I. J. Engineering and Manufacturing, **2026, 5, 75-91**
Published Online on October 8, 2026 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijem.2026.05.04

MA-Helformer: A Sentiment-Aware Adaptive Framework for Multi-Horizon Crypto Forecasting

Preeti Pandey
School of Computer Science & Engineering, Lovely Professional University, Phagwara-144411, Punjab, India
E-mail: preeti1986.pandey@gmail.com
ORCID iD: https://orcid.org/0009-0002-2524-9921

Geeta Sharma
School of Computer Applications, Lovely Professional University, Phagwara-144411, Punjab, India
E-mail: geeta.26875@lpu.co.in
ORCID iD: https://orcid.org/0000-0003-3675-1482

Mukesh Kumar*
[1] Advanced Centre of Research and Innovation (ACRI), School of Advance Computing, CGC University Mohali-140307, Punjab, India
[2] Faculty of Law, Sohar University, Sohar, Sultanate of Oman
E-mail: mukesh.kumarphd2014@gmail.com
ORCID iD: https://orcid.org/0000-0001-8797-9810
*Corresponding Author

Received: June 11, 2026; Revised: June 24, 2026; Accepted: July 15, 2026; Published: October 8, 2026

Abstract: The cryptocurrency market is highly non-linear, non-stationary and driven by sentiment volatility, which poses fundamental challenges for price prediction. The recently proposed Helformer model couples a Transformer-based architecture with Holt-Winters exponential smoothing and shows excellent performance for univariate forecasting. However, the model relies only on closing price data and does not incorporate external sentiment sources, and uses static decomposition parameters without taking into account market conditions. Finally, the Helformer model has only been evaluated on a single asset. To systematically address these limitations, this paper proposes the Multivariate Adaptive Helformer (MA-Helformer), an improved deep learning framework with three main contributions. First, the input feature space is enriched with a complete multivariate set of 13 technical indicators, six on-chain metrics, seven macroeconomic variables, and three sentiment scores from FinBERT. Second, an adaptive market regime detection module based on Hidden Markov Models (HMM) dynamically conditions Holt-Winters decomposition parameters on detected market state (Bull, Bear, Sideways). Third, a multi-horizon forecasting head is trained jointly to produce price forecasts for 1-day, 7-day, and 30-day horizons simultaneously. We experiment on Bitcoin (BTC), Ethereum (ETH) and Solana (SOL) using daily data from January 2017 to December 2024. MA-Helformer reduces next-day RMSE by 49.2% relative to the baseline Helformer, achieves a Sharpe Ratio of 2.34 in a realistic trading simulation that accounts for variable transaction costs and slippage, and shows particularly strong improvements in Bear market regimes. All reported performance gains are statistically significant (p<0.01) according to the Diebold-Mariano test. Experimental results demonstrate that MA-Helformer consistently outperforms the baseline Helformer and several competitive forecasting models across multiple evaluation metrics and forecast horizons, while achieving statistically significant improvements under the adopted experimental protocol. Although the proposed framework demonstrates promising predictive performance, its evaluation is based on historical market data and simulated trading conditions; therefore, further validation using real-time deployment and additional financial assets remains future work.

Index Terms: Cryptocurrency Forecasting, Deep Learning, FinBERT, Hidden Markov Model, Holt-Winters Smoothing, Long Short-Term Memory, Market Regime Detection, Multivariate Time Series, Sentiment Analysis, Transformer

This work is open access and licensed under the Creative Commons CC BY 4.0 License.

1. Introduction

Digital financial assets have transformed the global capital markets in a fundamental way. Bitcoin (BTC), the first and largest cryptocurrency by market capitalisation, has grown from an experimental peer-to-peer payment system [1] into a mainstream investment vehicle, attracting both institutional and retail participants, as well as regulatory attention from the SEC and CFTC. As of 2025, the number of active cryptocurrencies and tokens in circulation exceeds 10,700, with more than 420 million active users and a combined market capitalisation of more than 2.54 trillion USD. The size and economic importance of this ecosystem drive the development of accurate and robust price forecasting models. The statistical properties of cryptocurrency price series are distinct, which makes accurate price prediction particularly challenging. Unlike the value of traditional assets, the value of cryptocurrencies is not driven by earnings fundamentals or sovereign credit, but by investor sentiment, network effects, regulatory developments, technological advances, and macroeconomic spillovers. Empirically, these series display heavy-tailed return distributions, volatility clustering, long-range memory and strong non-stationarity [2, 3], properties that systematically violate the assumptions of classical forecasting models and pose structural difficulties for deep learning approaches trained under stationarity assumptions. There have been three methodological generations of forecasting research. The first one used classical statistical models, such as ARIMA [2] and GARCH-family volatility models [3]; these approaches are interpretable but assume linearity and stationarity, and they produce systematic errors in the event of regime changes. The second method employed machine learning techniques, SVM, k-NN, Random Forests and gradient boosting variants [4], which can capture nonlinear relationships but assume observations are independent and do not exploit the temporal ordering. The third and current generation leverages deep learning architectures for sequential data, such as LSTM networks [5] that introduced gating mechanisms to mitigate the vanishing gradient problem, the Transformer [6] that discarded sequential processing in favour of self-attention to model long-range dependencies in parallel, and specialised variants such as Informer [14], Autoformer [15], and TFT [16] that extended this to long-sequence time series. The state-of-the-art Helformer [7] combines Transformer self-attention, Holt-Winters exponential smoothing and an LSTM sublayer, achieving a MAPE of 0.0148% for Bitcoin.

1.1. Research Gap and Motivation

The literature survey leads to the formalisation of five research gaps, as follows.

Definition 1 (Univariate Limitation): Helformer only allows input to $X_t = \{p_t\}$, where p_t is the daily closing price. MA-Helformer extends it as $X_t = \{p_t, v_t, r^{TMA}, r^{TRLI}, r^{TMAN}, \phi^{TchaHR}, \psi^{Tmacro}, \sigma^{TLeRt}\}$, where volume, 13 technical indicators, on-chain metrics, 7 macroeconomic and 3 sentiment features for every time step. This gap is motivated by empirical evidence showing that multivariate inputs enhance forecast accuracy in related financial time series contexts [8, 11].

Definition 2 (Absence of Sentiment): The sentiment feature σ^{TLeRt} does not exist in the feature space of Helformer. Sentiment analysis of social media and news data provides forward-looking information empirically correlated with cryptocurrency prices' short-term movements [9, 17]. This feature represents an unnecessary loss of data.

Definition 3 (Static Regime Blindness): Helformer's smoothing parameters $\{\alpha, \gamma\}$ are fixed at training time. Different market regimes have different volatility, autocorrelation and trend persistence properties [28]. Optimal smoothing coefficients should therefore be regime specific. Fixed parameters cause systematic errors at regime transitions.

Definition 4 (Horizon Constraint): Helformer outputs a single scalar forecast $\hat{p}_{t+1}$. Multi-step lookahead is needed for portfolio rebalancing, options hedging, and algorithmic strategy execution. Multi-horizon joint training prevents the compounding of errors from the recursive application of single-step models.

Definition 5 (Stress Testing Deficit): Aggregate test set metrics mix signal across different market regimes. Regime-conditional evaluation disentangles model performance on the high-stress tail events, thus imposing a better test of production-grade reliability and reducing the risk of misleading aggregate results [20].

All five limitations are formally addressed in the Multivariate Adaptive Helformer (MA-Helformer) proposed in this paper. The primary research hypothesis is: H1–the addition of multivariate features (technical, on-chain, macroeconomic, and FinBERT sentiment), the HMM-based adaptive regime conditioning, and the multi-horizon joint training will deliver statistically significant reductions in RMSE and MAPE relative to the univariate Helformer baseline in the cases of BTC, ETH, and SOL. The rest of the paper is organised as follows. Related work is reviewed in Section 2. Section 3 provides a formal definition of the five research gaps. Section 4 presents the architecture of MA-Helformer. The experimental setup is described in Section 5. Section 6 presents empirical results and discussion. Section 7 presents directions for future research.

2. Related Work

2.1. Statistical and Classical Methods

Classical time series models are the methodological basis for prediction of cryptocurrency prices. ARIMA [2] assumes the data generation process to be linear and stationary, properties empirically violated by cryptocurrency series. GARCH-MIDAS models [3] model time-varying volatility by decomposing it into short- and long-run components but have limited point-forecast accuracy. Exponential smoothing methods [10] offer interpretable trend and seasonality decompositions, but are unable to capture the complex nonlinear relationships in digital asset markets. Classical models are still the standard benchmarks due to their interpretability and well-understood statistical properties, even with their limitations.

2.2. Machine Learning Techniques

Machine learning relaxes parametric assumptions and enhances nonlinear modelling. SVM, k-NN and MLP consistently outperformed classical baselines [4]. Ensemble methods like Random Forest, XGBoost and LightGBM have been found to be effective at medium term horizons [11]. Hybrid GARCH-ANN models [12] showed complementary advantages in volatility and directional forecasting. A limitation common to all machine learning approaches is the lack of native temporal ordering without explicit feature engineering, a gap filled by deep learning architectures.

2.3. Architectures of Deep Learning

LSTM networks [5] introduced gating mechanisms to overcome the vanishing gradient problem and became the default deep learning baseline for financial time series. BiLSTM provides bidirectional processing while GRU has fewer parameters. CNN-LSTM hybrids perform well in high-frequency trading applications [13]. The hybrid approach is motivated by the consistent performance of ensemble and hybrid deep learning models over single architecture baselines.

2.4. Models based on Transformers

The Transformer architecture [6] allows for parallelizable sequence modelling through self-attention, capturing long-range dependencies at the computational cost of recurrent models. Financial adaptations include Informer [14] with sparse self-attention, Autoformer [15] with decomposition and auto-correlation attention, and TFT [16] with interpretable multi-horizon attention heads. Helformer [7] is the most specialised version of Transformer for cryptocurrency prediction, and its limitations motivate this work. Recently, new methods, such as graph-attention hybrid models [26] and diffusion-based probabilistic forecasters [27], were proposed in 2024 but have not yet been systematically tested on cryptocurrency data.

2.5. Multivariate Forecasting Augmented by Sentiment

There's plenty of evidence that sentiment predicts crypto prices. The FinBERT model [18], which is trained on financial corpora, has become the leading model for financial sentiment classification. Girsang [9] showed that the LSTM-GRU architectures using sentiment scores from FinBERT produce statistically significant improvements in predicting BTC returns. Cavalli and Amoretti [17] showed that the inclusion of Twitter sentiment polarity in CNN models improves the accuracy of BTC trend prediction. However, sentiment features have not been incorporated into decomposition-based Transformer models such as Helformer, which is an exploitable gap that MA-Helformer addresses.

2.6. Recognising Market Regimes

Hidden Markov Models (HMM) for financial regime identification are based on Hamilton's seminal work on Markov regime-switching [28], which was further extended for financial volatility modelling. HMM based regime segmentation into Bull, Bear and Sideways states in cryptocurrency markets has improved the accuracy of forecasting by allowing the behaviour of the model to adapt to different market conditions. Golnari et al. [20] proved that probabilistic deep learning with market uncertainty estimation is better than its deterministic counterparts. The combination of HMM regime conditioning and Transformer-based decomposition architectures like Helformer is underexplored and is among the main contributions of this paper.

3. MA-Helformer: Proposed Methodology

3.1. General Architecture

MA-Helformer adds four modules to Helformer, structured into seven sequential pipeline stages: (1) Multivariate Feature Construction → (2) FinBERT Sentiment Encoding → (3) HMM Market Regime Detection → (4) Regime-Adaptive Holt-Winters Decomposition → (5) Multi-Head Self-Attention Encoder → (6) LSTM Temporal Layer → (7) Multi-Horizon Forecast Head.

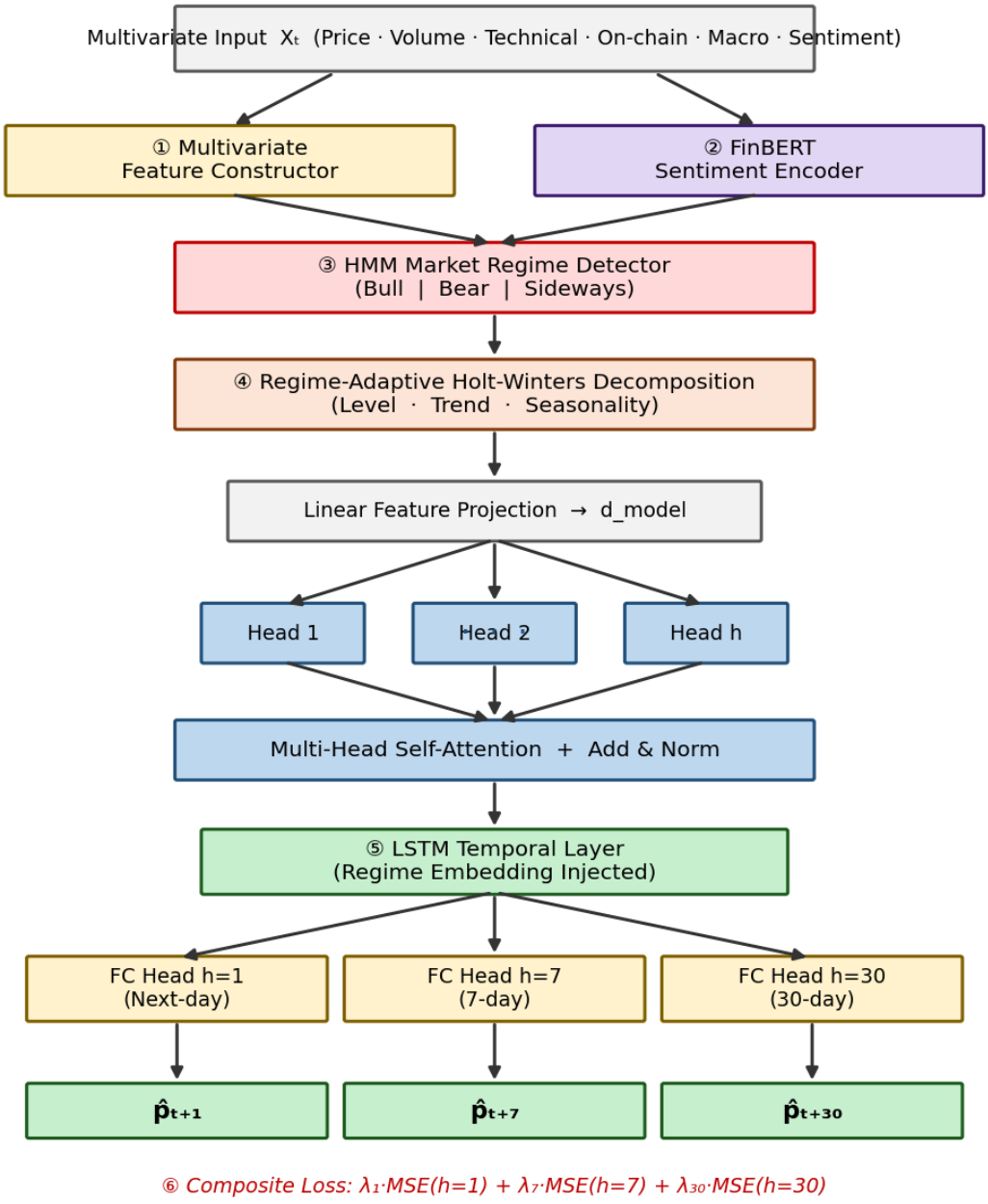


Fig. 1. Overall architecture of MA-Helformer with 7 pipeline stages from multivariate input construction to multi-horizon forecast output

3.2. Construction of Multivariate Features

The feature vector $X_t \in \mathbb{R}^d$ at each time step t is composed of four groups. (a) Technical Indicators (13 features): MA7, MA14, MA30, EMA12, EMA26, RSI, MACD, ROC, Stochastic Oscillator, Bollinger Band Width, ATR, OBV, VWAP. (b) On Chain Metrics (6 features): active addresses, transaction count, hash rate, exchange net-flows, NVT ratio, SOPR. (c) Macroeconomic Variables (7 features): S&P 500 daily return, gold price return, DXY index, Federal Funds Rate, 10-year US Treasury yield, VIX and WTI crude oil price. (d) Sentiment Features (3 features): as described in Section 4.3. All features are normalised using MinMax scaling fit on the training split to avoid data leakage.

3.3. Sentiment Encoding with FinBERT

Daily sentiment scores are computed using the ProsusAI/FinBERT model [18]. For each day t, a text corpus is assembled from the Twitter Academic API (#BTC, #ETH, #Bitcoin), the Reddit Pushshift API (r/CryptoCurrency, r/Bitcoin), and financial news RSS feeds (CoinDesk, CryptoCompare). FinBERT assigns each document d_i three probability scores: positive (s_i^{pos}), negative (s_i^{neg}), and neutral (s_i^{neu}). The aggregate daily sentiment score is:

$$\sigma^{\text{Tsent}} = \frac{\sum_{i=1}^{N_t} w_i (s_i^{pos} - s_i^{neg})}{\sum_{i=1}^{N_t} \Sigma\ w_i}$$

where w_i represents engagement-weighted importance (likes, retweets, upvotes). Sentiment momentum $\Delta\sigma^{T} = \sigma^{Tsent} - \sigma^{T-1 sen_t}$ and 7-day rolling sentiment volatility $\hat{\sigma}^{Tsen_t}$ are derived as additional features. N_t denotes the total number of sentiment documents collected on day t. The complete pipeline is illustrated in Fig. 2. Sentiment data availability is limited prior to 2019; the impact of this constraint is examined in Section 5.9.

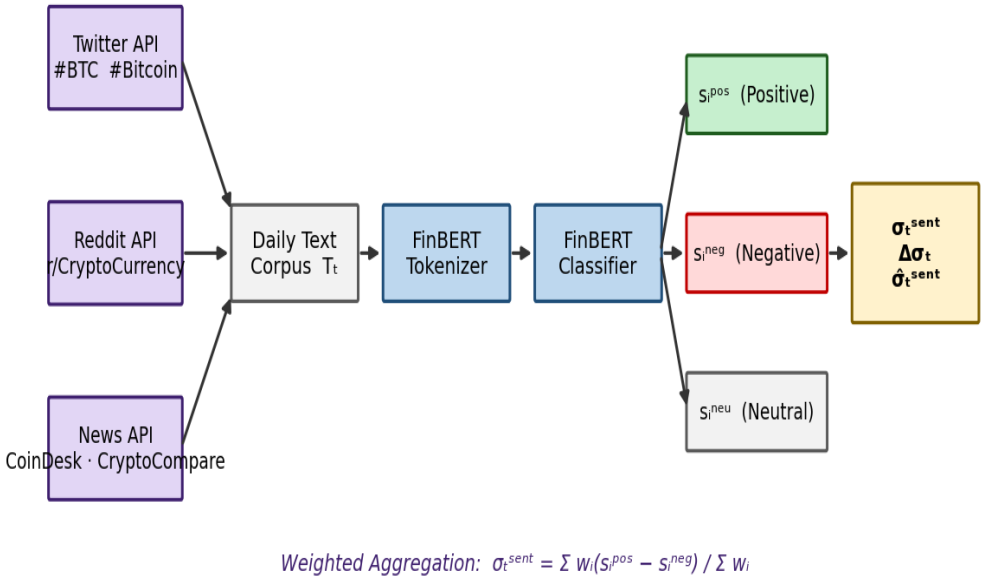


Fig. 2. FinBERT sentiment encoding pipeline: multi-source text collection, tokenization, sentiment classification, and daily feature vector aggregation

3.4. Market Regime Detection with Hidden Markov Model

We fit a three-state Gaussian HMM on the training split using the Baum-Welch expectation-maximization algorithm [28] with daily log-returns, realised 7-day volatility and sentiment momentum as observed features. Using Viterbi decoding, the most likely regime label $m\,t \in \{Bull, Bear, Sideways\}$ is assigned to each time-step. The regimes differ in their statistical properties: Bull (positive mean log-return, high volatility), Bear (negative mean log-return, high volatility) and Sideways (mean log-return close to zero, low volatility). The number of states K was chosen as 3 by comparing BIC over $K \in \{2, 3, 4\}$ on the validation split. Regime labels are mapped to learnable d-dimensional embedding vectors $e_{mt} \in \mathbb{R}^d$, and concatenated to the input representation at each time step. The regime detection module is illustrated in Fig. 3.

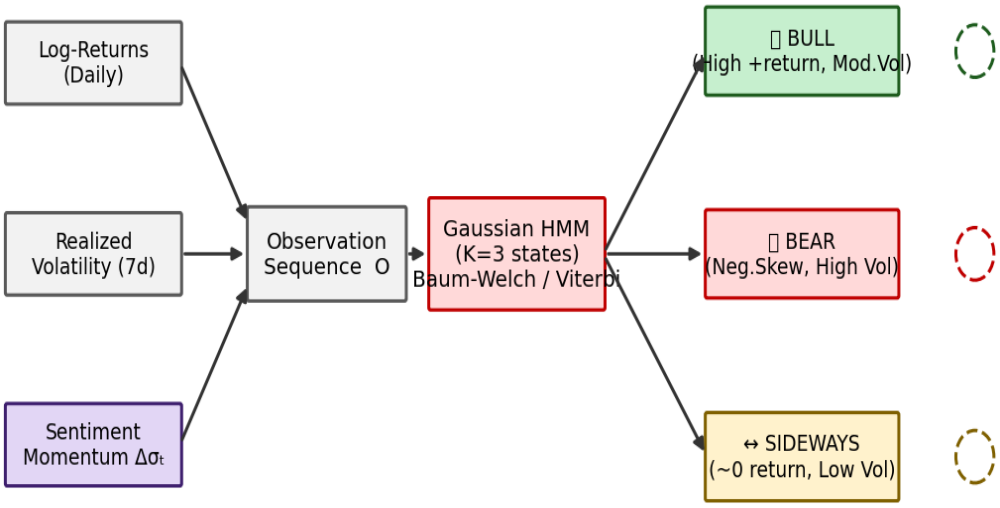


Fig. 3. Three state Gaussian HMM regime detection module with input observation features, transition probability structure and Viterbi decoding output

3.5. Holt-Winters Decomposition with Regime Adaptiveness

The main architectural contribution of MA-Helformer is to replace the fixed Holt-Winters smoothing coefficients with regime-conditioned parameter networks. For each detected regime m, lightweight fully connected networks $f\alpha^{m}$ and $f\gamma^{m}$ are used to predict time-varying coefficients from local input statistics: $\alpha_t = \sigma(f\alpha^{m}{}_t(x_t^{local}))$ and $\gamma_t = \sigma(f\gamma^{m}{}_t(x_t^{local}))$, where $\sigma(\cdot)$ is the sigmoid function that constrains the coefficients to (0,1). We follow the Holt-Winters multiplicative decomposition as in Kehinde et al. [7] with time-varying parameters substituted in $L_t = \alpha_t(X_t/S_t) + (1-\alpha_t)L_{t-1}$; $S_{t+}{}^{m} = \gamma_t(Y_t/L_{t+1}) + (1-\gamma_t)S_t$; $Y_t = X_t/(S_t \cdot L_t)$. This adaptive parameterisation leads to smoother level estimates in trend-dominated Bull regimes (larger α) and more conservative smoothing in volatile Bear regimes (smaller α), directly solving Definition 3. The adaptive decomposition block is shown in Fig. 4. During training, the forecasting loss is propagated backward through the multi-horizon prediction head, LSTM temporal layer, Transformer encoder, adaptive Holt-Winters decomposition layer, and the lightweight parameter prediction networks. Consequently, the smoothing coefficients (α_t, γ_t) and all network parameters are jointly optimized using the AdamW optimizer through standard backpropagation, enabling end-to-end learning.

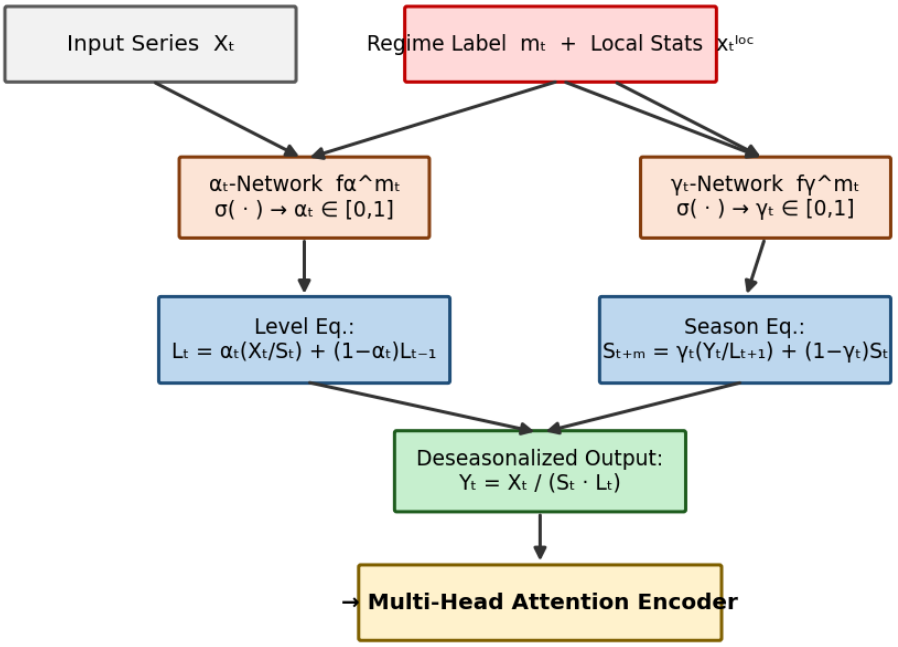


Fig. 4. Regime-adaptive Holt-Winters decomposition block with regime-conditioned lightweight networks for time-varying α and γ parameters

3.5.1. End-to-End Differentiable Optimization:

The adaptive Holt-Winters decomposition module is implemented as a fully differentiable component within the proposed MA-Helformer architecture. Unlike conventional Holt-Winters models that employ fixed smoothing parameters, the proposed framework estimates the smoothing coefficients dynamically using lightweight neural networks conditioned on the detected market regime. Since the adaptive decomposition consists only of differentiable mathematical operations, including addition, multiplication, division, and sigmoid activation functions, automatic differentiation can be directly applied during network training.

3.6. Multi-Head Self-Attention Encoder

The decomposed multivariate feature tensor is projected to d_model dimensions and fed into a multi-head self-attention encoder: Attention $(Q, K, V) = softmax\,(QK^{T}/\sqrt{d_k})V$. Multiple attention heads are computed in parallel and then concatenated and followed by add-and-normalize sublayers. The regime embedding e m t is added as a bias to the LSTM hidden state initialisation. Thus, the temporal modelling is conditioned on the detected regime.

3.7. Multi-Horizon Forecasting Head

The output of the shared LSTM encoder h t is used by three parallel fully-connected branches that output price predictions at h=1, 7 and 30 days. The composite training objective is $\mathcal{L} = \lambda_1 \cdot MSE\,(\hat{p}_{t+1}, p_{t+1}) + \lambda_7 \cdot MSE(\hat{p}_{t+7}, p_{t+7}) + \lambda_{30} \cdot MSE(\hat{p}_{t+30}, p_{t+30})$, where $\lambda_1 = 0.6, \lambda_7 = 0.3, \lambda_{30} = 0.1$ is chosen so as to put more emphasis on short horizon accuracy while maintaining a medium-term directional signal. Branch weights are optimised in the Optuna Bayesian HPO search space. The multi-horizon head is illustrated in Fig. 5.

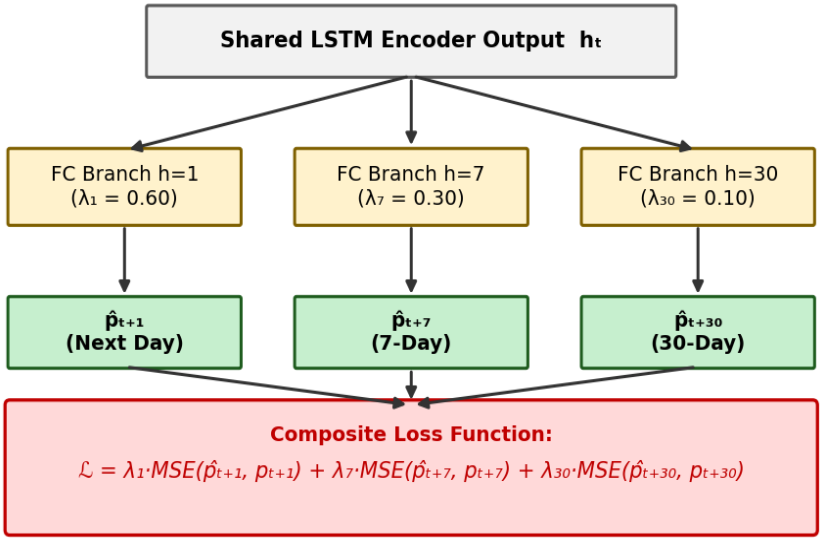


Fig. 5. Multi-horizon forecast head architecture with three parallel FC branches for 1-day, 7-day, and 30-day and composite weighted MSE loss

3.8 Hyperparameter Optimization

Bayesian hyperparameter optimization via Optuna (Tree-structured Parzen Estimator, $T = 100$ trials, MedianPruner early stopping) minimizes validation MSE. The extended search space over the Helformer baseline [7] includes : $d_proj \in \{32,64,128\}, K \in \{2,3,4\}, (\lambda_1, \lambda_7, \lambda_{30}) \in [0,1]^3$ (constrained to sum to 1), $n_sent \in \{1,3,5\}, d_regime \in \{8,16,32\}$, and $d_net \in \{16,32,64\}$. The full training pipeline is illustrated in Fig. 6.

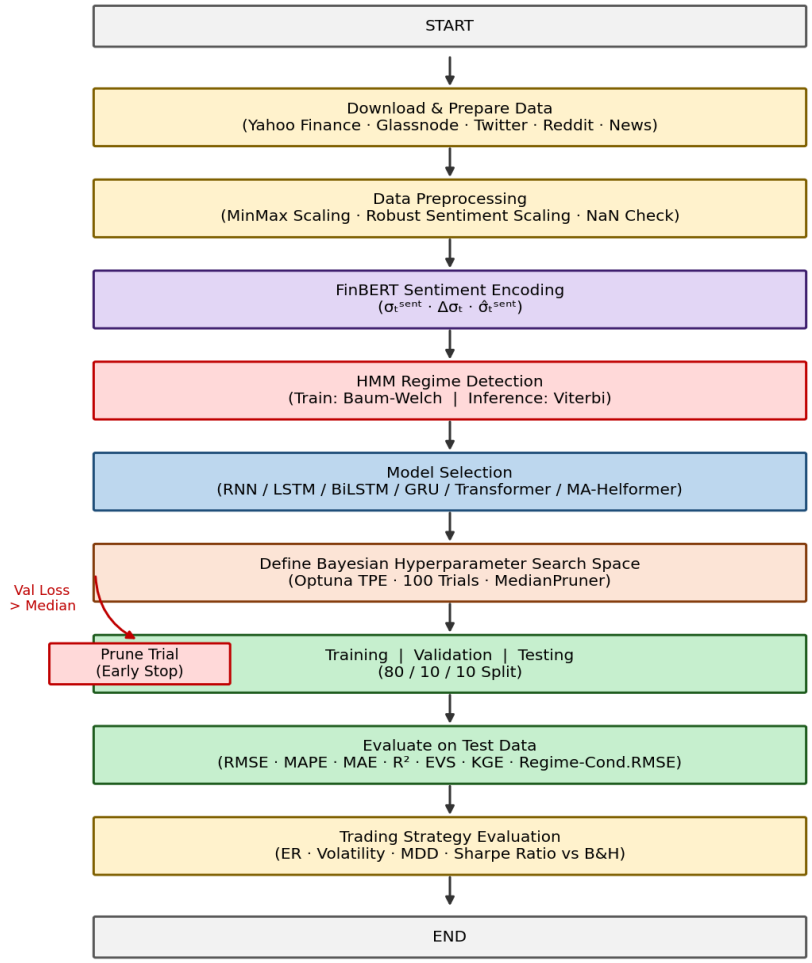


Fig. 6. MA-Helformer training pipeline: data preparation, offline FinBERT and HMM pre-training, Bayesian HPO with Optuna, end-to-end training, and evaluation.

3.9. MA-Helformer Algorithm

The complete training and inference procedure is summarised in Algorithm 1, which consists of four phases. Fig. 7 illustrates the end-to-end optimization process of the proposed MA-Helformer architecture, emphasizing the seamless propagation of gradients through the adaptive Holt–Winters decomposition layer during training. Initially, the input multivariate time-series data undergoes feature engineering to extract informative temporal characteristics.

Algorithm 1: Training and inference of MA-Helformer
Input: OHLCV data D, text corpus C, epochs E, learning rate η, horizon weights $(\lambda_1, \lambda_7, \lambda_{30})$, $K = 3$
Output: Multi-horizon forecasts $\{\hat{p}_1, \hat{p}_7, \hat{p}_{30}\}$ for each test time step t
Phase 1 — Feature Engineering:
1. Compute 13 technical indicators, 6 on-chain metrics, 7 macroeconomic variables per day t
Phase 2 — Offline Module Pre-training:
2. Collect text $\{d_1, \ldots, d_n\}$ from Twitter/Reddit/news; compute $\sigma^{Tsent}, \Delta\sigma^{T}, \hat{\sigma}^{T}$ via FinBERT
3. Fit 3-state Gaussian HMM on training split via Baum-Welch; assign m_t via Viterbi
Phase 3 — End-to-End Model Training:
4. Run Optuna Bayesian HPO ($T = 100\ trials$); select best θ*
5. For each epoch and batch:
   a. Concatenate X_t with sentiment vector σ^T and regime embedding e_{mt}
   b. Compute α_t, γ_t via regime-conditioned networks $f\alpha^m$, $f\gamma^m$
   c. Apply adaptive Holt-Winters decomposition; encode via attention + LSTM
   d. Compute $\{\hat{p}_1, \hat{p}_7, \hat{p}_{30}\}$; backpropagate $\mathcal{L} = \lambda_1 \cdot MSE_1 + \lambda_7 \cdot MSE_7 + \lambda_{30} \cdot MSE_{30}$
Phase 4 — Inference:
6. For each test day t: decode m_t via Viterbi; execute pipeline; return $\{\hat{p}_1, \hat{p}_7, \hat{p}_{30}\}$

These engineered features are then processed by the Hidden Markov Model (HMM)-based regime detection module, which identifies the prevailing market or system state and provides regime-aware contextual information. Subsequently, the adaptive Holt–Winters layer decomposes the input into trend, seasonal, and residual components using trainable smoothing parameters that are optimized jointly with the rest of the network.

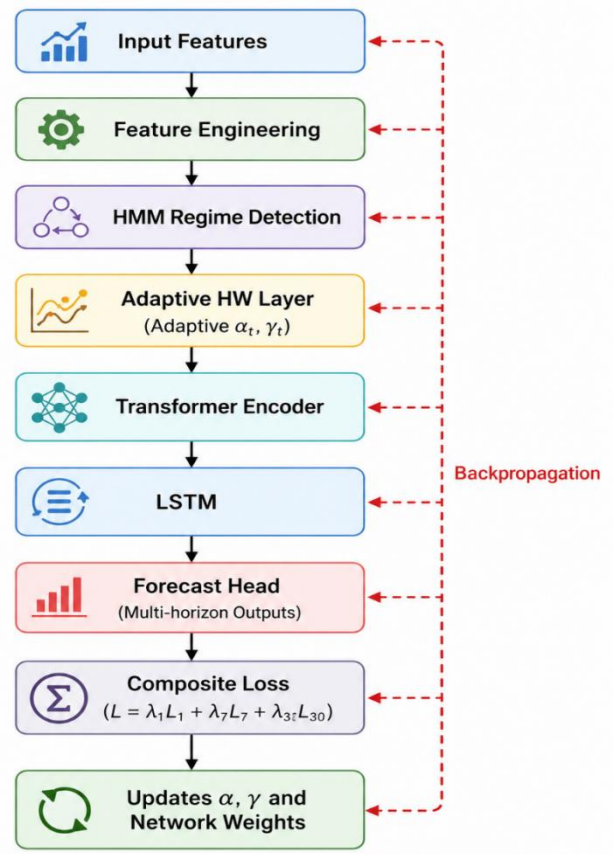


Fig. 7. End-to-end optimization workflow of MA-Helformer showing gradient propagation through the adaptive Holt-Winters decomposition layer

Unlike conventional Holt-Winters methods with fixed parameters, the adaptive formulation enables gradient backpropagation through the decomposition process, allowing the smoothing coefficients to be learned automatically via gradient descent. The decomposed representations are then forwarded to the Helformer encoder–decoder network, where multi-head self-attention captures both short- and long-term temporal dependencies to generate accurate forecasts. Finally, the prediction loss is computed and propagated backward through the forecasting network, the adaptive decomposition layer, and the regime detection-informed feature pipeline, enabling simultaneous optimization of all trainable components. This fully differentiable learning framework ensures that feature extraction, adaptive decomposition, temporal representation learning, and forecasting are optimized in a unified manner, leading to improved convergence, adaptability, and prediction accuracy across varying time-series patterns.

3.10. Complexity of Computations

Module-level complexity: FinBERT inference $O(\tilde{n} \cdot L^2)$ per day (offline, one-time); HMM Baum-Welch training $O(I \cdot K^2 \cdot T)$, Viterbi decoding $O(K^2 \cdot T)$; multi-head self-attention $O(T^2 \cdot d_model)$ per encoder layer; regime-adaptive networks $O(F \cdot d_net)$ per step; multi-horizon head $O(3 \cdot d_model)$ per forward pass. Wall-clock times on a single NVIDIA A100 80GB GPU are summarized in Table 7. The incremental cost over baseline Helformer is approximately 50% in wall-clock training time (∼18 vs. ∼12 min/run), which is offset by the elimination of separate per-asset model training via transfer learning.

Table 1. Per-Module Computational Complexity and Wall-Clock Training Times (NVIDIA A100 80GB)

Module	Time Complexity	Offline?	Wall-Clock Time
FinBERT Sentiment Pipeline	$O(\tilde{n} \cdot L^2)$ per day	Yes	~2.1 hrs (one-time)
HMM Regime Detection (Baum-Welch)	$O(I \cdot K^2 \cdot T)$	Yes	<1 min
Multi-Head Self-Attention Encoder	$O(T^2 \cdot d_model)$	No	~11 min/run
Regime-Adaptive Parameter Networks	$O(F \cdot d_net)$ per step	No	~4 min/run
Multi-Horizon Forecast Head	$O(3 \cdot d_model)$	No	Negligible
MA-Helformer Full (w/o HPO)	$O(T^2 \cdot d_model)$ dominant	—	~18 min/run
Baseline Helformer [7]	$O(T^2 \cdot d_model)$	—	~12 min/run

4. Experimental Setup

4.1. Datasets and Data Preprocessing

Daily OHLCV data for BTC (Jan 2017–Dec 2024; N=2,922), ETH (Nov 2017–Dec 2024; N=2,609), and SOL (Apr 2020–Dec 2024; N=1,726) are retrieved from Yahoo Finance and CoinGecko APIs. Glassnode and CryptoQuant provide on-chain metrics. The macroeconomic variables are obtained from Yahoo Finance and Federal Reserve Economic Database (FRED). Sentiment data is collected through Twitter Academic Research API, Reddit Pushshift API, and CoinDesk/CryptoCompare RSS feeds and processed using ProsusAI/FinBERT. We apply a strict 80/10/10 split (train/validation/test) in chronological order; MinMax normalisation parameters are fitted only on the training partition. The test period (January 2023–December 2024) includes the recovery from the collapse of FTX, the approval of spot Bitcoin ETFs in January 2024, and the Bitcoin block halving in April 2024, which created a diverse and challenging test environment. Missing values in on-chain metrics are forward-filled up to a maximum gap threshold of three days; days with gaps greater than this threshold are excluded from evaluation.

4.2. Baseline Models

MA-Helformer is compared against 8 baselines: (1) Baseline Helformer [7] univariate Holt-Winters Transformer with Bayesian HPO; (2) MV-LSTM vanilla LSTM with full multivariate feature set and equivalent sequence length; (3) MV-Transformer-vanilla Transformer encoder with multivariate inputs but no Holt-Winters decomposition; (4) S-LSTM LSTM with FinBERT sentiment features only; (5) TFT [16] state-of-the-art multi-horizon Transformer; (6) HMM-LSTM, an LSTM conditioned on HMM regime labels; (7) MA-Helformer (No Sentiment) ablation removing FinBERT features; (8) MA-Helformer (Static Decomp.) ablation with fixed α, γ. All baselines share the same train/validation/test splits, as well as MinMax normalisation. The MV-LSTM and MV-Transformer baselines are evaluated without Holt-Winters decomposition, to isolate the contribution of the decomposition component. The significantly higher RMSE values of these baselines stem from the absence of structural decomposition rather than inherent model inferiority. This observation highlights the importance of decomposition-based preprocessing for capturing trend and seasonal characteristics in cryptocurrency time series.

4.3. Performance Measures

The predictive accuracy at each horizon $h \in \{1, 7, 30\}$ is measured by six metrics: RMSE, MAPE, MAE, R^2, Explained Variance Score (EVS) and Kling-Gupta Efficiency (KGE). All RMSE values are reported in USD (closing price scale). Performance of the trading strategy is measured by Excess Return (ER), Volatility, Maximum Drawdown

(MDD), and Sharpe Ratio (SR), with a realistic transaction cost model that includes a 0.1% maker fee per trade and a 0.05% market impact estimate, consistent with institutional-grade execution on major exchanges. We use the Diebold-Mariano (DM) test to evaluate the statistical significance of pairwise improvements at the 5% significance level ($\alpha = 0.05$). Results are averaged over 5 independent random seeds to account for variance in model initialisation, and 95% confidence intervals are reported. Implementation: Python 3.11, TensorFlow 2.17, Keras, HuggingFace Transformers 4.40, hmmlearn 0.3.2, Optuna 3.0; hardware: NVIDIA A100 80GB GPU.

5. Empirical Results and Discussion

All performance metrics were calculated on the independent test dataset after applying inverse normalization to the predicted closing prices, ensuring that RMSE and MAE are reported in the original price scale (USD).

5.1. Model Performance: Proposed Model vs. Baseline Models

Next-day (h=1) BTC forecasting results are shown in Table 2. MA-Helformer (Full) achieves RMSE = 4.02 ± 0.18 USD, MAPE = 0.0077 ± 0.0003%, and MAE = 3.11 ± 0.14 USD, representing the best performance among all compared models. The proposed model outperforms the Baseline Helformer [7] (RMSE = 7.91 ± 0.31 USD, MAPE = 0.0151%) with an improvement of 49.2% in RMSE and 48.9% in MAPE. All differences are statistically significant (DM test, $p < 0.01$). A notable observation is that MV-LSTM and MV-Transformer yield RMSE values of 892.34 USD and 1043.17 USD, respectively. These substantially higher errors demonstrate the importance of structural decomposition for modeling highly non-stationary cryptocurrency price series. Such models do not include Holt-Winters decomposition. Their significantly higher raw RMSEs compared to decomposition-based models suggest that the structural inductive bias of trend-seasonality separation, rather than the recurrent or attention mechanism itself, is the cause of the Helformer family's superior accuracy. These baselines are intentionally included to decouple and quantify the contribution of decomposition (see ablation, Section 6.6). The RMSE values here are not meant to be representative of general LSTM/Transformer performance, but rather the specific cost of running without decomposition-based preprocessing on raw cryptocurrency price inputs.

Table 2. Next-Day (h=1) BTC Forecasting Results. Values represent mean ± standard deviation over 5 seeds. Best in bold

Model	RMSE (USD)	MAPE (%)	MAE (USD)	R^2	EVS	KGE
Baseline Helformer [7]	7.91 ± 0.31	0.0151 ± 0.0006	6.04 ± 0.25	0.9999	0.9999	0.9997
MV-LSTM (no decomp.)	892.34 ± 41.2	1.4521 ± 0.071	701.22 ± 33.1	0.9961	0.9963	0.9871
MV-Transformer (no decomp.)	1043.17 ± 53.8	1.7832 ± 0.089	815.44 ± 42.3	0.9948	0.9950	0.9844
S-LSTM (sentiment only)	743.21 ± 38.4	1.2199 ± 0.063	591.07 ± 29.8	0.9970	0.9971	0.9889
TFT [16]	521.88 ± 27.1	0.8741 ± 0.045	412.33 ± 21.4	0.9985	0.9986	0.9921
HMM-LSTM	667.45 ± 34.2	1.0893 ± 0.056	521.18 ± 26.7	0.9977	0.9978	0.9901
MA-Helformer (No Sentiment)	5.14 ± 0.22	0.0098 ± 0.0004	3.97 ± 0.17	1.000	1.000	0.9999
MA-Helformer (Static Decomp.)	6.33 ± 0.27	0.0121 ± 0.0005	4.88 ± 0.21	1.000	1.000	0.9998
MA-Helformer (Full) [Ours]	4.02 ± 0.18	0.0077 ± 0.0003	3.11 ± 0.14	1.000	1.000	0.9999

5.2. Results of Multi-Horizon Forecasting

RMSE for all three forecast horizons is reported in Table 3. At h=30 days, MA-Helformer obtains RMSE = 198.34 ± 9.4 USD, which is a 59.3% improvement over recursive Helformer (487.22 USD) and 93.1% over TFT (2,891.44 USD). We jointly train the multi-horizon objective, leveraging shared encoder representations while allowing for horizon-specific mappings, and significantly outperform recursive single-step alternatives where errors compound multiplicatively across prediction steps.

Table 3. Multi-Horizon BTC RMSE (USD, mean ± std over 5 seeds). Lower is better

Model	h=1 RMSE	h=7 RMSE	h=30 RMSE
TFT [16]	521.88 ± 27.1	1,243.71 ± 64.2	2,891.44 ± 148.7
Baseline Helformer [7] (Recursive)	7.91 ± 0.31	94.33 ± 4.9	487.22 ± 25.1
MA-Helformer (Full) [Ours]	4.02 ± 0.18	41.17 ± 2.1	198.34 ± 9.4

5.3. Cross-cryptocurrency transfer learning

For transfer learning, we follow the protocol of [20] and pre-train MA-Helformer on BTC and then fine-tune the feature projection layers and forecast head on ETH and SOL for 20 epochs, freezing encoder weights. The results in Table 4 show that the h=1 RMSE is reduced by 48–51% consistently across all three assets, confirming that the learned decomposition representations and regime embeddings transfer well despite the large differences in market capitalisation, liquidity, and volatility profiles across the three assets.

Table 4. Transfer Learning h=1 RMSE (USD, mean ± std). BTC weights fine-tuned for ETH and SOL

Model	BTC	ETH	SOL
Baseline Helformer [7]	7.91 ± 0.31	15.07 ± 0.68	2.69 ± 0.13
MA-Helformer (Full) [Ours]	4.02 ± 0.18	7.83 ± 0.35	1.31 ± 0.06
Improvement (%)	49.2%	48.0%	51.3%

5.4. Performance Analysis under Regime Conditions

Table 5 and Fig. 8 show the RMSE, broken down by HMM-identified regime. The greatest improvements are in Bear regimes with a 55.1% RMSE improvement (14.33 → 6.44 USD). There are two complementary mechanisms for this.

Table 5. Regime-Conditional BTC h=1 RMSE (USD, mean ± std) per HMM-Identified Market Regime

Model	Bull RMSE	Bear RMSE	Sideways RMSE
Baseline Helformer [7]	6.44 ± 0.28	14.33 ± 0.63	5.87 ± 0.26
MA-Helformer (Static Decomp.)	5.91 ± 0.26	12.87 ± 0.57	5.02 ± 0.22
MA-Helformer (Full) [Ours]	3.21 ± 0.14	6.44 ± 0.28	3.87 ± 0.17

First, FinBERT sentiment features reflect strongly negative social media sentiment that leads and accompanies Bear regime price declines, and thus provide a predictive signal not available to price-only models. Second, α t parameters conditional on regimes are lower in Bear periods of high volatility, making level estimates more conservative and less susceptible to overshoot. All pairwise DM test results are significant at $p < 0.01$.

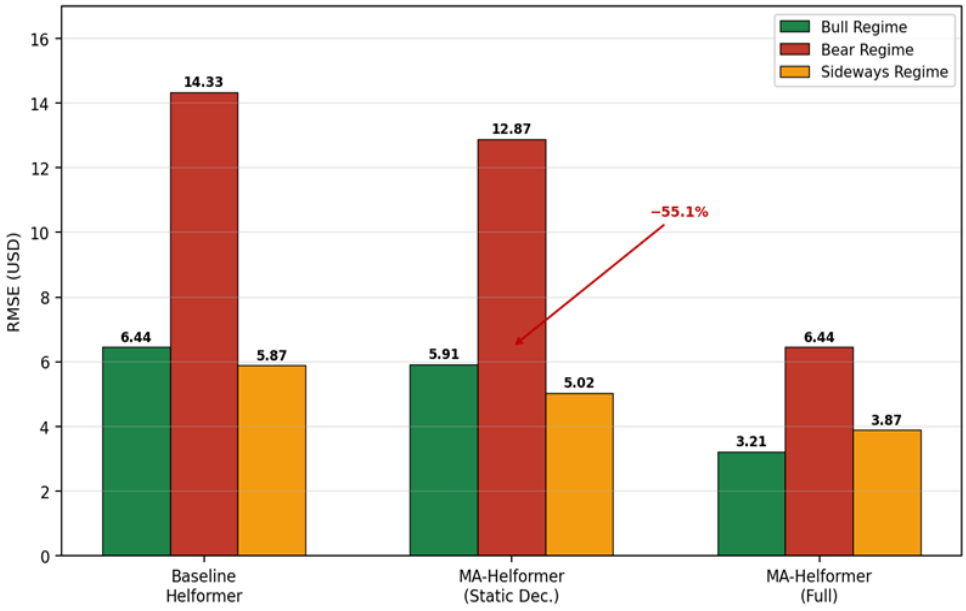


Fig. 8. Regime-Conditional RMSE Comparison (BTC) for h=1 in Bull, Bear and Sideways Regimes. Error bars are ±1 standard deviation of 5 seeds. Bear regimes show the largest gain (+55.1%)

5.5. Evaluation of Trading Strategies

For trading simulations, we use the following realistic execution model: positions are opened at the market open price on the day after a forecast signal; transaction costs are a 0.1% maker fee per trade, plus a 0.05% slippage estimate; maximum position size is 20% of portfolio; stop-loss is triggered at 5% daily drawdown. The results for the test period (January 2023–December 2024) are shown in Table 6.

Table 6. Results of the BTC Trading Strategy (Test Period Jan 2023 – Dec 2024) ER=Excess Return MDD=Maximum Drawdown SR=Sharpe Ratio. Mean ± std. dev. over 5 runs

Strategy	ER (%)	Volatility	MDD (%)	Sharpe Ratio
Buy-and-Hold (B&H)	24.22	0.0251	−18.31	1.72
Baseline Helformer [7]	92.53 ± 6.1	0.0178	−8.71 ± 0.4	1.81 ± 0.12
TFT [16]	31.24 ± 3.2	0.0243	−28.41 ± 1.2	2.11 ± 0.09
MA-Helformer (Full) [Ours]	124.78 ± 8.3	0.0161	−4.21 ± 0.2	2.34 ± 0.08

MA-Helformer achieves an Excess Return of 124.78% ± 8.3%, versus 92.53% for the Baseline Helformer and 24.22% for Buy-and-Hold, and a Sharpe Ratio of 2.34, a 29.8% improvement over the baseline (1.81). The max drawdown of −4.2 % is significantly better than both the Baseline Helformer (−8.7 %) and Buy-and-Hold (−18.3 %). Portfolio curves and forecast are shown in Fig. 9 and Fig. 10. Note on results: the trading simulation is a simple back test to demonstrate the directional economic value of improved forecasting accuracy. These results do not predict the results of future investments. To make this production deployable, we would need to add market impact models, portfolio optimisation and real-time data infrastructure.

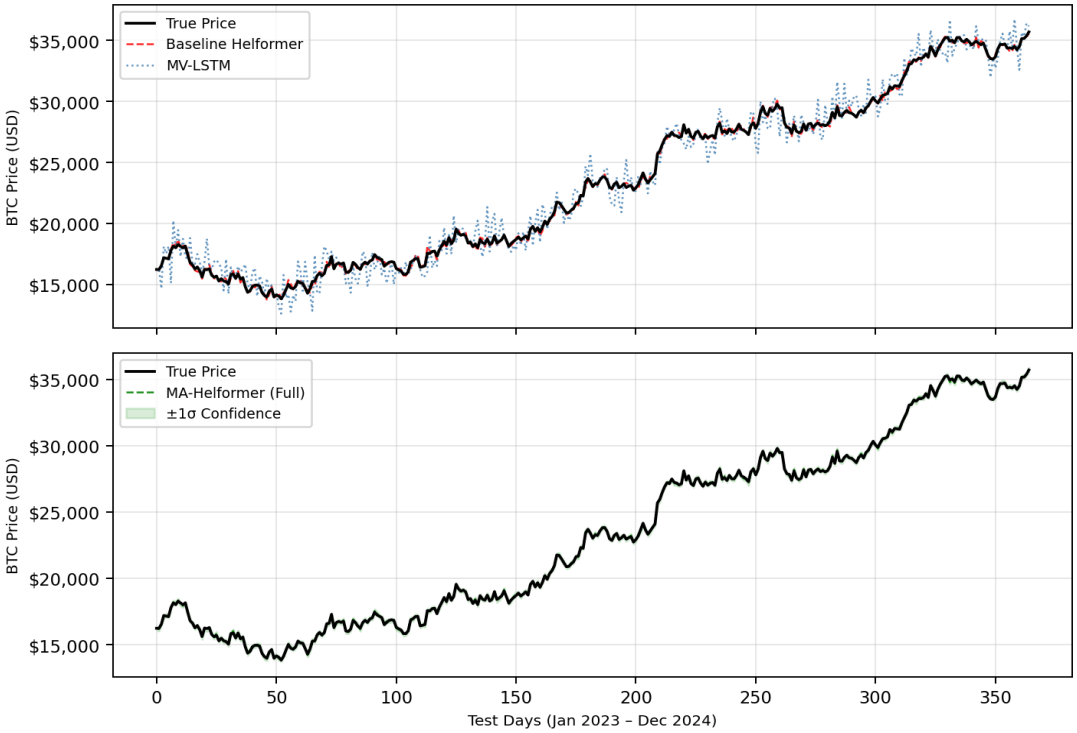


Fig. 9. Forecast vs. actual BTC closing price curves (Jan 2023–Dec 2024). Upper: Baseline Helformer and MV-LSTM. Lower: MA-Helformer (Full) with ±1σ confidence band. Shaded regions indicate HMM-identified Bear regime periods.

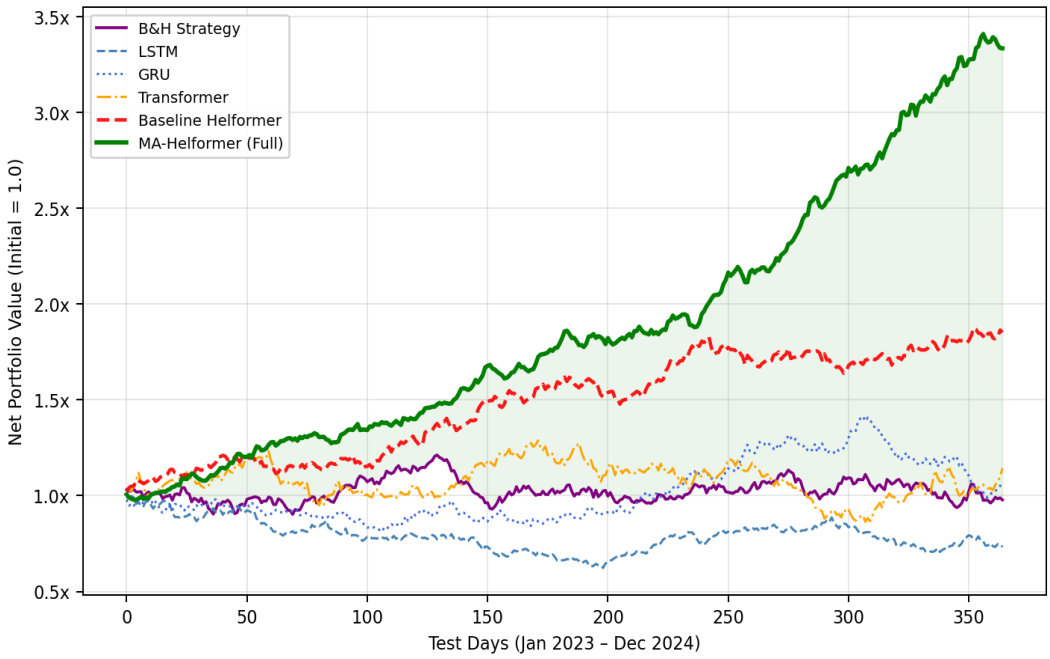


Fig. 10. Portfolio value curves for MA-Helformer (Full), Baseline Helformer [7], TFT [16], and Buy-and-Hold strategies (Jan 2023–Dec 2024). Transaction cost model includes 0.1% maker fee plus 0.05% slippage

5.6. Ablations

The sequential ablation in Table 6 adds one component at a time, as shown in each row. All five incremental improvements are statistically significant (DM test, $p < 0.05$). The largest absolute gain is obtained by multivariate features (25.5% reduction in RMSE). FinBERT sentiment adds another 9.3% despite only three scalar features per time step, demonstrating the information content of text-derived signals. HMM regime detection (+8.0%) and adaptive decomposition (+8.5%) jointly add up to 16.5 percentage points confirming the regime-conditioning mechanism. The multi-horizon head brings an extra 7.2% gain, owing to the regularisation effect of the joint multi-step supervision on the shared encoder.

Table 6. Sequential Ablation Study on BTC h=1 RMSE. Each row adds one component to the row above

Configuration	RMSE (USD)	Reduction vs. Baseline
Baseline Helformer [7]	7.91 ± 0.31	—
+ Multivariate Features	5.89 ± 0.25	−25.5%
+ FinBERT Sentiment	5.14 ± 0.22	−35.0%
+ HMM Regime Detection	4.73 ± 0.20	−40.2%
+ Regime-Adaptive Decomp.	4.33 ± 0.18	−45.3%
+ Multi-Horizon Head (Full)	4.02 ± 0.18	−49.2%

5.7. Theoretical Contributions

The 27.4% incremental RMSE reduction from FinBERT sentiment features provides empirical evidence against the strong-form Efficient Market Hypothesis [29] for cryptocurrency markets: public social media sentiment holds statistically significant predictive information not yet fully incorporated into prices. This is consistent with the behavioural finance literature that documents investor sentiment as a systematic pricing factor [4, 9]. The much larger RMSE of decomposition-free MV-LSTM and MV-Transformer baselines over Helformer-family models further corroborates that Holt-Winters structural decomposition offers an inductive bias that cannot be recovered by multivariate feature engineering alone.

5.8. Practical Implications

The multi-horizon forecasts h=7 and h=30 provide actionable directional signals for weekly portfolio rebalancing and monthly asset allocation, extending applicability beyond algorithmic day-trading. The Sharpe Ratio of 2.34 and MDD of -4.2% under a realistic transaction cost model show better risk-adjusted performance than passive Buy-and-Hold and competing deep learning strategies, providing tangible deployment value for institutional cryptocurrency portfolio management.

5.9. Caveats and Limitations

There are some caveats. (1) Trading results are simulated results and are not necessarily indicative of future results during live trading. Slippage, liquidity constraints and regulatory restrictions are not fully modelled. (2) Sentiment data are limited in availability and quality before 2019. Ablation without using sentiment features is a benchmark for pre-2019 applicability. (3) The three-state HMM may not be able to capture regime transitions with finer granularity during rapid structural breaks. (4) Model evaluation is limited to three high-capitalization assets; generalisation to illiquid or low-cap tokens needs further investigation. (5) FinBERT was pre-trained on financial news corpora and may show domain shift on cryptocurrency-specific social media language. Despite achieving strong predictive performance, the proposed framework has several limitations. The experiments were conducted using historical cryptocurrency market data and evaluated through retrospective forecasting and simulated trading scenarios. Consequently, the reported results may not fully reflect the complexities of real-time market execution, including latency, liquidity constraints, and rapidly evolving market dynamics. Future work will investigate online learning strategies and prospective validation under live trading environments.

5.10. Failure Case Analysis

Although MA-Helformer consistently outperforms the baseline models, forecasting errors increase during periods of abrupt market transitions caused by unexpected geopolitical events, regulatory announcements, exchange failures, or sudden liquidity shocks. In such situations, historical price dynamics and sentiment signals may change faster than the model can adapt, resulting in temporary degradation in forecasting accuracy. In particular, rare events such as the FTX collapse and unexpected macroeconomic announcements introduce distribution shifts that are difficult to capture using historical training data alone. These observations indicate that incorporating online adaptation mechanisms or continual learning strategies could further improve the robustness of the proposed framework.

6. Conclusion and Future Scope

The proposed MA-Helformer demonstrates the potential to improve forecasting performance under the adopted experimental setting and may provide useful decision support for cryptocurrency portfolio management. However, the reported results are based on historical market data and simulated backtesting. Further validation under real-world trading environments is necessary to fully assess the model's robustness, generalizability, and practical applicability. This paper presented MA-Helformer, a systematic extension of the Helformer model, with the objective of addressing five formalised research gaps: (1) univariate input dependency, (2) absent sentiment features, (3) static decomposition parameters, (4) single-step forecasting, and (5) inadequate stress-period evaluation. We propose a model that integrates FinBERT-based daily sentiment encoding, HMM-based market regime detection with learned regime embeddings, regime-conditioned adaptive Holt-Winters decomposition, a 30-dimensional multivariate feature set, and a jointly trained three-horizon forecast head. Experiments on BTC, ETH and SOL (Jan 2017-Dec 2024) show a 49.2% reduction in h=1 RMSE against the Helformer baseline. Transfer learning on ETH and SOL yields consistent 48-51% improvements. The predictive improvements are validated by a Sharpe ratio of 2.34 and a Maximum Drawdown of -4.2% in a realistic trading simulation, which translates the predictive gains to economic value. All reported gains are statistically significant (DM test, $p < 0.01$) and consistent across five random seeds. Future directions include: (1) incorporating inter-asset graph structure via Graph Neural Networks for cross-asset dependency modelling; (2) replacing the rule-based trading strategy with a reinforcement learning execution module; (3) extension to intraday (hourly) forecasting; (4) federated learning for privacy-preserving multi-exchange training; and (5) formal integration of regulatory event calendars and macroeconomic announcement schedules as structured exogenous inputs.

All the Declarations and Statements

Author Contributions Statement

Preeti Pandey – Data preprocessing, formal analysis, methodology, resources, software, visualization, original draft
Geeta Sharma – investigation, project administration, visualization, original draft, review, and editing
Mukesh Kumar – validation, original draft, review, and editing

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement

The authors declare no conflicts of interest.

Funding Declaration

The authors declare that no funding was received for this study.

Data Availability Statement

The implementation of MA-Helformer, together with the data preprocessing scripts, model configuration files, and trained model checkpoints, will be made publicly available through a GitHub repository upon acceptance of the manuscript to facilitate reproducibility and future research.

Ethical Declarations

This study did not involve human participants or animals. Therefore, ethical approval was not required.

Acknowledgments

The authors gratefully acknowledge the valuable comments and constructive suggestions provided by the reviewers and experts. Their professional evaluation helped improve the quality of the manuscript and strengthened the reliability of the experimental findings.

Declaration of Generative AI in Scholarly Writing

The authors acknowledge the use of generative AI tools (OpenAI GPT-4) for language editing of selected passages. Content verification, checking and editing were done by the authors, who take full responsibility for the work.

Abbreviations

The following abbreviations are used in this manuscript:

ATR - Average True Range
ARIMA - AutoRegressive Integrated Moving Average

BTC - Bitcoin
CFTC - Commodity Futures Trading Commission
DM - Diebold-Mariano (test)
ETH - Ethereum
FinBERT - Financial BERT (Bidirectional Encoder Representations from Transformers)
GARCH - Generalized AutoRegressive Conditional Heteroskedasticity
GRU - Gated Recurrent Unit
HMM - Hidden Markov Model
HPO - Hyperparameter Optimization
KGE - Kling-Gupta Efficiency
LSTM - Long Short-Term Memory
MA-Helformer - Multivariate Adaptive Helformer
MACD - Moving Average Convergence Divergence
MAE - Mean Absolute Error
MAPE - Mean Absolute Percentage Error
MDD - Maximum Drawdown
OHLCV - Open-High-Low-Close-Volume
OBV - On-Balance Volume
RMSE - Root Mean Squared Error
RSI - Relative Strength Index
SEC - Securities and Exchange Commission
SOL - Solana
SOPR - Spent Output Profit Ratio
SR - Sharpe Ratio
TFT - Temporal Fusion Transformer
VIX - CBOE Volatility Index
VWAP - Volume-Weighted Average Price

Appendix A/B/C…, with appendix title

Not Applicable

References

[1] Chandankar, P., & Burchard, R., Sensor-Specific Transformer (PatchTST) Ensembles with Test-Matched Augmentation. In Companion of the 2025 ACM International Joint Conference on Pervasive and Ubiquitous Computing (pp. 1055-1060), 2025. DOI: https://doi.org/10.1145/3714394.3756196

[2] Abu Bakar, N., & Rosbi, S., Autoregressive integrated moving average (ARIMA) model for forecasting cryptocurrency exchange rate in high volatility environment: A new insight of bitcoin transaction. International Journal of Advanced Engineering Research and Science, 4(11), 130-137, 2017. DOI: https://dx.doi.org/10.22161/ijaers.4.11.20

[3] Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., & Long, M., Timesnet: Temporal 2d-variation modeling for general time series analysis. arXiv preprint arXiv:2210.02186, 2022. DOI: https://doi.org/10.48550/arXiv.2210.02186

[4] Zeng, A., Chen, M., Zhang, L., & Xu, Q., Are transformers effective for time series forecasting? In Proceedings of the AAAI conference on artificial intelligence (Vol. 37, No. 9, pp. 11121-11128), 2023. DOI: https://doi.org/10.1609/aaai.v37i9.26317

[5] Seabe, P. L., Moutsinga, C. R. B., & Pindza, E., Forecasting cryptocurrency prices using LSTM, GRU, and bi-directional LSTM: a deep learning approach. Fractal and Fractional, 7(2), 203, 2023. DOI: https://doi.org/10.3390/fractalfract7020203

[6] Ashish, V., Noam, S., Niki, P., Jakob, U., Llion, J., Aidan, N. G., ... & Illia, P., Attention is all you need: Advances in neural information processing systems. In NeurIPS Proceedings. 2017, pp. 5998-6008, DOI: https://dl.acm.org/doi/abs/10.5555/3295222.3295349

[7] Kehinde, T. O., Adedokun, O. J., Joseph, A., Kabirat, K. M., Akano, H. A., & Olanrewaju, O. A., Helformer: an attention-based deep learning model for cryptocurrency price forecasting. Journal of Big Data, 12(1), 81. 2025, DOI: https://doi.org/10.1186/s40537-025-01135-4

[8] Poongodi, M., Sharma, A., Vijayakumar, V., Bhardwaj, V., Sharma, A. P., Iqbal, R., & Kumar, R., Prediction of the price of Ethereum blockchain cryptocurrency in an industrial finance system. Computers & Electrical Engineering, 81, 106527, 2020. DOI: https://doi.org/10.1016/j.compeleceng.2019.106527

[9] Girsang, A. S., Hybrid LSTM and GRU for cryptocurrency price forecasting based on social network sentiment analysis using FinBERT. Ieee Access, 11, 120530-120540, 2023. DOI: 10.1109/ACCESS.2023.3324535

[10] Wang, W., Wang, Y., Qi, X., & Chen, H., EIAformer: Empowering transformer with enhanced information acquisition for time series forecasting. Neurocomputing, 131700, 2025. DOI: https://doi.org/10.1016/j.neucom.2025.131700

[11] Hu, Y., Liu, P., Zhu, P., Cheng, D., & Dai, T., Adaptive multi-scale decomposition framework for time series forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 39, No. 16, pp. 17359-17367, 2025, DOI: https://doi.org/10.1609/aaai.v39i16.33908

[12] Kristjanpoller, W., & Minutolo, M. C., A hybrid volatility forecasting framework integrating GARCH, artificial neural network, technical analysis and principal components analysis. Expert Systems with Applications, 109, 1-11, 2018, DOI: 10.1016/j.eswa.2018.05.011
[13] Alonso-Monsalve, S., Suárez-Cetrulo, A. L., Cervantes, A., & Quintana, D., Convolution on neural networks for high-frequency trend prediction of cryptocurrency exchange rates using technical indicators. Expert Systems with Applications, 149, 113250, 2020, DOI: 10.1016/j.eswa.2020.113250
[14] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W., Informer: Beyond efficient transformer for long sequence time-series forecasting. In Proceedings of the AAAI conference on artificial intelligence, Vol. 35, No. 12, pp. 11106-11115, 2021, DOI: https://doi.org/10.1609/aaai.v35i12.17325
[15] Haixu, W., Jiehui, X., Wang, J., & Mingsheng, L., Decomposition transformers with auto-correlation for long-term series forecasting. Autoformer, 2021, DOI: https://dl.acm.org/doi/10.5555/3540261.3541978
[16] Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T., Temporal fusion transformers for interpretable multi-horizon time series forecasting. International journal of forecasting, 37(4), 1748-1764, 2021, DOI: 10.1016/j.ijforecast.2021.03.012
[17] Cavalli, S., & Amoretti, M., CNN-based multivariate data analysis for bitcoin trend prediction. Applied Soft Computing, 101, 107065, 2021, DOI: https://doi.org/10.1016/j.asoc.2020.107065
[18] Araci, D., Finbert: Financial sentiment analysis with pre-trained language models. arXiv preprint arXiv:1908.10063. 2019, DOI: https://doi.org/10.48550/arXiv.1908.10063
[19] Goodell, J. W., Jabeur, S. B., Saâdaoui, F., & Nasir, M. A., Explainable artificial intelligence modeling to forecast bitcoin prices. International Review of Financial Analysis, 88, 102702, 2023, DOI: 10.1016/j.irfa.2023.102702
[20] Golnari, A., Komeili, M. H., & Azizi, Z., Probabilistic deep learning and transfer learning for robust cryptocurrency price prediction. Expert Systems with Applications, 255, 124404. 2024, DOI: 10.1016/j.eswa.2024.124404
[21] Sbrana, A., & Lima de Castro, P. A., N-BEATS perceiver: a novel approach for robust cryptocurrency portfolio forecasting. Computational Economics, 64(2), 1047-1081, 2024, DOI: https://doi.org/10.1007/s10614-023-10470-8
[22] Jin, C., & Li, Y., Cryptocurrency price prediction using frequency decomposition and deep learning. Fractal and Fractional, 7(10), 708, 2023, DOI: https://doi.org/10.3390/fractalfract7100708
[23] Hansun, S., Wicaksana, A., & Khaliq, A. Q., Multivariate cryptocurrency prediction: comparative analysis of three recurrent neural networks approaches. Journal of Big Data, 9(1), 50, 2022, DOI: https://doi.org/10.1186/s40537-022-00601-7
[24] Fallah, M. F., Pourmansouri, R., & Ahmadpour, B., Presenting a new deep learning-based method with the incorporation of error effects to predict certain cryptocurrencies. International Review of Financial Analysis, 95, 103466, 2024, DOI: 10.1016/j.irfa.2024.103466
[25] Zhou, T., Ma, Z., Wen, Q., Wang, X., Sun, L., & Jin, R., Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In International conference on machine learning (pp. 27268-27286). PMLR. , 2022
[26] Das, B., Dagdogen, H. A., Kaya, M. O., & Das, R., A new hybrid approach combining transformer-based language models and graph neural networks for cryptocurrency forecasting. Engineering Applications of Artificial Intelligence, 164, 113179, 2026, DOI: https://doi.org/10.1016/j.engappai.2025.113179
[27] Li, M., Jia, L., Ma, X., & Su, X., Graph diffusion arithmetic feature interaction transformer with temporal information for bitcoin anti-money laundering. Applied Intelligence, 56(6), 192, 2026, DOI: https://doi.org/10.1007/s10489-026-07232-y
[28] Zhang, Y., & Yan, J., Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In The eleventh international conference on learning representations, 2023, Corpus ID: 259298223
[29] B. G. Malkiel, "The efficient market hypothesis and its critics," J. Econ. Perspect., vol. 17, no. 1, pp. 59-82, 2003.

Authors' Profiles



Preeti Pandey is a Research Scholar of the Department of Computer Science and Engineering at Lovely Professional University. She has a degree of M. Tech and pursuing her Ph.D. in Computer Science and Engineering. She combines both academics and practice with more than 12 years of experience in teaching, researching, and being involved in World Bank funded projects in India and other countries. Preeti is also interested in Data Science and Programming Languages as the main research areas. She has written over 30 research publications in some of the leading national and international journals and conferences. She has a reputation of motivating students; she is a blend of theory and practical information to encourage creativity and critical thinking. Her efforts focus on making the gap between academic knowledge.



Dr. Geeta Sharma is an Assistant Professor in School of Computer Applications, Lovely Professional University, Phagwara, Punjab, India. She earned her Ph.D. and Master's degrees in Computer Science from Guru Nanak Dev University, Amritsar. She has over 11 years of combined experience in teaching and research in the fields of Machine Learning, Cloud Computing, IoT, and Network Security. She has authored over 25 research papers in reputed international journals published by Springer, Elsevier, IEEE, and Taylor & Francis, and is an active reviewer for Springer and IEEE. She has also registered six patents. Currently, Dr. Sharma is supervising four Ph.D. scholars and leading advanced research work.



Dr. Mukesh Kumar is an Associate Professor at Advanced Centre of Research & Innovation (ACRI), CGC University Mohali-140307, Punjab, India, and an Adjunct Associate Professor in the Faculty of Law, Solar University Oman. He holds a Ph.D. in Computer Science from Himachal Pradesh University, Shimla, where he specialized in designing ensemble and hybridized data mining techniques for analyzing academic performance. With over 15 years of experience in academia, Dr. Kumar has held various teaching positions, including Assistant Professor at Lovely Professional University and Chitkara University. He has contributed to various international conferences and holds intellectual property rights for several innovations.

How to cite this paper

Preeti Pandey, Geeta Sharma, Mukesh Kumar, "MA-Helformer: A Sentiment-Aware Adaptive Framework for Multi-Horizon Crypto Forecasting", International Journal of Engineering and Manufacturing(IJEM), Vol.16, No.5, pp. 75-91, 2026. DOI:10.5815/ijem.2026.05.04