IJICTR
International Journal of Information & Communication Technology Research

Volume 17- Number 4 – 2025 (27-39)

A Framework for Cryptocurrency Volatility Prediction Based on Cross-Correlation Analysis Using Deep Learning

Masoud Omidvari Abarghouie
Department of Computer Engineering, SR.C.
Islamic Azad University
Tehran, Iran
masoud.omidvari@iran.ir

Sasan H. Alizadeh*
ICT Research Institute (ITRC)
Tehran, Iran
s.alizadeh@itrc.ac.ir

Ahmad Khademzadeh
ICT Research Institute (ITRC)
Tehran, Iran
Zadeh@itrc.ac.ir

Abbas Koochari
Department of Computer Engineering, SR.C.
Islamic Azad University
Tehran, Iran
koochari@iau.ac.ir

Received: 29 January 2025 - Revised: 16 March 2025 - Accepted: 19 July 2025

Abstract— **The popularity of cryptocurrencies has intensified the need for accurate volatility prediction models. This research proposes a novel approach to enhance conditional variance predictions for cryptocurrencies. By leveraging a feature selection technique that selects strong features based on price, return, and volatility cross-correlation analysis, we effectively select the most relevant features for an LSTM-based prediction model. To obtain initial volatility estimates, various GARCH family models (GARCH, EGARCH, and GJR-GARCH) were fitted to the dataset, with the best-fitting model selected based on minimum MSE and RMSE. Subsequently, the proposed MGF-LSTM model was applied to the top eight cryptocurrencies by market capitalization. Experimental results demonstrate that our model significantly reduces prediction errors, providing valuable insights for risk management and investment decision-making in the cryptocurrency market.**

Keyword: cryptocurrency, volatility, cross-correlation, prediction, GARCH, LSTM.

Article type: Research Article

I. Introduction

The rapid expansion of the cryptocurrency market has been accompanied by significant price volatility, posing challenges for investors and traders. Accurate prediction of cryptocurrency price and volatility is crucial for informed decision-making and risk mitigation. Researchers are continuously exploring innovative techniques to improve prediction accuracy within this complex market [1,2]. Moreover, time series analysis has gained significant attention from

* Corresponding Author

researchers, particularly in the domain of identifying fraud activities within e-commerce platforms [3].

Bitcoin, as the first decentralized digital currency, has seen widespread adoption due to various factors. These include legal recognition in several countries, increasing interest from central banks, the launch of Bitcoin futures contracts, and initiatives like Facebook's Libra. The rise in total market capitalization of cryptocurrencies reflects this growing global interest and has sparked extensive academic inquiry into their price dynamics and potential market risks [2].

The growing interest in cryptocurrencies necessitates a deeper understanding of their relationship with traditional finance. One important topic in crypto research is how volatility in cryptocurrency markets relates to traditional financial assets. Understanding this connection helps investors manage risk and assists regulators in maintaining financial stability. The European Central Bank's past concerns about Bitcoin's significant volatility in 2012 highlight the potential for systemic disruptions. Governments also need this knowledge to ensure financial stability and navigate potential policy decisions regarding central bank digital currencies (CBDCs). Existing research primarily focuses on Bitcoin's cross-correlation with a limited number of other investments, highlighting the need for more comprehensive studies examining its broader impact on the financial system [4].

The dynamic nature of the cryptocurrency market and its trading venues necessitates a deeper understanding of its underlying mechanisms. A crucial area of research focuses on the interconnections between major cryptocurrencies and their evolving volatility patterns. By unraveling these relationships, we can gain valuable insights into the market's fairness, stability, and long-term viability. As regulators closely monitor the growth of cryptocurrencies, research in this field provides essential information for informed decision-making. Recent studies [5,6] have suggested potential interconnections between different cryptocurrencies, including evidence of Tether purchases potentially influencing the prices of other cryptocurrencies. These findings underscore the need for further investigation into the complex relationships within cryptocurrency markets.

Existing research [7] has primarily focused on analyzing basic relationships using daily data, neglecting the complex dynamics of volatility interactions over time. However, understanding these dynamics is crucial for various financial applications and risk management strategies. Additionally, knowledge of cryptocurrency correlations aids investors in assessing portfolio risk.[8]

The relative novelty of cryptocurrency markets, coupled with their inherent volatility and influence by unknown factors, presents unique challenges for analysis.1 The online nature of cryptocurrency trading, characterized by 24/7 operation, widespread accessibility, and susceptibility to misinformation, further complicates the landscape.2 Additionally, the complex and diverse nature of cryptocurrency market data, including non-traditional sources like social media sentiment, poses additional challenges for predictive modeling. [9]

Recent studies have demonstrated the potential of news sentiment analysis to predict stock price volatilities. However, these studies have often yielded mixed results, with some suggesting that negative sentiment and a combination of positive and negative sentiment are more predictive than positive sentiment alone. Among various machine learning models, random forest has emerged as a robust choice for both regression and classification tasks in this domain. The disparity between sensitivity and specificity often observed in such analyses highlights the need for a balanced dataset with equal proportions of positive and negative news. Additionally, the comparability of human and machine-based tagging techniques suggests the feasibility of automated, real-time sentiment analysis, which could significantly reduce costs and processing time [10].

Given the challenges outlined above, researchers are exploring innovative techniques for predicting cryptocurrency volatility. Statistical models, known for their efficiency in capturing relationships between multiple cryptocurrencies, are increasingly being explored. Neural networks, while capable of learning from historical data, can be prone to instability during training. Long Short-Term Memory (LSTM) networks, a type of Recurrent Neural Network, address this issue. Combining statistical models with LSTMs offers a powerful approach that leverages the strengths of both techniques: statistical models capture relationships between cryptocurrencies, while LSTMs effectively handle complex patterns in historical data. This synergistic approach has the potential to significantly improve the accuracy of cryptocurrency volatility predictions, as supported by recent research [11].

By analyzing these studies, we aim to gain a comprehensive understanding of the current state of cryptocurrency volatility prediction and its implications for investment strategies. Our objective is to refine the input selection process for a hybrid cryptocurrency volatility prediction model. By focusing on the cross-correlation between price series, returns, and volatility, we aim to identify the most effective input variables. We hypothesize that this selection process will lead to a significant reduction in the mean squared error (MSE) of the cryptocurrency volatility prediction by Multivariate GARCH-Family LSTM (MGF-LSTM) model.

In this paper, we conduct a literature review to explore the strengths and limitations of various approaches to cryptocurrency volatility prediction, identifying promising avenues for future research. Section III presents our proposed model, while Section IV details the results obtained. Finally, the conclusion summarizes our findings and offers valuable suggestions for future researchers.

To ensure a comprehensive understanding of cryptocurrency volatility prediction, the following section reviews previous research in this domain. It focuses on both statistical modeling and the recent adoption of deep learning methods in financial predicting.

II. Literature Review

A. Predicting Cryptocurrency Price and Volatility

Cryptocurrency markets are notoriously volatile, making accurate price prediction a significant challenge. Researchers have explored various techniques to address this issue, with a growing focus on leveraging information from other cryptocurrencies. One study investigated this concept by employing six different methods to create models that combine information from multiple sources. They found that scaled principal component analysis (SPCA) yielded the most accurate predictions for Bitcoin volatility. This improvement was consistent across different tests and even when the researchers made adjustments to their analysis. This suggests that incorporating information from other cryptocurrencies can significantly improve Bitcoin volatility predictions, potentially benefiting investors who can use these predictions to make informed decisions [1].

However, the dynamic nature of cryptocurrency price data poses a significant challenge for prediction. These patterns often exhibit extreme observations, asymmetries, and several nonlinear characteristics that traditional models struggle to capture. To address this complexity, another study developed a new model that can account for long-term trends, asymmetries in volatility, and variations in the frequency of extreme volatility events. Their analysis, applied to over 600 cryptocurrencies, highlights the importance of a reliable filter for cryptocurrency volatility. The results showed that incorporating these variations in extreme observations significantly improved predictions of volatility, the overall price distribution, and specific price points at different future time horizons [2].

Given these challenges in price prediction, machine learning has emerged as a powerful tool for developing cryptocurrency investment strategies. A study employing machine learning to predict returns for over 3700 cryptocurrencies over an eight-year period yielded promising results. Their model achieved a daily return of 7.1% and demonstrated a significant ability to predict future returns based on historical data. This suggests that, unlike traditional currencies heavily influenced by production costs, cryptocurrencies exhibit price behavior that aligns more closely with traditional investments. The study also found that returns for less established cryptocurrencies were generally easier to predict than those for well-established ones, potentially offering additional investment opportunities [12].

Additionally, another study developed two strategies for Ethereum price prediction using real-world data on market indicators and online trends, excluding technical analysis of historical price data. Their models, one using recurrent neural networks (RNNs) and another using support vector machines (SVMs), aimed to predict future prices at various time horizons and overall price trends. Notably, sentiment analysis, which analyzes the emotional tone of online conversations, did not significantly improve the predictive accuracy of these models [9].

It's important to note that no single machine learning method is a one-size-fits-all solution for cryptocurrency volatility prediction. Another research study examined different machine learning methods (NNETAR, CSS, GMDH-NN) to predict volatility in four major cryptocurrencies (Bitcoin, Ethereum, XRP, Tether) between April 2017 and October 2020. They found that different techniques performed best for specific cryptocurrencies (Bitcoin & XRP: CSS, Ethereum: NNETAR, Tether: GMDH-NN). This highlights the need for adaptable prediction models that consider the unique characteristics of each cryptocurrency [13].

Another research investigates the application of deep learning and quantum computing techniques, integrated with EGARCH models, to predict buy/sell signals for cryptocurrencies. The study compares the performance of neural networks and genetic algorithms, specifically highlighting the effectiveness of the Adaptive Genetic Algorithm with Fuzzy Logic. The findings reveal that the X2Y2 cryptocurrency consistently demonstrates the highest prediction accuracy across both methodologies. The integration of EGARCH enhances the predictive capabilities, providing valuable insights for investors, regulators, and developers in the cryptocurrency market [14].

likewise, research investigates the volatility dynamics of financial time series using a variety of GARCH models, including SGARCH, EGARCH, GJRGARCH, and FIGARCH. The study highlights the importance of considering asymmetry, fat-tailedness, and long-memory effects in financial data [15].

B. Deep Learning for Cryptocurrency Prediction

Building upon these advancements in predictive capabilities, researchers are exploring the potential of deep learning, a subfield of machine learning, for cryptocurrency price prediction. A study delves into this area by proposing a Jordan Neural Network model. This model harnesses the power of deep learning to effectively predict future prices for major cryptocurrencies like Bitcoin, Ripple, and Ethereum, outperforming traditional methods in terms of accuracy [16].

Recent research has introduced a novel deep learning model that merges established techniques like ARIMA with Convolutional Neural Networks (CNNs) to capture intricate price patterns. This innovative approach goes beyond traditional methods by incorporating diverse data sources, including exchange data, sentiment analysis, and blockchain data, to provide a holistic view of the market. The model prioritizes robustness and accuracy through hyperparameter optimization, cross-validation, and real-time data integration. Compared to traditional methods like ARIMA, this model demonstrates superior performance, potentially benefiting traders, investors, and decision-makers by offering more comprehensive insights for informed decision-making [17].

Another study has explored the use of chatbots for financial applications. It proposed a novel Persian stock-market chatbot based on the ParsBERT model. A dedicated dataset was set to train the model, and the resulting ChatParse app facilitates multi-turn conversations on stock-market topics [18].

TABLE I. PREVIOUS RESEARCH METHODS

Ref	Method	Data	Category
[2]	HAR	BTC, ETH, XRP, LTC, S&P500	Price
[9]	RNN, SVM	ETH	
[17]	ARIMA, CNN	BTC	
[12]	OLS, XGB	3703 Crypto, 65 Stock	Return
[1]	Multivariate HAR	BTC, ETH, XRP	Volatility
[4]	Asymmetric Diagonal BEKK	BTC, 8 Stock	
[11]	MLP, LSTM, LSTM-GARCH	BTC, ETH, LTC, USDT, EOS, BNB, XLM, TRX	
[13]	NNETAR, GMDH-NN	BTC, ETH, XRP, USDT	
[16]	Jordan Neural Network	BTC, ETH, XRP	
[19]	GARCH, MLP-GARCH	S&P500, CAC40, FTSE100, DAX30, MIB40, TSK60, NK225	
[20]	GARCH, MLP-GARCH	BTC	

Furthermore, deep learning's effectiveness extends to volatility prediction. Researchers compared various deep learning models for predicting volatility in leading cryptocurrencies around the time of the 2020 pandemic. The study found that a deep learning model (LSTM) outperformed traditional GARCH models in predicting volatility. This deep learning model also yielded superior investment returns compared to a stablecoin for long-term investments. Incorporating transaction volume data further enhanced the models' ability to predict risk. Notably, the investment strategy based on these models shifted away from Bitcoin after the pandemic declaration, highlighting the need for even the most advanced models to account for unforeseen global events [11].

This research advances the field of cryptocurrency volatility prediction by demonstrating the superiority of deep learning models over traditional statistical models. Moreover, we highlight the importance of incorporating cross-correlations between cryptocurrencies to improve prediction accuracy. By integrating price, return, and volatility cross-correlation analysis into the model's input design, we achieve enhanced prediction results. Table 1 provides a summary of relevant previous research.

The reviewed literature reveals a growing consensus on the importance of combining multiple data sources and modeling techniques to improve prediction accuracy. In the next section, we introduce a hybrid methodology that leverages these insights to build a robust predicting model.

III. PROPOSE METHOD

Building on the insights from previous studies, we propose a hybrid approach that combines the strengths of traditional econometric models and deep learning. Our framework aims to capture both the temporal dependencies and inter-cryptocurrency relationships in volatility behavior. In this section, we detail the proposed methodology. We begin by selecting the eight cryptocurrencies with the highest market capitalization. For each chosen cryptocurrency, historical closing price data is extracted and partitioned into training and testing sets.

To predict volatility, we first prepared the data through several steps. We calculated log returns from price data to handle volatilities, then tested the weak stationarity of returns using standard statistical tools before applying GARCH models. Then, a Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model is estimated for each cryptocurrency. GARCH models effectively capture the time-varying nature of conditional variance in financial data, which is crucial for accurate volatility prediction.

Subsequently, we examine the cross-correlation among the price, return, and conditional variance series. Features exhibiting strong correlation are selected as potential inputs for the prediction model.

The proposed method for cryptocurrency volatility prediction utilizes a Long Short-Term Memory (LSTM) deep neural network. The LSTM architecture is specifically designed to learn temporal dependencies within data, making it well-suited for tasks involving time series data such as cryptocurrency volatility prediction. The selected features are fed into the LSTM deep neural network, and the cryptocurrency's volatility prediction is designated as the target output from the LSTM. The specific structure of the proposed method is illustrated in Fig.1.

A. *Pre-Processing*

To ensure robust model evaluation in time series prediction, a common practice is to split the data into training and testing sets based on temporal order. A popular approach is the 70/30 split, where 70% of the data is used for training the model and the remaining 30% is reserved for testing its predictive performance. This split has been empirically shown to be effective in time series prediction compared to other splits like 60/40 or 80/20.

B. *Return Series*

To obtain the return series of each cryptocurrency, the Log-Return transformation is used, which follows the formula:

$$r_{i,t} = \log\left(\frac{p_{i,t}}{p_{i,t-1}}\right) \tag{1}$$

where $r_{i,t}$ is return series of each $Crypto_i$ at time t and $p_{i,t}$ is price series of each $Crypto_i$ at time t.

C. *GARCH Modelling*

1) Weak Stationary: To assess weak stationarity in the return series, which is crucial before employing GARCH models for volatility prediction, we utilize the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots. If the squared return series exhibits significant autocorrelation, while the raw returns do not, it suggests the presence of conditional heteroskedasticity, a characteristic of weak stationarity. The autocorrelation coefficient and its significance interval, calculated using the equations above, can be used to confirm the absence of randomness ("whiteness") in the squared returns. This approach is particularly valuable for financial time series, which often exhibit non-stationary behavior. Equations that use for this part is shown in (2,3).

$$\rho_{i,h} = \frac{\sum_{t=h+1}^{N}\left(r_{i,t} - \overline{r_i}\right)\left(r_{i,t-h} - \overline{r_i}\right)}{\sum_{t=1}^{N}\left(r_{i,t} - \overline{r_i}\right)^2} \quad (2)$$

Where $\rho_{i,h}$ is autocorrelation coefficient of each $Crypto_i$ return series at interval time h and N is total number of observations.

$$SE_{i,\rho} = \sqrt{\frac{\left(1 + 2\sum_{i=1}^{h-1}\rho_i{}^2\right)}{N}} \quad (3)$$

Where $SE_{i,\rho}$ is the significance interval of the autocorrelation coefficient for each $Crypto_i$ return series at interval time h with a confidence level of 95%.

2) Conditional Heteroskedasticity Prediction: GARCH models are widely used to predict changing volatility in financial data. They use past price behavior to estimate how volatility might change in the future, making them useful for modeling risks in cryptocurrencies. The core GARCH equation uses past squared returns and volatility estimates to calculate future volatility. The model's order determines how many past terms are considered. To ensure positive volatility, the model imposes conditions on the coefficients. Our approach involves fitting the GARCH model to training data and using the estimated coefficients to predict future volatility. A goodness-of-fit test is then performed to assess the model's adequacy. If the residuals (differences between predicted and actual values) do not resemble random noise, the model's order is adjusted, and the process is repeated until a satisfactory fit is achieved. Equations that use for this part is shown in (4,5).

$$\sigma_{i,t}^2 = \alpha_{i,0} + \sum_{s=1}^{q}\alpha_{i,s} r_{i,t-s}^2 + \sum_{s=1}^{p}\beta_{i,s}\sigma_{i,t-s}^2 \quad (4)$$

Where $r_{i,t}^2$ is squared of each $Crypto_i$ return series at time t and p,q are the order of model that determines the number of lagged terms and $\alpha_{i,0}$ is constant value and α_i is residual coefficient and β_i is residual coefficient of the conditional variance for $Crypto_i$ return series.

$$\begin{cases} p \geq 0, & q > 0 \\ \alpha_{i,0} > 0, & \alpha_i \geq 0 \quad i = 1, \ldots, q. \\ \beta_i \geq 0 & \qquad\quad i = 1, \ldots, p. \end{cases} \quad (5)$$

Above conditions are considered to ensure the positiveness of the conditional heteroskedasticity and to ensure the model remains static, it is crucial that $\alpha_i + \beta_i < 1$.

The Exponential Generalized AutoRegressive Conditional Heteroskedasticity (EGARCH) model is a variant of the GARCH family, specifically designed for modeling conditional volatility. Unlike traditional GARCH models, EGARCH employs the logarithm of conditional variance rather than the variance itself. This characteristic enables the model to capture asymmetric volatility and the differential impacts of positive and negative shocks with greater precision. The EGARCH model's key advantages include the elimination of the positive variance constraint and its ability to model asymmetry. By utilizing the logarithm of conditional variance, the model ensures that the variance remains positive, regardless of the parameter values. Moreover, the model's coefficients allow for the incorporation of asymmetric shock effects. The conditional variance is calculated using the EGARCH model according to equation (6). Where σ_t^2 denotes the conditional variance at time t, ω represents the constant term of the model, β_i are the coefficients associated with lagged conditional variances, α_j are the parameters capturing the influence of absolute shocks, γ_j are the asymmetry parameters that account for the differential impact of positive and negative shocks, and z_t signifies the standardized innovation.

The Glosten-Jagannathan-Runkle Generalized AutoRegressive Conditional Heteroskedasticity (GJRGARCH) model is a specialized extension of the GARCH family that is particularly suited to capturing asymmetric responses to shocks. Unlike standard GARCH models, GJRGARCH explicitly differentiates between the impacts of positive and negative shocks, positing that negative shocks exert a more pronounced influence on conditional variance. This model offers enhanced flexibility in modeling volatility dynamics, especially in the presence of asymmetric effects. The conditional variance in the GJRGARCH model is determined by equation (7), which incorporates parameters that specifically account for the differential effects of positive and negative shocks. where σ_t^2 represents the conditional variance at time t, ω denotes the constant term of the model, β_i signifies the impact of past conditional variances, α_i captures the effect of previous shocks (measured by their absolute values), λ_j represents the asymmetric effect of negative shocks, and $\mathbb{I}\left(\varepsilon_{t-j} < 0\right)$ is an indicator function that assumes a value of 1 if the shock at time $t - j$ is negative and 0 otherwise. Following the fitting of statistical models to each crypto's time series data, a set of predicted conditional volatility series is generated for each model. To construct the final set of predicted conditional volatility series for cryptocurrencies, the model exhibiting the best fit for each cryptocurrency is selected and subsequently included in the output set.

$$\ln(\sigma_t^2) = \omega + \sum_{i=1}^{q}\beta_i \ln(\sigma_{t-1}^2) + \sum_{j=1}^{p}\alpha_j \frac{|z_{t-j}| - \sqrt{\frac{2}{\pi}}}{\sigma_{t-j}} + \sum_{j=1}^{p}\gamma_j z_{t-j} \quad (6)$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{q}\beta_i\sigma_{t-i}^2 + \sum_{j=1}^{p}\alpha_i\varepsilon_{t-j}^2 + \sum_{j=1}^{p}\lambda_j\varepsilon_{t-j}^2\,\mathbb{I}(\varepsilon_{t-j} < 0) \quad (7)$$

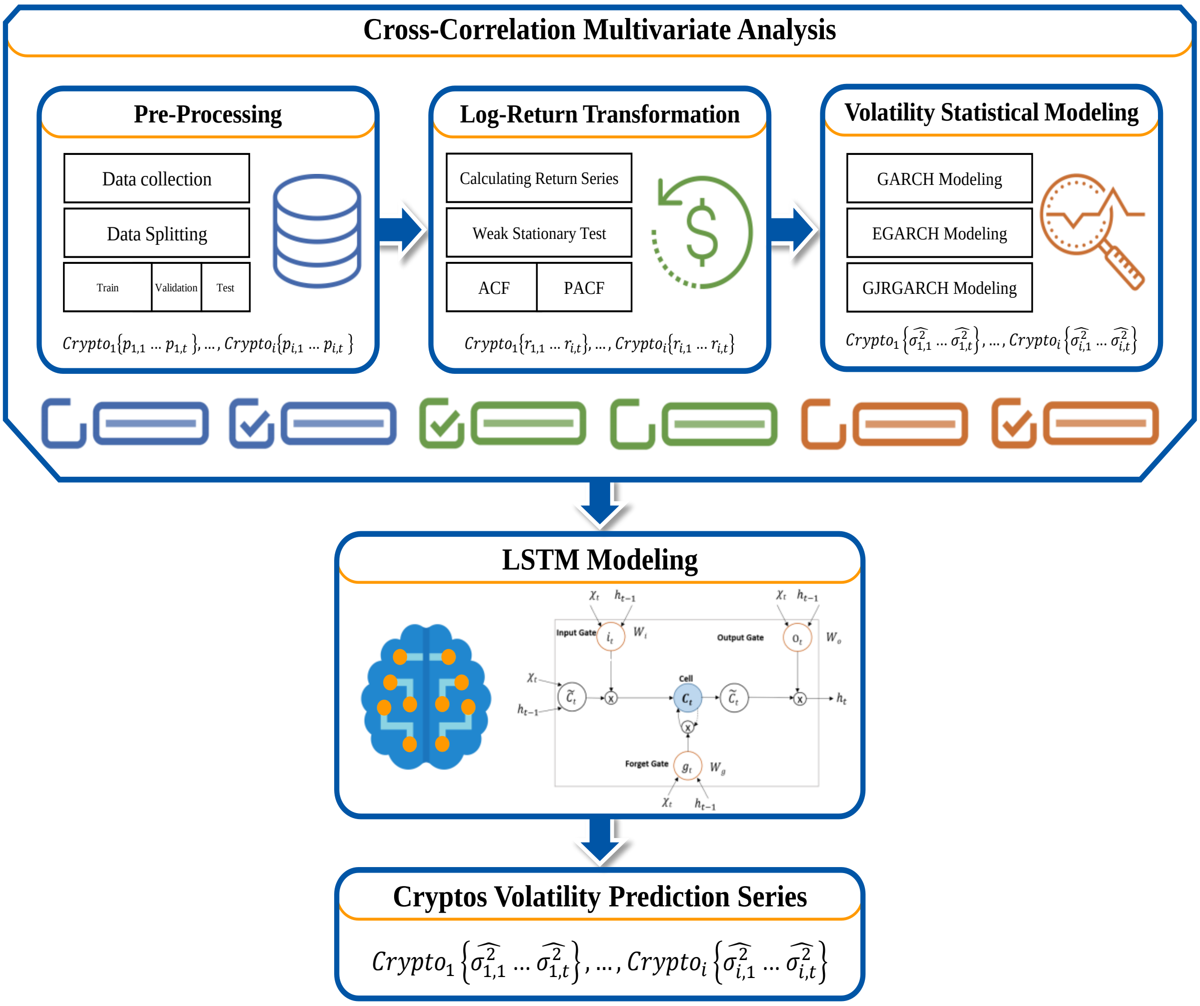

Figure 1. Structure of proposed method

D. Multivariate Analysis with Cross-Correlation:

Research has demonstrated that incorporating cross-correlations between financial factors can enhance prediction accuracy. Our previous work highlighted these interrelationships by examining how the volatility of one financial variable influences the volatility of others, thereby improving future volatility predictions [19]. In this research, we further investigate the cross-correlation between price, return, and volatility series using the Cross-Correlation Function (CCF), as shown in equation (8).

$$CCF(X_i, X_j, k) = \frac{\sum_t \left(X_{i,t} - \bar{X}_i\right)\left(X_{j,t+k} - \bar{X}_j\right)}{\sqrt{\sum_t \left(X_{i,t} - \bar{X}_i\right)^2 \sum_t \left(X_{j,t+k} - \bar{X}_j\right)^2}} \qquad (8)$$

Where $CCF(X_i, X_j, k)$ is cross-correlation between two crypto series at lag k and $X_{i,t}, X_{j,t}$ are the $Crypto_i$ and $Crypto_j$ series at time t and $\bar{X}_i, \bar{X}_j$ are mean of the each crypto series. By examining the degree of cross-correlation between price, return, and volatility series, as calculated using the Cross-Correlation Function (CCF), we categorized these values based on specific criteria outlined in equation (9) and selected the strongly correlated data points. These selected features constitute the input for our prediction model.

$$\begin{cases} CCF(X_i, X_j, k) \le 0.3 & Weak \\ 0.3 < CCF(X_i, X_j, k) < 0.5 & Medium \\ CCF(X_i, X_j, k) \ge 0.5 & Strong \end{cases} \qquad (9)$$

Where $CCF(Xi, Xj, k)$ is cross-correlation between two crypto series at lag k. As it is shown in Fig. 1 features within price, return, and volatility domains were categorized based on the strength of their relationships. Only features demonstrating strong associations were selected for subsequent analysis. These selected features then served as the input for the LSTM model.

E. Volatility Prediction With LSTM:

LSTM networks are specifically designed to address the challenge of processing sequential data by effectively capturing long-term dependencies. They achieve this through specialized gates that regulate the flow of information within the network, enabling it to focus on relevant details for future predictions. Unlike traditional Recurrent Neural Networks (RNNs), LSTMs have a unique architecture where each neuron contains distinct layers that facilitate information filtering. This architecture helps overcome the challenges encountered during RNN training, specifically when calculating derivatives for optimization. Fig. 2 illustrates the structure of one unit LSTM network [11].

Where x_t is input series for i_t as input gate and h_t is output of network from O_t as output gate and C_t is memory gate and forget gate is denoted by g_t at time t. Fig. 2 illustrates the input x_t, the current hidden state h_t, and a value that determines which information is kept or discarded $\overline{C_t}$ within the memory cell at a specific point in time t. The internal operations of an LSTM unit can be mathematically represented by a set of equations (10-15).

$$g_t = \sigma\left(U_g x_t + W_g h_{t-1} + b_f\right) \tag{10}$$

$$i_t = \sigma(U_i x_t + W_i h_{t-1} + b_i) \tag{11}$$

$$\tilde{C}_t = tanh(U_x x_t + W_c h_{t-1} + b_c) \tag{12}$$

$$C_t = g_t * C_{t-1} + i_t * \tilde{C}_t \tag{13}$$

$$O_t = \sigma(U_O x_t + W_O h_{t-1} + b_O) \tag{14}$$

$$h_t = O_t * tanh(C_t) \tag{15}$$

Where U, W is weight matrices, b is bias term, and * is element-wise multiplication.

While we employed Multilayer Perceptron (MLP) for volatility prediction in previous studies [19,20], this research leverages the superior time series modeling capabilities of LSTM. Building upon the insights from prior research, we incorporate the leverage effect and cross-correlation of financial time series as key factors to enhance predictive accuracy. Consequently, the proposed model's inputs include price, return, and conditional heteroskedasticity series exhibiting strong cross-correlations. The LSTM network is tasked with predicting future cryptocurrency volatility. Prediction accuracy is evaluated using Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) metrics, which quantify the deviation between predicted and actual values, as outlined in equations (16) and (17).

$$MSE = \frac{1}{N}\sum_{t=1}^{N}(x_t - h_t)^2 \tag{16}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(x_t - h_t)^2} \tag{17}$$

Where x_t is real observed series and h_t is predicted value at time t.

Table 2. shows the parameters of proposed method.

IV. Results and Discussion

A. Dataset

The dataset used in this study is obtained from CoinMetrics and includes daily closing price data of the top 8 cryptocurrencies by market cap, covering the time period from September 2018 to January 2024. The selected cryptocurrencies include BTC, ETH, USDT, BNB, USDC, XRP, DOGE and ADA [21]. These data were subsequently partitioned into training, validation, and testing sets, comprising 1364, 292, and 293 observations, respectively (Table 3). Fig.3 provides a visual representation of the Bitcoin price series as an illustrative example.

B. Log-Return

To analyze price volatility, log returns were calculated for each cryptocurrency. The resulting return series for all eight cryptocurrencies are visualized in Fig. 4.

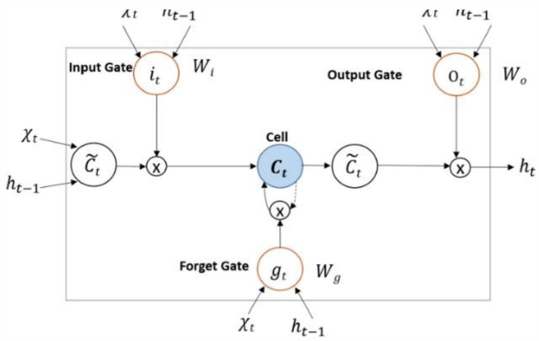

Figure 2. Structure of one unit LSTM

TABLE II. PROPSED METHOD PARAMETERS

Section	Preprocessing			Weak Stationarity		GARCH		EGARCH		GJRGARCH			LSTM				
Parameters	Training set	Validation set	Test set	Lags for ACF	Lags for PACF	p	q	p	q	p	q	o	Input units	Hidden units	Output units	Epochs	Batch size
Value	70%	15%	15%	40	40	1	1	1	1	1	1	1	50	50	1	50	32

TABLE III. DATASET SPLITTING

Set	%	Number of Observation
Training	70	[1364,8]
Validation	15	[292,8]
Test	15	[293,8]

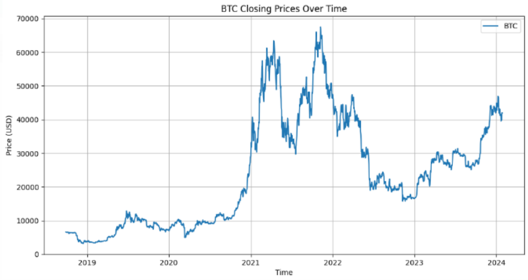

Figure 3. BTC CLOSING PRICE TIME SERIES

C. GARCH Modeling

To prepare for GARCH modeling, the return series of each cryptocurrency were examined for evidence of weak stationarity using autocorrelation function (ACF) and partial autocorrelation function (PACF) plots (Fig. 5 and 6). These analyses suggested potential conditional heteroskedasticity (time-varying volatility) in the squared return series. Further investigation, as presented in Table 4, confirmed this intuition by revealing significant autocorrelation in the squared returns, indicating a departure from a random walk process. These findings provide empirical support for the suitability of GARCH models in capturing the time-varying volatility characteristics inherent in cryptocurrency data. See Table 5 for detailed results.

D. Cross-Correlation Analysis

To identify the most relevant features for the predictive model, the cross-correlation between cryptocurrency price series, returns, and conditional variance was investigated. The results of this analysis are presented in Tables 6, 7, and 8. Based on these findings, a robust set of strongly correlated features was selected through an iterative process and summarized in Table 9.

TABLE IV. ACF AND PACF OUT OF BOUNDS SQUARED RETURNS

Crypto	**ACF Out of Bounds Squared Returns**	**PACF Out of Bounds Squared Returns**
BTC	[1, 2, 4, 5, 7, 11, 18]	[1, 4, 7]
ETH	[1, 2, 3, 4, 5, 6, 7, 10, 11, 16, 18, 25, 29, 35]	[1, 4, 7, 25]
USDT	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40]	[1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 14, 15, 20, 23, 26, 30, 31, 36, 38, 39, 40]
BNB	[1, 2, 3, 4, 5, 7, 9, 10, 11, 14, 18]	[1, 2, 4, 7, 9, 10, 18]
USDC	[1, 2]	[1, 2, 3, 4]
XRP	[1, 2, 5, 23, 30, 32, 33, 37, 38]	[1, 2, 5, 23, 30, 32, 38]
DOGE	[26]	[1, 2, 7, 26]
ADA	[1, 2, 3, 4, 5, 6, 7, 10, 11, 14, 15, 25, 30, 34, 35, 36, 39]	[1, 2, 4, 5, 7, 15]

TABLE V. CRYPTOS GARCH FAMILY MODELLING RESULTS

	MSE			**RMSE**		
	GARCH	EGARCH	GJRGARCH	GARCH	EGARCH	GJRGARCH
BTC	0.001214139	**0.001180083**	0.001220555	0.034844504	**0.03435233**	0.034936447
ETH	**0.001904187**	0.001908472	0.001911343	**0.043636995**	0.043686064	0.043718912
USDT	4.20462E-06	111.0049857	**4.16649E-06**	0.002050517	10.53589036	**0.002041199**
BNB	0.002276851	**0.002254046**	0.002467026	0.047716357	**0.047476795**	0.04966917
USDC	**1.07349E-06**	1.666144837	1.0961E-06	**0.001036094**	1.290792329	0.001046949
XRP	**0.003329679**	0.00530635	0.003411502	**0.057703367**	0.072844695	0.05840806
DOGE	0.054596345	0.069893516	**0.054203263**	0.233658607	0.264373818	**0.232815942**

TABLE VI. CRYPTOS PRICE TIME SERIES CROSS-CORRELATION ANALYSIS

	BTC	**ETH**	**USDT**	**BNB**	**USDC**	**XRP**	**DOGE**	**ADA**	**Mean**
BTC	1	0.924252219	0.20835317	0.865865389	-0.054945999	0.819838274	0.753927427	0.839250481	**0.66956762**
ETH	0.924252219	1	0.183535239	0.953835175	-0.061198562	0.840334838	0.831522943	0.861625389	**0.691738405**
USDT	0.20835317	0.183535239	1	0.182902989	-0.101472838	0.066479446	0.139692694	0.137498949	0.227123706
BNB	0.865865389	0.953835175	0.182902989	1	-0.073566559	0.81548833	0.836255643	0.80904915	**0.673728765**
USDC	-0.054945999	-0.061198562	-0.101472838	-0.073566559	1	-0.032550788	-0.055505887	-0.039091064	0.072708538
XRP	0.819838274	0.840334838	0.066479446	0.81548833	-0.032550788	1	0.873982234	0.861179473	**0.655593976**
DOGE	0.753927427	0.831522943	0.139692694	0.836255643	-0.055505887	0.873982234	1	0.872127297	**0.656500294**

ADA	0.839250481	0.861625389	0.137498949	0.80904915	-0.039091064	0.861179473	0.872127297	1	**0.667704959**
Mean	**0.66956762**	**0.691738405**	0.227123706	**0.673728765**	0.072708538	**0.655593976**	**0.656500294**	**0.667704959**	

TABLE VII. CRYPTOS RETURN SERIES CROSS-CORRELATION ANALYSIS

	BTC	**ETH**	**USDT**	**BNB**	**USDC**	**XRP**	**DOGE**	**ADA**	**Mean**
BTC	1	0.824586448	0.246210713	0.65164289	0.067505725	0.579409116	0.357293219	0.688096486	**0.551843075**
ETH	0.824586448	1	0.236684134	0.679718191	0.052579313	0.627302351	0.341796464	0.748084843	**0.563843968**
USDT	0.246210713	0.236684134	1	0.148629833	0.07180894	0.146631368	0.085239817	0.205056228	0.267532629
BNB	0.65164289	0.679718191	0.148629833	1	0.043280029	0.517185414	0.248286153	0.603095597	0.486479763
USDC	0.067505725	0.052579313	0.07180894	0.043280029	1	0.03304029	0.025146745	0.07121938	0.170572553
XRP	0.579409116	0.627302351	0.146631368	0.517185414	0.03304029	1	0.276205848	0.617765889	0.474692535
DOGE	0.357293219	0.341796464	0.085239817	0.248286153	0.025146745	0.276205848	1	0.333557291	0.333440692
ADA	0.688096486	0.748084843	0.205056228	0.603095597	0.07121938	0.617765889	0.333557291	1	**0.533359464**
Mean	**0.551843075**	**0.563843968**	0.267532629	0.486479763	0.170572553	0.474692535	0.333440692	**0.533359464**	

TABLE VIII. CRYPTOS VARIANCES SERIES CROSS-CORRELATION ANALYSIS

	BTC	**ETH**	**USDT**	**BNB**	**USDC**	**XRP**	**DOGE**	**ADA**	**Mean**
BTC	1	0.779078295	0.250438882	0.624801727	0.200340943	0.32769946	0.300194072	0.69920416	**0.522719692**
ETH	0.779078295	1	0.252089784	0.666738345	0.159466097	0.291715712	0.323262884	0.667448326	**0.51747493**
USDT	0.250438882	0.252089784	1	0.196742793	0.353142272	0.047895059	-0.051319955	0.236057472	0.285630788
BNB	0.624801727	0.666738345	0.196742793	1	0.089105862	0.320771598	0.472134584	0.638919919	**0.501151854**
USDC	0.200340943	0.159466097	0.353142272	0.089105862	1	0.052003926	-0.051087179	0.131283002	0.241781865
XRP	0.32769946	0.291715712	0.047895059	0.320771598	0.052003926	1	0.320590124	0.46758901	0.353533111
DOGE	0.300194072	0.323262884	-0.051319955	0.472134584	-0.051087179	0.320590124	1	0.372341044	0.335764447
ADA	0.69920416	0.667448326	0.236057472	0.638919919	0.131283002	0.46758901	0.372341044	1	**0.526605367**
Mean	**0.522719692**	**0.51747493**	0.285630788	**0.501151854**	0.241781865	0.353533111	0.335764447	**0.526605367**	

TABLE IX. SELECTED FEATURES

Set	**Prices**	**Returns**	**Variances**
Cryptos	BTC, ETH, BNB, XRP, DOGE, ADA	BTC, ETH, ADA	BTC, ETH, BNB, ADA

TABLE X. COMPARISON OF MODEL RESULTS BASED ON RMSE FOR PREDICTING FUTURE CONDITIONAL HETEROSKEDASTICITY

	BTC	**ETH**	**USDT**	**BNB**	**USDC**	**XRP**	**DOGE**	**ADA**
GARCH	0.028306	0.030838	0.000685	0.028273	0.000631	0.062706	0.057045	0.043767
EGARCH	0.028191	0.028586	10.41173	0.026843	1.288028	0.140529	0.043196	0.040788
GJRGARCH	0.028318	0.029443	0.000593	0.028016	0.000606	0.064294	0.055596	0.042465
GARCH-LSTM	0.015734	0.009293	0.007629	0.016724	0.048364	0.055036	**0.007375**	0.025210
SARIMA-GARCH-CNN-BiLSTM [22]	0.022531	0.032275	0.001237	0.037404	0.006549	0.164380	0.026432	0.042754
MGF-LSTM	**0.001108**	**0.001488**	**9.26332E-06**	**0.00328**	**2.57698E-06**	**0.032261**	0.061556	**0.005514**

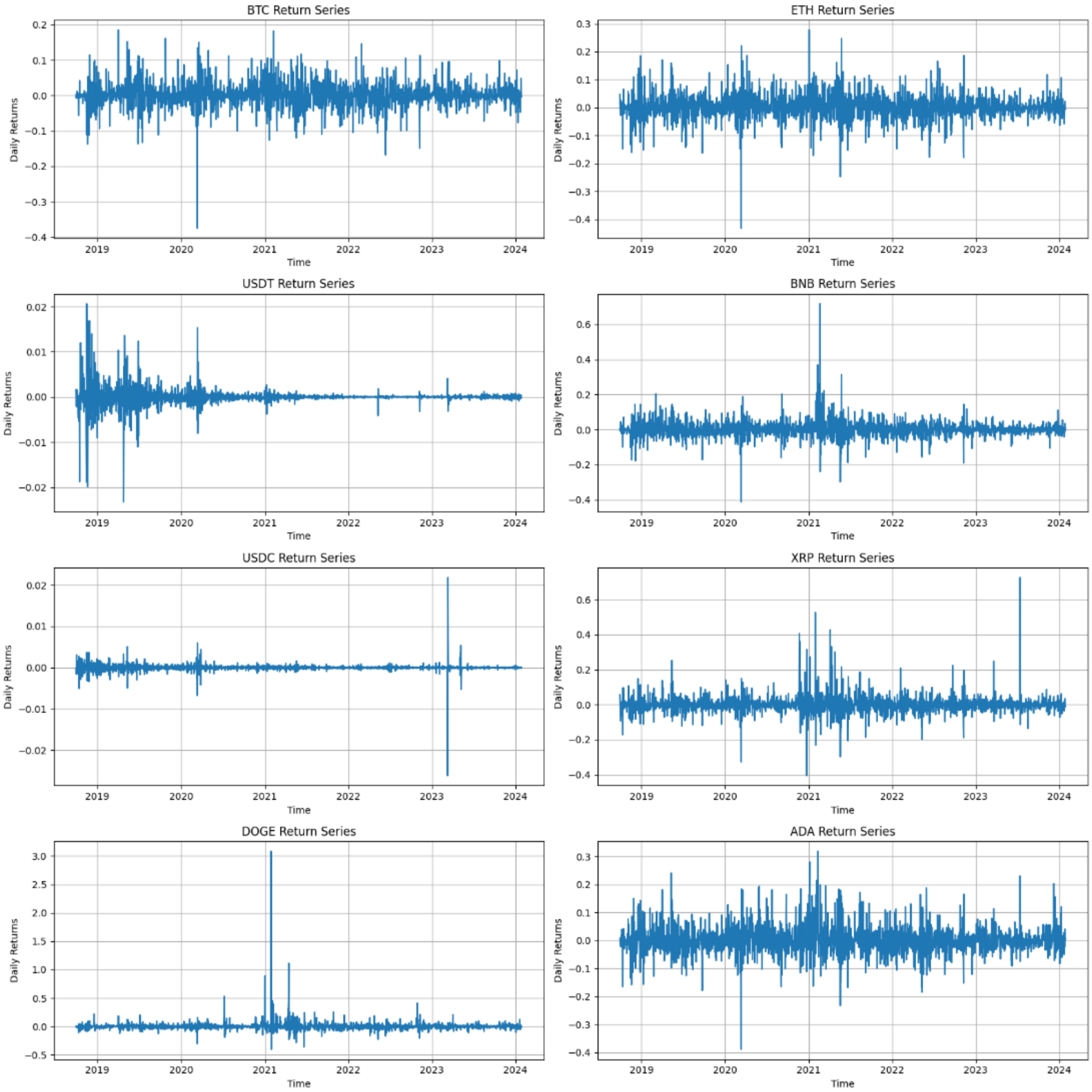

Figure 4. CRYPTOS RETURN SERIES

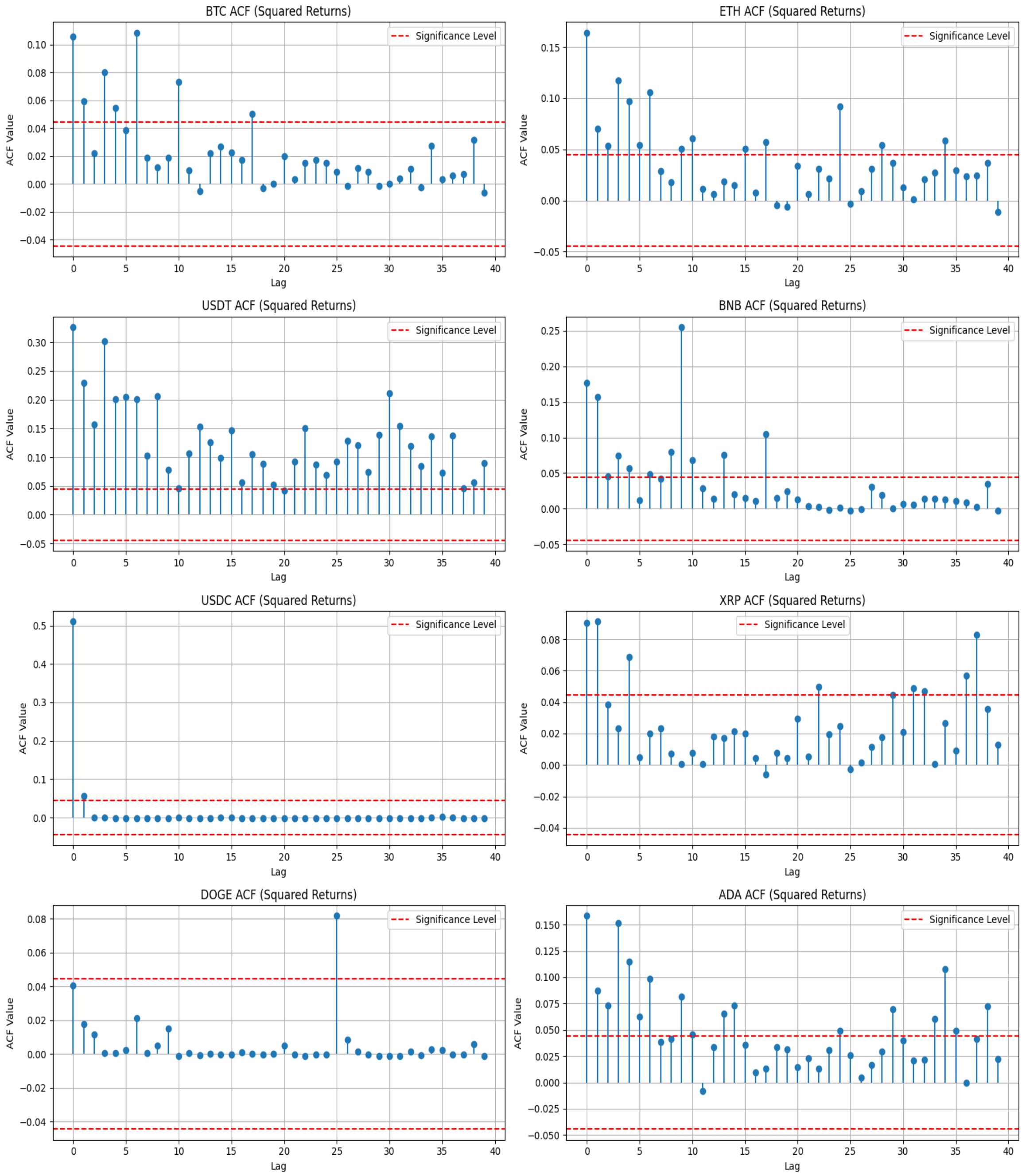

Figure 5. ACF on Cryptos Squared Return Series

E. *LSTM-GARCH Prediction*

After selecting the optimal data for our prediction model, we applied our new method to predict top eight cryptocurrency volatility. Interpretability is essential for applying predictive models in practical financial settings. In our approach, we used a cross-correlation-based feature selection method that allows us to identify key interdependencies between cryptocurrencies. These relationships are not only statistically significant but also economically meaningful, as they help investors understand how volatility in one asset can influence others. Additionally, the model structure ensures that the selected features have a direct contribution to output behavior, improving trust in the prediction process. This transparency helps stakeholders make informed decisions regarding asset allocation and risk management. The accuracy of our predictions was evaluated using common metric: root mean squared error (RMSE). Table 10 compares our results to a well-established GARCH family models. The results demonstrate the superior predictive accuracy of our new method called MGF-LSTM.

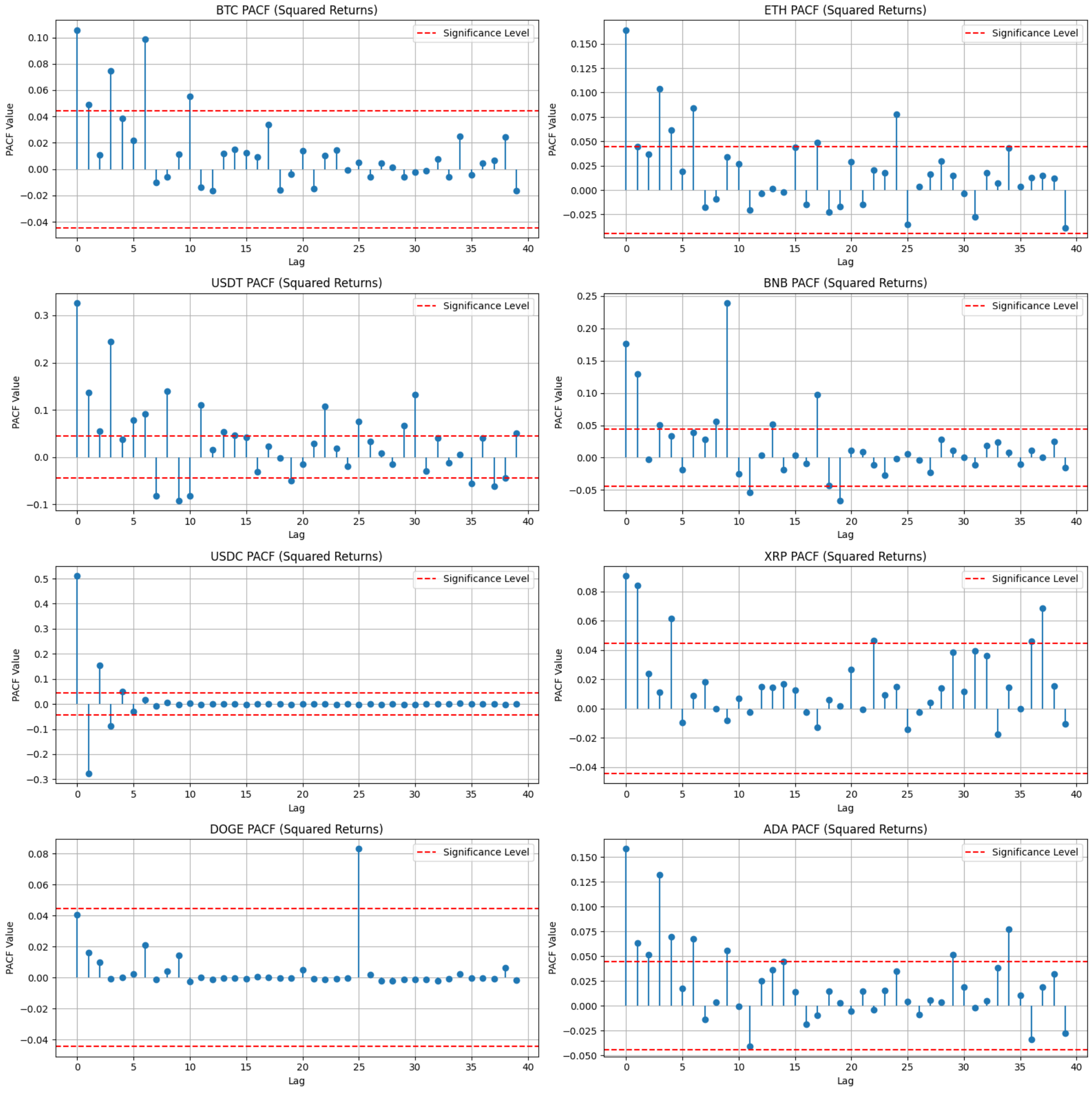

Figure 6. PACF on Cryptos Squared Return Series

V. CONCLUSION

This research presents a novel approach for improved cryptocurrency volatility prediction. We leverage the Long Short-Term Memory (LSTM) method as the predictive model. The model's inputs are carefully selected based on cross-correlation analysis of price series, return, and volatility of the eight leading cryptocurrencies by market capitalization. The results show that carefully selecting input features has a strong impact on improving prediction accuracy. Our model reduced predicting errors and can support better investment strategies and risk decisions. Utilizing data from 2018 to 2024, the research substantially reduces the error of prediction. This research offers valuable insights for future exploration. The proposed method can be adapted to predict the price and volatility of a wider range of cryptocurrencies. Additionally, it holds promise as a tool for selecting stronger member cryptocurrency portfolios.

Despite the promising performance of the proposed MGF-LSTM model, there are several limitations to consider. First, like most deep learning-based approaches, our model is susceptible to overfitting, especially when trained on limited or unbalanced data. Second, the model's performance may degrade under extreme market anomalies or structural shifts, such as regulatory changes, flash crashes, or geopolitical events. Future work could explore ensemble techniques to improve generalization, incorporating external indicators like social sentiment or news feeds, and applying real-time anomaly detection to dynamically adjust model parameters.

REFERENCES

[1] Y. Yi, M. He, and Y. Zhang, "Out-of-sample prediction of Bitcoin realized volatility: Do other cryptocurrencies help?," *The North American Journal of Economics and Finance*, vol. 62, p. 101731, 2022.

[2] L. Catania and S. Grassi, "Forecasting cryptocurrency volatility," *International Journal of Forecasting*, vol. 38, no. 3, pp. 878-894, 2022.

[3] Seyedhossein, L. and M.R. Hashemi, "A timelier credit card fraud detection by mining transaction time series ,"*ITRC*, vol.2, no.3, pp. 21-28, 2010.

[4] E. Bouri, M. Das, R. Gupta, and D. Roubaud, "Spillovers between Bitcoin and other assets during bear and bull markets," *Applied Economics*, vol. 50, no. 55, pp. 5935-5949, 2018.

[5] P. Katsiampa, S. Corbet, and B. Lucey, "Volatility spillover effects in leading cryptocurrencies: A BEKK-MGARCH analysis," *Finance Research Letters*, vol. 29, pp. 68-74, 2019.

[6] J. Griffin and A. Shams, "Is Bitcoin really un-tethered *Journal of Finance*, , vol. 75, pp. 1913-1964, 2018.

[7] J. Fry and E.-T. Cheah, "Negative bubbles and shocks in cryptocurrency markets," *International Review of Financial Analysis*, vol. 47, pp. 343-352, 2016.

[8] P. Katsiampa, S. Corbet, and B. Lucey, "High frequency volatility co-movements in cryptocurrency markets," *Journal of International Financial Markets, Institutions and Money*, vol. 62, pp. 35-52, 2019.

[9] A. Viéitez, M. Santos, and R. Naranjo, "Machine learning Ethereum cryptocurrency prediction and knowledge-based investment strategies," *Knowledge-Based Systems*, p. 112088, 2024.

[10] Ranjbaran, G., et al., Investigating sentiment analysis of news in stock market prediction utilizing machine learning technique. *ITRC*, vol.16, no, pp. 54-60, 2024.

[11] A. García-Medina and E. Aguayo-Moreno, "LSTM–GARCH hybrid model for the prediction of volatility in cryptocurrency portfolios," *Computational Economics*, vol. 63, no. 4, pp. 1511-1542, 2024.

[12] Y. Liu, Z. Li, R. Nekhili, and J. Sultan, "Forecasting cryptocurrency returns with machine learning," *Research in International Business and Finance*, vol. 64, p. 101905, 2023.

[13] F. U. Khan, F. Khan, and P. A. Shaikh, "Forecasting returns volatility of cryptocurrency by applying various deep learning algorithms,"*Future Business Journal*, vol. 9, no. 1, p. 25, 2023.

[14] Alaminos, D., M.B. Salas-Compás, and Á. Callejón-Gil, "Managing extreme cryptocurrency volatility in algorithmic trading: EGARCH via genetic algorithms and neural networks," *Quantitative Finace and Economics*, vol.8, no.1, pp.153-209, 2024.

[15] Ampountolas, A., "Enhancing forecasting accuracy in commodity and financial markets: Insights from garch and SVR models." *International Journal of Financial Studies*, vol.12, no**.**3, p. 59, 2024.

[16] V. D'Amato, S. Levantesi, and G. Piscopo, "Deep learning in predicting cryptocurrency volatility," *Physica A: Statistical Mechanics and its Applications*, vol. 596, p. 127158, 2022.

[17] J. Sabeena and P. L. Sagar, "Enhancing Predictive Accuracy for Real Time Cryptocurrency Market Prices with Machine Learning Techniques," in *Conf. Sustainable Communication Networks and Application (ICSCNA)*, 2023: IEEE, 2023, pp. 1288-1292.

[18] Shahedi, A. and S. Seyedin, "ChatParse: A New Application for Persian Bourse Chatbot." *ITRC*, vol.15, no.2, pp. 49-58.2023.

[19] M. Omidvari Abarghouie, S. H. Alizadeh, and A. Khademzadeh, "Cross Relation Modeling for Financial Time Series Conditional Heteroskedasticity Using MLP Neural Network and Genetic Algorithm," in *Conf. Web Research (ICWR2023)*, Tehran, 2023.

[20] M. Omidvari Abarghouie, S. H. Alizadeh, and A. Khademzadeh, "A study of the leverage effect in Bitcoin Conditional Heteroskedasticity prediction," *in Conf. Artificial Intelligence and Robotics (QICAR)*, IEEE, 2024, pp. 180-185.

[21] https://charts.coinmetrics.io/crypto-data

[22] H. Araya, J. Aduda, and T. Berhane, "A Hybrid GARCH and Deep Learning Method for Volatility Prediction," *Journal of Applied Mathematics*, 2024.

Masoud Omidvari Abarghouie received his B.Sc. in Computer Software Engineering from Islamic Azad University, Shiraz, Iran. Also, he received his M.Sc. degree in Artificial Intelligence from the Science and Research Branch University, Qazvin, Iran. Now, he is a Ph.D. candidate in Artificial Intelligence and Robotics at the Science and Research Branch University, Tehran, Iran. His interests lie in the fields of Artificial Intelligence, Econometrics, Volatility Prediction and Cryptocurrency.

Sasan H. Alizadeh received his B.Sc. in Computer Hardware Engineering from Shiraz University, Shiraz, Iran. Also, he received his M.Sc. and the Ph.D. degrees in Computer Science from Amirkabir University of Technology, Tehran, Iran. He is a faculty member of the Department of Information Technology at ICT Research Institute (Iran Telecommunication Research Center), Tehran, Iran. His research interests include Statistical Machine Learning, Stochastic Processes, Bayesian Networks, Software Engineering, Social Network Analysis and Recommender Systems.

Ahmad Khademzadeh was born in Mashhad, Iran, in 1943. He received his B.Sc. degree in Applied Physics from Ferdowsi University, Mashhad, Iran, in 1969 and the M.Sc., Ph.D. degrees respectively in Digital Communication and Information Theory and Error Control Coding from the University of Kent, Canterbury, U.K. He is currently the Head of Education and National Scientific and Informational Scientific Cooperation Department at Iran Telecom Research Center (ITRC). He was the head of Test Engineering Group and the director of Computer and Communication Department at ITRC. He is also a lecturer at Tehran Universities and he is a committee member of Iranian Computer society and also a committee member of the Iranian Electrical Engineering Conference Permanent Committee. Dr. Khademzadeh has been received four distinguished national and international awards including Kharazmi International Award, and has been selected as the National outstanding researcher of the Iran Ministry of Information and Communication Technology.

Abbas Koochari received his PhD in Computer Science with a focus on Artificial Intelligence in 2012. He is currently an Assistant Professor and a faculty member at the Science and Research Branch of Azad University. His research interests include image processing, machine vision, speech processing, natural languages, and deep learning.