition on LSTM prediction accuracy—A case study with wind turbine power signals. *Renewable Energy*, *245*, 122863. https://doi.org/10.1016/j.renene.2025.122863

[14] Zhang, Z., Liu, Q., Hu, Y., & Liu, H. (2025). Multi-feature stock price prediction by LSTM networks based on VMD and TMFG. *Journal of Big Data*, *12*(1), 74. https://doi.org/10.1186/s40537-025-01127-4

[15] Fakhfekh, M., & Jeribi, A. (2020). Volatility dynamics of crypto-currencies' returns: Evidence from asymmetric and

long memory GARCH models. *Research in International Business and Finance*, *51*, 101075. https://doi.org/10.1016/j.ribaf.2019.101075
[16] Bouteska, A., Abedin, M. Z., Hajek, P., & Yuan, K. (2024). Cryptocurrency price forecasting—A comparative analysis of ensemble learning and deep learning methods. *International Review of Financial Analysis*, *92*, 103055. https://doi.org/10.1016/j.irfa.2023.103055
[17] Zennaro, F. M., & Chen, K. (2018). Towards understanding sparse filtering: A theoretical perspective. *Neural Networks*, *98*, 154–177. https://doi.org/10.1016/j.neunet.2017.11.010
[18] Lahmiri, S., & Bekiros, S. (2019). Cryptocurrency forecasting with deep learning chaotic neural networks. *Chaos, Solitons & Fractals*, *118*, 35–40. https://doi.org/10.1016/j.chaos.2018.11.014
[19] Kim, J.-M., Jun, C., & Lee, J. (2021). Forecasting the volatility of the cryptocurrency market by GARCH and stochastic volatility. *Mathematics*, *9*(14), 1614. https://doi.org/10.3390/math9141614
[20] Abumohsen, M., Owda, A. Y., & Owda, M. (2023). Electrical load forecasting using LSTM, GRU, and RNN algorithms. *Energies*, *16*(5), 2283. https://doi.org/10.3390/en16052283
[21] Niu, M., Hu, Y., Sun, S., & Liu, Y. (2018). A novel hybrid decomposition-ensemble model based on VMD and HGWO for container throughput forecasting. *Applied Mathematical Modelling*, *57*, 163–178. https://doi.org/10.1016/j.apm.2018.01.014
[22] Dragomiretskiy, K., & Zosso, D. (2014). Variational mode decomposition. *IEEE Transactions on Signal Processing*, *62*(3), 531–544. https://doi.org/10.1109/TSP.2013.2288675
[23] Zhang, J., Ye, L., & Lai, Y. (2023). Stock price prediction using CNN-BiLSTM-Attention model. *Mathematics*, *11*(9), 1985. https://doi.org/10.3390/math11091985
[24] Wu, S., & Cai, H. (2025). Short-term power load prediction of VMD-LSTM based on ISSA optimization. *Applied Sciences*, *15*(9), 5037. https://doi.org/10.3390/app15095037
[25] Soomro, A. H., Larik, A. S., Mahar, M. A., Sahito, A. A., Soomro, A. M., & Kaloi, G. S. (2021). Dynamic voltage restorer—A comprehensive review. *Energy Reports*, *7*, 6786–6805. https://doi.org/10.1016/j.egyr.2021.09.004
[26] Kazanci, O. B., & Shukuya, M. (2022). A theoretical study of the effects of different heating loads on the exergy performance of water-based and air-based space heating systems in buildings. *Energy*, *238*, 122009. https://doi.org/10.1016/j.energy.2021.122009
[27] Sun, Y., Tang, X., & Jiang, Y. (2025). Bitcoin volatility forecasting based on time series decomposition and deep learning model. In *Proceedings of the 2nd International Conference*, *404*, 194–201. https://doi.org/10.3233/FAIA250121
[28] Shan, Y., Li, W., Li, J., & Ma, D. (2024). Short-term wind power prediction model based on MIC-CEEMDAN-VMD-CNN-BiLSTM. In *2024 2nd Power Electronics and Power System Conference*, 184–188. https://doi.org/10.1109/PEPSC63375.2024.10823539
[29] Zhang, J., & Duan, H. (2023). Enhanced LFTSformer: A novel long-term financial time series prediction model using advanced feature engineering and the ds encoder informer architecture. *arXiv Preprint: 2310.01884.*
[30] Yuan, Z., Lin, X., Xu, Y., Dai, R., Yang, C., Zhao, L., & Han, Y. (2025). The VMD-Informer-BiLSTM-EAA hybrid model for predicting zenith tropospheric delay. *Remote Sensing*, *17*(4), 672. https://doi.org/10.3390/rs17040672
[31] Xu, X., & Zhang, Y. (2021). Corn cash price forecasting with neural networks. *Computers and Electronics in Agriculture*, *184*, 106120. https://doi.org/10.1016/j.compag.2021.106120
[32] Hai, T., Basem, A., Alizadeh, A., Sharma, K., Jasim, D. J., Rajab, H., . . . , & Maleki, H. (2024). Optimizing Gaussian process regression (GPR) hyperparameters with three metaheuristic algorithms for viscosity prediction of suspensions containing microencapsulated PCMs. *Scientific Reports*, *14*(1), 20271. https://doi.org/10.1038/s41598-024-71027-9
[33] Xu, X., & Zhang, Y. (2023). Price forecasts of ten steel products using Gaussian process regressions. *Engineering Applications of Artificial Intelligence*, *126*, 106870. https://doi.org/10.1016/j.engappai.2023.106870
[34] Jin, B., & Xu, X. (2024). Machine learning coffee price predictions. *Journal of Uncertain Systems*, *17*(4), 2450023. https://doi.org/10.1142/S1752890924500235
[35] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*, *38*(4), 1346–1364. https://doi.org/10.1016/j.ijforecast.2021.11.013
[36] Bantis, E., Clements, M. P., & Urquhart, A. (2023). Forecasting GDP growth rates in the United States and Brazil using Google Trends. *International Journal of Forecasting*, *39*(4), 1909–1924. https://doi.org/10.1016/j.ijforecast.2022.10.003
[37] Kim, D.-H., Kim, D.-J., & Choi, S.-Y. (2025). A Variational-Mode-Decomposition-Cascaded long short-term memory with attention model for VIX prediction. *Applied Sciences*, *15*(10), 5630. https://doi.org/10.3390/app15105630
[38] Li, Y., Jiang, S., Li, X., & Wang, S. (2022). Hybrid data decomposition-based deep learning for Bitcoin prediction and algorithm trading. *Financial Innovation*, *8*(1), 31. https://doi.org/10.1186/s40854-022-00336-7
[39] Raj, P., Bera, K., & Selvaraju, N. (2025). Power of decomposition in volatility forecasting for Bitcoins. *Pacific-Basin Finance Journal*, *93*, 102839. https://doi.org/10.1016/j.pacfin.2025.102839
[40] Zhou, Z., Chen, J., Yang, Z., Wu, W., & Ding, H. (2024). Research on urban power load forecasting based on improved LSTM. *Frontiers in Energy Research*, *12*, 1443814. https://doi.org/10.3389/fenrg.2024.1443814
[41] Fernández-Avilés, G., Mattera, R., & Scepi, G. (2024). Factor-augmented autoregressive neural network to forecast NOx in the city of Madrid. *Socio-Economic Planning Sciences*, *95*, 101958. https://doi.org/10.1016/j.seps.2024.101958
[42] Mahjoub, S., Chrifi-Alaoui, L., Marhic, B., & Delahoche, L. (2022). Predicting energy consumption using LSTM, multi-layer GRU and Drop-GRU neural networks. *Sensors*, *22*(11), 4062. https://doi.org/10.3390/s22114062
[43] Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, 1724–1734. https://doi.org/10.3115/v1/D14-1179

How to Cite: Maatallah, M., Fariss, M., Asaidi, H., & Bellouki, M. (2026). VMD-RDIC-DL: A Composite Relevance-Driven Hybrid Decomposition and Deep Learning Framework for Cryptocurrency Forecasting. *Journal of Computational and Cognitive Engineering.* https://doi.org/10.47852/bonviewJCCE62029131