electronics

MDPI

Article

A Hybrid Frequency Decomposition–CNN–Transformer Model for Predicting Dynamic Cryptocurrency Correlations

Ji-Won Kang [1], Daihyun Kwon [2] and Sun-Yong Choi [1,*]

[1] Department of Finance and Big Data, Gachon University, Seongnam 13120, Gyeonggi, Republic of Korea; kangjiwon5961@gachon.ac.kr
[2] Quantum Intelligence Corp., 31F, One IFC, 10 Gukjegeumyung-ro, Yeongdeungpo-gu, Seoul 07326, Republic of Korea; daihyun.kwon@qic.ai
* Correspondence: sunyongchoi@gachon.ac.kr; Tel.: +82-31-750-5387

Abstract

This study proposes a hybrid model that integrates Wavelet frequency decomposition, convolutional neural networks (CNNs), and Transformers to predict correlation structures among eight major cryptocurrencies. The Wavelet module decomposes asset time series into short-, medium-, and long-term components, enabling multi-scale trend analysis. CNNs capture localized correlation patterns across frequency bands, while the Transformer models long-term temporal dependencies and global relationships. Ablation studies with three baselines (Wavelet–CNN, Wavelet–Transformer, and CNN–Transformer) confirm that the proposed Wavelet–CNN–Transformer (WCT) consistently outperforms all alternatives across regression metrics (MSE, MAE, RMSE) and matrix similarity measures (Cosine Similarity and Frobenius Norm). The performance gap with the Wavelet–Transformer highlights CNN's critical role in processing frequency-decomposed features, and WCT demonstrates stable accuracy even during periods of high market volatility. By improving correlation forecasts, the model enhances portfolio diversification and enables more effective risk-hedging strategies than volatility-based approaches. Moreover, it is capable of capturing the impact of major events such as policy announcements, geopolitical conflicts, and corporate earnings releases on market networks. This capability provides a powerful framework for monitoring structural transformations that are often overlooked by traditional price prediction models.

Keywords: Wavelet frequency decomposition; convolutional neural networks; Transformers; correlation structure; cryptocurrencies

check for updates

Academic Editor: Arkaitz Zubiaga

Received: 1 September 2025
Revised: 10 October 2025
Accepted: 17 October 2025
Published: 22 October 2025

Citation: Kang, J.-W.; Kwon, D.; Choi, S.-Y. A Hybrid Frequency Decomposition–CNN–Transformer Model for Predicting Dynamic Cryptocurrency Correlations. *Electronics* **2025**, *14*, 4136. https://doi.org/10.3390/electronics14214136

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

1. Introduction

The cryptocurrency market has matured into a viable alternative asset class with the approval of spot Bitcoin ETFs by the U.S. Securities and Exchange Commission in early 2024 [1]. However, cryptocurrencies exhibit substantially higher volatility than traditional assets and display time-varying correlation structures that shift rapidly in response to market events [2–4]. These dynamic correlations intensify during market stress, precisely when diversification is most needed, creating critical challenges for portfolio risk management.

Traditionally, statistical models such as Vector Autoregression (VAR) [5] and Vector Exponential Smoothing (VES) [6] have been widely used for multivariate financial time series analysis, but they struggle to capture nonlinear dynamics and rapidly changing correlation structures, challenges that are amplified in the highly volatile cryptocurrency markets.

Following recent successes of machine learning techniques in various fields including computer vision, natural language processing, and drug discovery, researchers have been exploring machine learning applications in time series analysis [7,8].

Leveraging the Transformer architecture [9], which was designed to overcome recurrent neural network limitations through self-attention mechanisms enabling long-term dependency modeling and parallel processing, researchers have developed various Transformer-based architectures for time series analysis [10–13]. Yet, as demonstrated in [14], the variants of Transformer models specifically designed for long-term time series forecasting (LTSF) may not be ideal.

More sophisticated hybrid architectures combining multiple neural components have emerged as effective approaches for complex financial forecasting. Ref. [15] integrated variational LSTM with vine copula structures to model cross-market dependencies, while others have combined convolutional neural networks (CNNs) with long short-term memory networks (LSTMs) for financial prediction [16–18], an approach that has been applied across multiple domains. However, LSTMs suffer from limited parallelization due to their sequential structure, and their ability to capture long-term dependencies deteriorates as sequence length increases. Moreover, CNN-LSTM architectures are limited in their ability to explicitly represent complex structural interactions among multiple assets [19].

Despite the importance of correlation structures for risk management and portfolio optimization [20,21], existing machine learning approaches focus primarily on price prediction rather than correlation dynamics.

Building upon these prior studies, we propose a hybrid architecture integrating three components: (1) Wavelet-based frequency decomposition to separate multi-scale patterns (short-, medium-, and long-term), (2) CNNs to extract spatial structure from correlation matrices at each frequency band, and (3) Transformers to model temporal evolution and long-range dependencies. This approach addresses the challenge of predicting time-varying correlations in cryptocurrency markets by explicitly modeling multi-scale structural relationships.

This study makes the following contributions. First, it extends the existing literature on correlation forecasting by incorporating machine learning methodologies. Since correlation prediction inherently requires multivariate forecasting, we depart from traditional econometric approaches such as GARCH (Generalized Autoregressive Conditional Heteroskedasticity) [22] and employ machine learning techniques instead. Second, we propose a novel Transformer-based hybrid model as the core methodological contribution. This architecture is expected to be applicable to multivariate time series forecasting tasks across various domains beyond the scope of this study. Third, unlike prior studies that used frequency bands mainly as separate features, this study inputs inter-asset correlation matrices for each band directly into the model, enabling it to learn structural relationships on short-, medium-, and long-term scales. Following Barbierato et al. [23], this study applies machine learning methodologies to financial time series, specifically focusing on cryptocurrency correlation forecasting.

The remaining paper is organized as follows. The following section reviews prior studies on Transformer architectures and multivariate time series forecasting. Section 3 provides a brief description of the cryptocurrency data used in this study. In Section 4, we review the foundational models underlying our hybrid approach—frequency decomposition, CNN, and Transformer—and present the proposed methodology. Section 5 reports the empirical results of forecasting cryptocurrency correlations. Finally, Section 6 offers a summary of the study and concluding remarks.

2. Literature Review

This section provides a brief review of prior studies on Transformer models and the existing literature on multivariate time series forecasting, which form the methodological foundation of this study.

2.1. Multivariate Time Series Forecasting

Traditionally, multi-dimensional time series prediction has relied on statistical methods such as Vector Autoregression (VAR) [5], Vector Exponential Smoothing (VES) [6], and multivariate Autoregressive Moving Average (ARIMA) [24,25]. These approaches explicitly model the interdependencies among multiple variables; however, their linear specifications limit their ability to capture complex nonlinear dynamics.

In recent years, the growing availability of large-scale datasets and advances in computational methods have driven increased interest in applying machine learning techniques to multivariate time series forecasting. Machine learning models offer the capability to capture nonlinear patterns and handle high-dimensional data, making them well-suited for scenarios where traditional models may underperform.

2.2. Transformer Models

As noted by Vaswani et al. [9], the dominant architectures in natural language processing at the time—recurrent neural networks and long short-term memory networks (and, in computer vision, convolutional neural networks)—suffered from notable computational limitations. Specifically, RNN-based models were constrained by their inherently sequential nature, preventing effective parallelization, while CNN-based models required increased computational complexity to capture long-range dependencies. The Transformer architecture [9] addressed these issues by replacing sequential computations with a self-attention mechanism. This innovation yielded the first transduction model to rely solely on attention for long-range dependency modeling, achieving $\mathcal{O}(1)$ sequential operations and $\mathcal{O}(1)$ path length, compared with $\mathcal{O}(n)$ and $\mathcal{O}(\log_k(n))$ in RNN and CNN, respectively. The Transformer demonstrated superior performance on machine translation tasks while requiring substantially fewer computational resources.

Time series data share key structural properties with natural language data, i.e., sequential ordering of observations and the presence of long-range dependencies, where historical patterns or events influence future values. These similarities have naturally motivated the adoption of Transformers for time series analysis [10–13,26–28]. However, despite their success in capturing long-term dependencies in natural language processing, vanilla Transformers face several challenges when applied to time series data. These include computational complexity when processing long sequences [10,13,27], a lack of locality and hierarchical modeling due to the equal treatment of all positions in the standard attention mechanism [10,27], inadequate periodic and trend modeling because of the absence of structural biases for such patterns [11,13,26], and insufficient mechanisms for explicitly capturing multivariate dependencies [28].

To address these limitations, various Transformer variants have been proposed. Li et al. [10] introduced a log-sparse attention mechanism, reducing complexity from $\mathcal{O}(n^2)$ to $\mathcal{O}(n \log n)$ through fixed sparse attention patterns, enabling more efficient modeling of long sequences. However, fixed sparsity may overlook certain relevant dependencies. Zhou et al. [12] proposed the ProbSparse self-attention mechanism, which also achieves $\mathcal{O}(n \log n)$ complexity, combined with a distilling operation for handling long inputs and a generative decoder for producing long outputs in a single pass. Yet, the probabilistic selection of attention weights can lead to loss of temporal information. Wu et al. [11] replaced self-attention with an auto-correlation mechanism, enabling discovery of period-

based dependencies and aggregation at the sub-series level, while incorporating series decomposition to extract trend and seasonal components; this process, however, can be computationally expensive for long sequences.

Similarly, Liu et al. [27] developed pyramidal attention with inter- and intra-scale connections to achieve $\mathcal{O}(n)$ complexity and capture multi-scale temporal dependencies through a hierarchical structure, though designing the optimal pyramid configuration remains challenging. Chen et al. [26] proposed a quaternion-based architecture with learning-to-rotate attention, encoding learnable period and phase information for complex periodic pattern modeling, albeit with potentially high computational costs. Zhou et al. [13] leveraged frequency-enhanced blocks operating in the frequency domain via Fourier transforms to capture global correlations, combining seasonal trend decomposition with frequency domain processing for improved performance and linear complexity, though at the cost of some temporal locality.

More recently, Zeng et al. [14] conducted a comprehensive evaluation of Transformer variants designed for long-term time series forecasting (LTSF), such as Informer, Autoformer, and Fedformer, and concluded that these architectures are fundamentally unsuitable for LTSF tasks. Their findings showed that a set of simple linear models (LTSF-Linear) consistently outperformed Transformer variants across nine benchmark datasets, attributing this to the inherent permutation-invariant tendencies of attention mechanisms, which result in the loss of temporal information.

2.3. Hybrid Models

Lu et al. [17] proposed a hybrid CNN-LSTM architecture that integrates convolutional neural networks for feature extraction from historical stock data with long short-term memory (LSTM) networks for temporal sequence modeling. Using data from the Shanghai Composite Index, the model forecasted next-day closing prices and outperformed several baselines, including MLP, CNN, RNN, LSTM, and CNN–RNN architectures. However, the model was trained on a limited set of eight indicators and employed a fixed 10-day window size, which may not be optimal for all forecasting horizons.

Widiputra et al. [16] addressed the problem of simultaneously forecasting multiple financial market indices by proposing a hybrid deep learning model that combines convolutional neural networks (CNNs) with long short-term memory networks (LSTMs). Their multivariate CNN–LSTM model takes multiple financial time series as input, extracts local patterns through CNN layers, and captures temporal dependencies using stacked LSTMs. The model achieved lower RMSE scores than standalone CNN or LSTM architectures, demonstrating its effectiveness in capturing cross-market interactions and improving predictive accuracy in multi-output forecasting tasks.

Kim and Park [29] introduced a hybrid CNN–Transformer architecture that employs dilated causal convolutional neural networks (DCCNNs) to extract temporal features from each variable, combined with Transformer attention mechanisms to capture inter-series dependencies. The approach was evaluated on four benchmark datasets—Traffic, Exchange Rate, Electricity, and Solar Energy—and demonstrated the ability to simultaneously learn temporal patterns and correlations across variables. Nevertheless, the model's effectiveness was most pronounced in datasets with clear periodicity and strong inter-variable correlations. As discussed above, this study builds on the widely adopted Transformer architecture to propose a novel hybrid model. In addition, by forecasting the dynamic correlations among cryptocurrencies, it contributes to the broader literature on multivariate time series forecasting.

3. Data Description

In this study, we collected historical price data for eight major cryptocurrencies—Bitcoin, Ethereum, Binance Coin, Neo, Litecoin, Qtum, Cardano, and XRP—covering the period from 1

January 2018, to 1 January 2025. The data collection was conducted in a Python 3 environment using the public API provided by the Binance exchange (https://github.com/ccxt/ccxt (accessed on 28 May 2025)). Among all available trading pairs, only those with USDT as the quote currency were selected to construct a list of major assets traded against USDT. From this list, we chose eight representative cryptocurrencies based on trading volume and market relevance. The data was retrieved on a daily basis (1-day OHLCV). Due to Binance API's limitations on the number of data points per request and time interval constraints, we implemented a loop-based approach that iteratively moved the time window to cover the entire target period. To comply with the API rate limits, a delay interval of 0.5 s was introduced between requests. A total of 2559 data points were collected. The retrieved daily price data was organized into a Pandas DataFrame and saved in CSV format.

Table 1 presents the descriptive statistics of the log returns for each cryptocurrency, while Table 1 visualizes the daily log return fluctuations over time.

Table 1. Log return summary statistics by cryptocurrency.

Asset	Mean	Std.Dev.	Min	Max	Skewness	Kurtosis
BTC	0.0009	0.0348	−0.5026	0.1784	−1.2519	20.4016
ETH	0.0006	0.0455	−0.5905	0.2338	−1.1268	15.1670
BNB	0.0016	0.0474	−0.5823	0.5324	−0.3706	21.6922
NEO	−0.0007	0.0562	−0.5024	0.3690	−0.3738	8.4099
LTC	−0.0002	0.0487	−0.4867	0.2635	−0.7146	9.9519
QTUM	−0.0008	0.0591	−0.6260	0.4043	−0.4986	11.5918
ADA	0.0004	0.0524	−0.5331	0.2864	−0.2299	7.0560
XRP	0.0004	0.0538	−0.5387	0.5487	0.5689	18.4507

Starting with Table 1, the mean returns of most cryptocurrencies lie within a narrow range between −0.001 and +0.001, indicating relatively small average daily returns. Among them, BNB stands out with a mean return of 0.0016, suggesting higher profitability compared to other assets. In contrast, NEO, LTC, and QTUM exhibit negative mean returns, implying potential long-term losses.

The standard deviation (Std), which represents the volatility of returns, generally ranges between 0.03 and 0.06 for most assets. This reflects a relatively high level of daily risk, especially when compared to traditional financial assets. In particular, QTUM and NEO show relatively higher standard deviations, indicating greater price fluctuations.

The minimum and maximum values represent the extreme ends of the return distribution. While the minimum returns are fairly consistent across assets, hovering around −0.5, the maximum returns vary significantly. BTC's maximum return is 0.1784, indicating a more limited upside, whereas BNB and XRP reach exceptionally high maximums of 0.5324 and 0.5487, respectively, suggesting significant upward spikes at certain points.

Skewness measures the asymmetry of the return distribution. Most assets exhibit negative skewness, meaning they are more prone to large downward movements. However, XRP is the only asset with positive skewness (0.5689), indicating a distribution with a heavier right tail. This suggests that XRP experienced more frequent or intense upward movements compared to the others.

Kurtosis indicates the "peakedness" of the return distribution and the frequency of extreme values. All assets show kurtosis values significantly exceeding the normal distribution benchmark of 3, with BNB (21.6922), BTC (20.4016), and XRP (18.4507) displaying particularly high values. This suggests that extreme return events occurred more frequently for these assets, implying elevated risk.

Looking at Figure 1, a common sharp decline can be observed across most cryptocurrencies during the first half of 2020, which can be attributed to the global financial shock

caused by the COVID-19 pandemic. Furthermore, consistent with the descriptive statistics, BNB and XRP show noticeable spikes at certain time points, with XRP exhibiting more frequent surges, highlighting its tendency toward sudden upward movements.

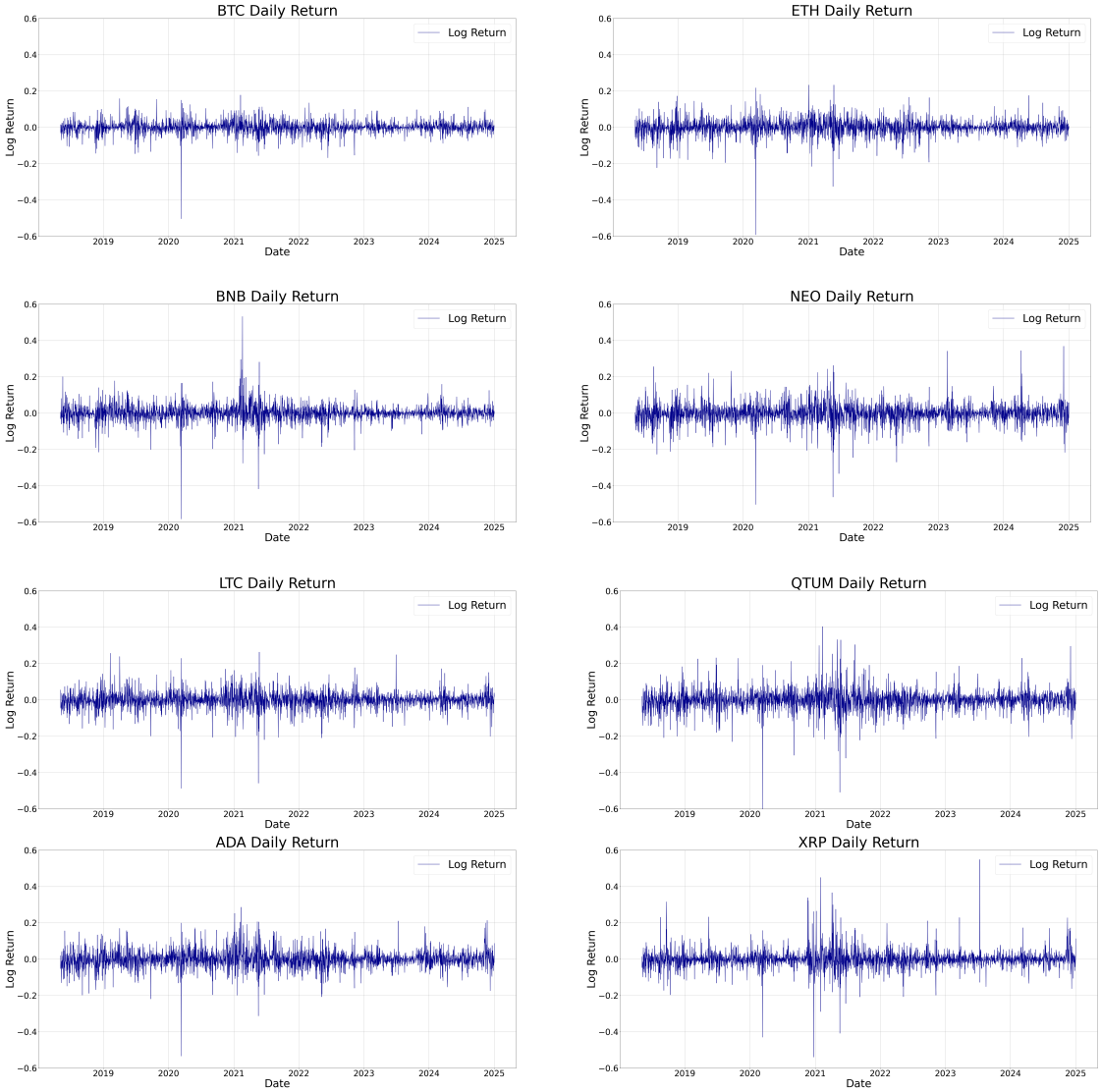


Figure 1. Time series of daily log returns for each cryptocurrency.

4. Methods

In this section, we present the proposed model, which integrates frequency decomposition, convolutional neural networks (CNNs), and Transformer architectures. Furthermore, we provide a brief review of the foundational components, frequency decomposition, CNN, and Transformer, that form the basis of our proposed framework. The implementation of our proposed model is publicly available at our GitHub repository (https://github.com/EarthWon/Wavelet-Frequency-Decomposed-CNN-Transformer.git) (accessed on 16 October 2025).

4.1. Frequency Decomposition

Frequency decomposition is a signal processing technique derived from the concept of frequency transform, which represents time or spatial input signals as a combination of multiple frequency components. This approach is one of the most fundamental methods for analyzing and understanding signals, based on the assumption that any input signal can be

decomposed into a sum of periodic functions with different frequencies. This perspective was first formalized in Fourier's analytical solution to the heat equation, which later led to the development of the Fourier Transform.

While the Fourier transform extracts the global frequency components of a signal, it has the inherent limitation of losing temporal information. To overcome this issue, the Short-Time Fourier transform (STFT) was proposed Allen [30], which analyzes the frequency content of signals over localized time intervals. However, STFT suffers from the trade-off between time and frequency resolution due to its fixed window size. To address these limitations, the Wavelet transform was introduced Mallat [31]. Wavelet transform is designed to consider both time and frequency information simultaneously, enabling the decomposition of signals into multiple scales. This allows for multiresolution analysis (MRA), where low-frequency components capture long-term trends while high-frequency components reveal short-term fluctuations. Wavelet frequency decomposition builds upon this core idea of the wavelet transform by decomposing time series data into various frequency bands for detailed analysis. This technique has been widely used in fields such as image and speech signal processing, time series analysis and anomaly detection, and more recently, in financial time series analysis [32–37].

In this study, we apply Wavelet decomposition to learn the correlation structure of financial markets across different frequency bands. In the low-frequency domain, the model captures the long-term co-movement patterns among assets, while in the high-frequency domain, it detects short-term abrupt changes in correlations. By simultaneously learning correlation structures across multiple time scales, the model can effectively capture a multi-scale view of market dependencies, allowing for a more comprehensive understanding and prediction of complex market dynamics, from long-term trends to short-term market shocks. Figure 2 provides an example of frequency decomposition applied to our dataset.

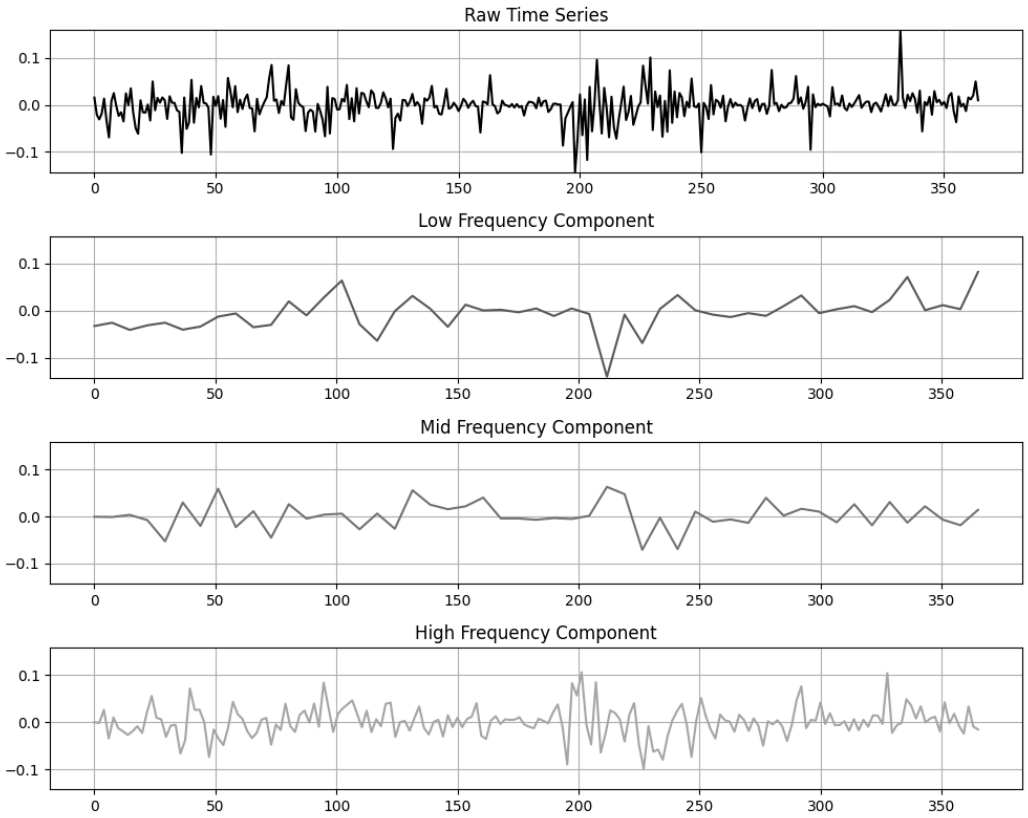


Figure 2. Example of Wavelet decomposition.

4.2. Convolutional Neural Networks

CNNs, originally proposed by [38] as a model inspired by the human visual cortex, have become a foundational architecture in deep learning, particularly excelling in tasks involving visual data such as image classification [39]. A typical CNN consists of an input layer, hidden layers, and an output layer. Figure 3 shows a typical CNN architecture.

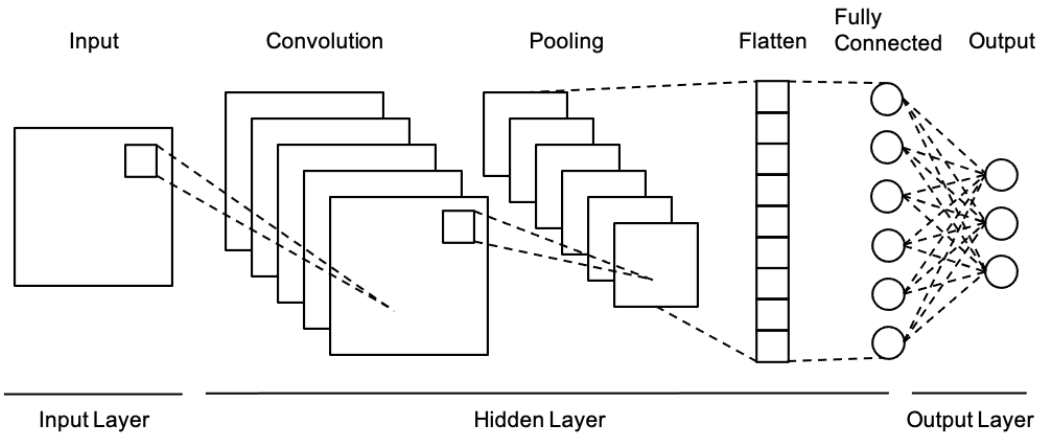


Figure 3. Standard CNN architecture.

The input layer receives image data, which is typically three-dimensional, composed of height, width, and depth. The depth, often referred to as the number of channels, corresponds to the Red, Green, and Blue (R, G, B) components in the case of color images. The output layer generates the final prediction by applying previously learned weights and mapping the resulting features to class scores. These scores represent a probability over the possible class labels, and the class with the highest probability is selected as the predicted label during inference.

The hidden layers generally include convolutional layers, pooling layers, and fully connected (FC) layers, which are combined in various configurations to form the network architecture.

A convolutional layer consists of multiple feature maps and is responsible for extracting meaningful features from the input data through convolution operations. The convolution operation involves sliding a matrix-shaped kernel (or filter) over the input with a fixed stride, computing point-wise multiplications at each position, and summing the results. Convolutional layers employ weight sharing, and each neuron is connected only to a local receptive field rather than the entire input. This allows the model to effectively capture local patterns while preserving the spatial structure of the data. As the network deepens, it enables the model to learn hierarchical representations, ranging from low-level to high-level features. The pooling layer, typically located after a convolutional layer, performs downsampling operations such as max pooling or average pooling. This reduces the spatial dimensions of the feature maps while retaining the most salient features, thereby lowering the computational burden and enhancing robustness to small spatial variations. The fully connected layer, located at the end of the hidden layers, connects every neuron to all neurons in the previous layer. It receives the flattened feature maps and is primarily responsible for the final classification task in the CNN [40–42]. Through this hierarchical structure, CNNs preserve the spatial configuration of the input data and are highly effective at learning and representing local patterns.

In this study, we treat the three-dimensional correlation coefficient matrices, decomposed by frequency bands, as image-like representations and feed them into a CNN to effectively learn local spatial patterns. The architectural characteristics of CNN —namely,

weight sharing and local connectivity— allow the model to precisely capture repetitive correlation structures, similar patterns, and local structures such as clusters among assets within each frequency band. This ability to extract spatial features is advantageous for learning subtle inter-asset relationships that are often overlooked by traditional models, thereby enabling a more effective reflection of the latent structural information embedded in the correlation matrices. In particular, CNNs are highly responsive to local patterns within each frequency-specific correlation matrix, allowing the model to hierarchically learn multidimensional correlation structures and ultimately derive richer and more expressive features [43].

4.3. Transformer

The Transformer was first introduced in the 2017 paper "Attention Is All You Need" Vaswani et al. [9], and since then, it has been actively studied in a wide range of AI fields, including natural language processing, computer vision, and time series analysis [44,45].

The Transformer is an advanced form of the traditional sequence-to-sequence architecture, consisting of an encoder that processes the input data and a decoder that generates the output sequence. The encoder takes the input sequence used for prediction, while the decoder learns to produce the next output by receiving part of the ground truth sequence as input. Figure 4 shows a typical Transformer architecture.

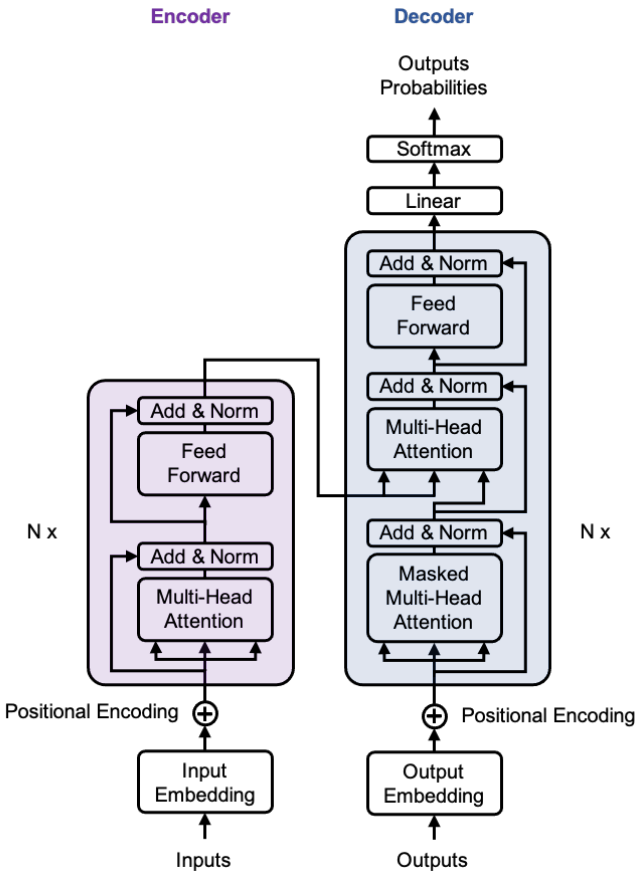


Figure 4. Standard Transformer architecture.

At the core of the Transformer is "positional encoding" and a "multi-head attention" mechanism called self-attention, which dynamically captures the dependencies between different elements in the input sequence, allowing the model to efficiently integrate information across the entire sequence. Since the Transformer processes sequences in parallel, positional information (time information) is not inherently preserved; to address this,

positional encoding is added to provide the model with the order of the sequence. In particular, multi-head attention enables the model to focus on different parts of the input simultaneously from multiple perspectives, effectively overcoming limitations found in RNN-based models, such as the long-term dependency problem, vanishing gradients, and gradient explosion [46].

These structural advantages make the Transformer highly suitable for time series modeling. In this study, we leverage the Transformer's self-attention mechanism and positional encoding to effectively learn global temporal interactions and long-term dependencies across the entire time series. This enables the model to capture not only short-term fluctuations but also long-term patterns and trends, modeling dependencies even between distant time points. As a result, the Transformer produces representations that reflect the full-range temporal context of time series data, allowing it to explain complex patterns that are difficult to capture using local models [47].

4.4. The Proposed Model

Based on the above, we propose a model called Wavelet decomposition + CNN + Transformer (WCT). The overall architecture of the proposed model is illustrated in Figure 5.

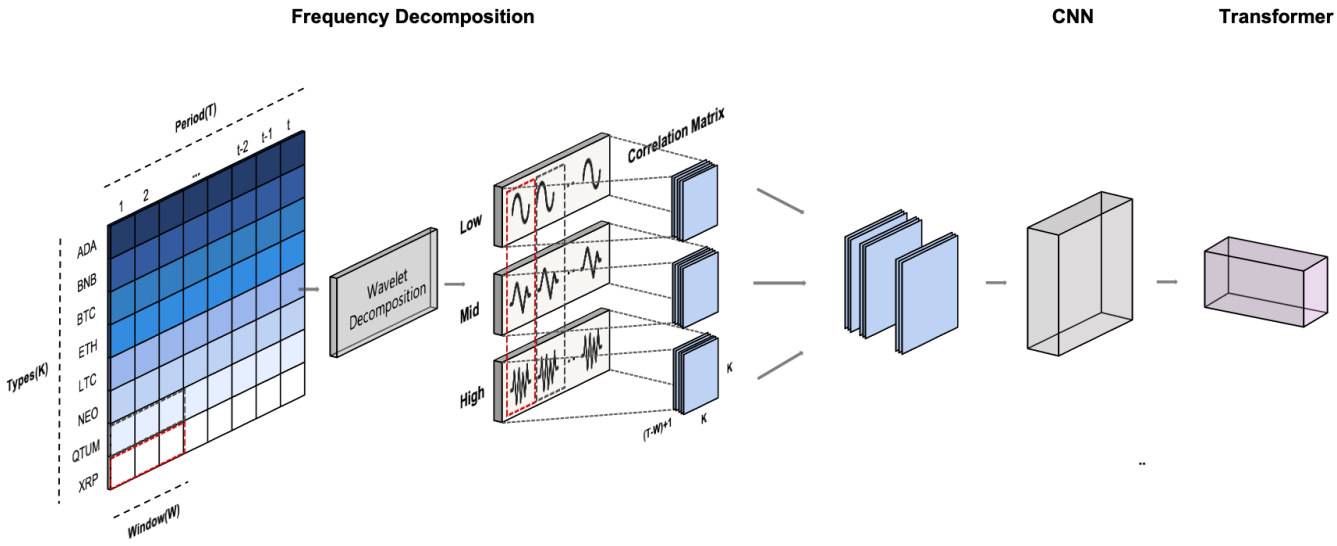


Figure 5. Overall architecture of the proposed model.

The objective of WCT is to predict future correlation matrices among multiple assets using cryptocurrency time series data. To this end, we design an architecture that captures multiresolution information, while jointly modeling spatial and temporal dependencies and local and global patterns based on the raw time series of each asset.

First, the time series data of each asset is segmented into sliding windows of 30 days. For each window, we apply a Wavelet transform to decompose using the Daubechies-4 basis with the signal into three frequency bands: low-frequency approximation, intermediate-frequency approximation, and high-frequency approximation [48]. Then, a correlation matrix is computed for each frequency band across all assets. These three matrices are stacked along the channel dimension, forming a tensor of shape $(3, 8, 8)$, which can be interpreted as a color (RGB) image containing frequency-specific spatial structures and local patterns of asset correlations.

This tensor is then passed into a CNN to extract features that capture both intra-band and inter-band correlation patterns. Following prior works such as [49–51], and to avoid overfitting due to model complexity, we design a CNN with two convolution layers, as

shown in Figure 6. The CNN output is summarized into a feature representation (global embedding vector) for each window.

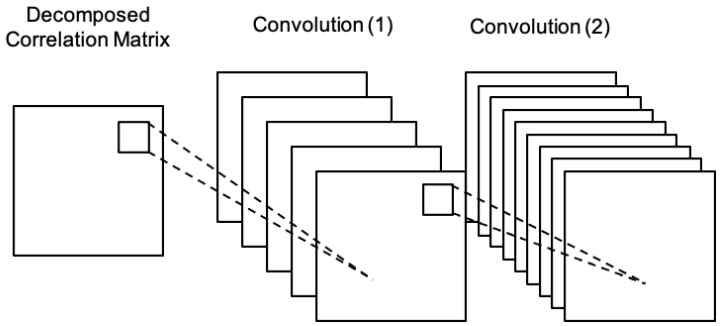


Figure 6. The proposed CNN architecture.

The resulting sequence of embedding vectors, corresponding to the CNN batch size (256 windows), is then fed into the Transformer encoder. Each window spans 30 days and is shifted sequentially by 1 day, so the full input sequence implicitly incorporates information from 285 days $(30 + (256 - 1))$. After enriching the sequence with positional encodings, the Transformer applies self-attention to capture long-term dependencies and global temporal interactions across all time steps. Finally, our model predicts the correlation matrix for the next 10-day horizon [17], serving as a short-term forecast of inter-asset correlations. We adopt an encoder-only Transformer design, following prior studies [52–54], as illustrated in Figure 7. This structure effectively models how past correlation patterns influence future dynamics and complements the CNN's role in extracting local features.

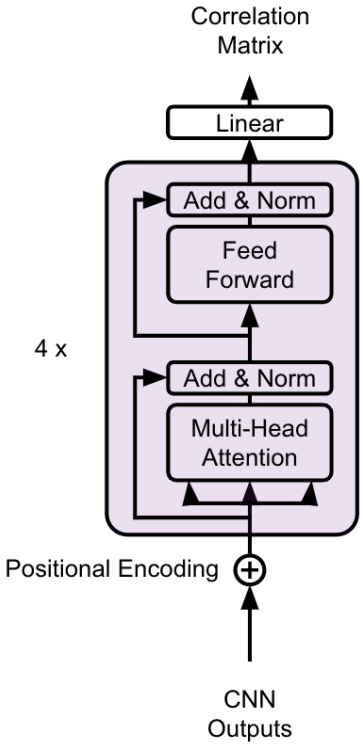


Figure 7. The proposed Transformer architecture.

Finally, the output of the Transformer is passed through an FC layer, which restores the representation to a matrix of shape (8×8). This serves as the prediction of the future correlation matrix among assets for the next 10-day period. The target matrix is the raw

(non-decomposed) correlation matrix, and the model is trained to leverage multi-frequency information to capture the dynamic structure of the market over time.

4.5. Model Training

Figure 8 presents the training mechanism of the proposed model. To determine the optimal hyperparameters that enhance model performance during training, we performed hyperparameter optimization using the grid search method.

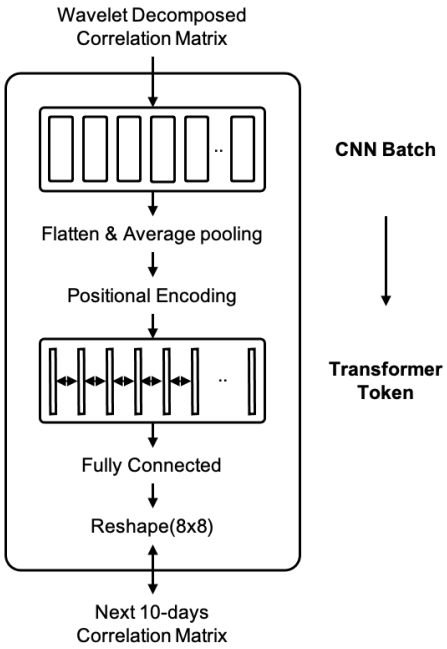


Figure 8. Proposed model training flow architecture.

The tuning process focused on key hyperparameters that govern both the architecture and the learning dynamics of the CNN-Transformer. For all training runs, the number of epochs was fixed at 70. Considering the dataset size and commonly adopted practices in related studies, we defined the following search ranges.

- **CNN parameters**: The kernel size (kernel_size) was set to [3, 5], and the embedding dimension (d_model)—which determines the output dimension of the second convolutional layer—was selected from [32, 64, 128]. Additionally, the batch size for the CNN output, which is subsequently used as the input embedding sequence size for the Transformer, was chosen from [64, 128, 256, 512].
- **Transformer parameters**: The number of heads in multi-head attention (nhead) was selected from [2, 4], and the number of Transformer layers (num_layers) from [2, 4]. The dimension of the feed-forward network (dim_feedforward) was set to [256, 512], and the activation function (activation) was selected between ReLU and GELU.
- **Optimization parameters**: The optimizer was chosen between Adam, AdamW, and RMSprop, and the learning rate (lr) was set to [0.001, 0.0005]. To prevent overfitting and promote stable training, a learning rate scheduler was employed, which automatically reduced the learning rate when the validation loss plateaued.

For model performance evaluation, the dataset was divided into training, validation, and test sets in an approximate 8:1:1 ratio. Out of a total of 2520 Wavelet-decomposed correlation matrices, we used 2016 samples for training, 252 samples for validation, and 252 samples for testing, maintaining strict chronological order to respect the temporal nature of the data.

For each hyperparameter combination, the model was initialized and trained using the same training and validation sets to ensure fair comparison. Model performance was evaluated based on the validation loss, and the combination that achieved the lowest validation loss was selected as the optimal set of hyperparameters, as shown in Table 2.

Table 2. Selected hyperparameter settings.

Hyperparameters	Setting
Kernel Size	3
d_{model}	128
Batch Size	256
Number of Heads (n_{head})	4
Number of Transformer Layers	4
Feedforward Dimension	256
Activation Function	GELU
Learning Rate	0.001
Optimizer	AdamW

Our proposed WCT model, regardless of its predictive performance, exhibited a limitation in that the resulting correlation matrices did not always satisfy the fundamental mathematical constraints of correlation matrices—namely, symmetry, unit diagonal, and positive semidefiniteness. To address this issue, we employed Higham's algorithm [55], which enforces these mathematical constraints through a post-processing step. Higham's algorithm is widely used in the financial domain to compute the nearest correlation matrix from a given symmetric matrix. The resulting matrix is defined as a symmetric positive semidefinite matrix with a unit diagonal. Specifically, we first enforced symmetry by averaging each predicted matrix with its transpose, and then we applied Higham's algorithm to obtain the nearest correlation matrix that satisfies all required mathematical properties.

5. Results

The Proposed Model

In this study, we evaluated the performance of the proposed model and verified the contribution of each component by conducting comparative experiments using identical datasets and under the same experimental conditions. Three variant models were selected as baselines, each composed of partial combinations of the proposed model's key components: *Wavelet-decomposed CNN*, *Wavelet-decomposed Transformer*, and *CNN Transformer*.

Each baseline model adopted the same architectural design principles as the proposed model, and identical hyperparameter search ranges and tuning procedures were applied. This ensured that differences in performance could be attributed solely to architectural variations, minimizing the influence of other confounding factors.

Furthermore, we compare the findings of this study with correlation forecasts obtained from a traditional econometric model, namely the DCC-GARCH (Dynamic Conditional Correlation—Generalized Autoregressive Conditional Heteroskedasticity) framework [56–58]. This comparison allows us to identify the differences between our machine learning–based results and those derived from conventional models, thereby providing evidence of the robustness and superiority of the proposed approach.

Performance was evaluated using standard regression metrics, including Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). Additionally, to measure the similarity between the predicted and ground truth matrices, Cosine Similarity and the Frobenius Norm were employed.

MSE represents the average of the squared differences between the predicted and actual values, placing greater penalties on larger errors. MAE calculates the average of

the absolute differences between the predicted and actual values, providing an intuitive measure of the average error magnitude while being less sensitive to outliers. RMSE, the square root of MSE, shares the same unit as the original data, allowing for a more intuitive interpretation of the error magnitude. Cosine Similarity measures the directional similarity between the predicted and actual vectors, where values closer to 1 indicate higher directional alignment. Finally, the Frobenius Norm quantifies the global distance between the predicted and actual matrices, enabling evaluation of prediction accuracy at the entire matrix level.

Let $\hat{Y} \in \mathbb{R}^{n \times n}$ denote the predicted matrix, $Y \in \mathbb{R}^{n \times n}$ the ground truth matrix, and n the matrix dimension. Each evaluation metric is defined as follows:

$$\text{MSE} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (y_{ij} - \hat{y}_{ij})^2 \tag{1}$$

$$\text{MAE} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} |y_{ij} - \hat{y}_{ij}| \tag{2}$$

$$\text{RMSE} = \sqrt{\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (y_{ij} - \hat{y}_{ij})^2} \tag{3}$$

$$\text{Cosine Similarity} = \frac{\text{Tr}(Y^\top \hat{Y})}{\|Y\|_F, \|\hat{Y}\|_F} \tag{4}$$

$$\text{Frobenius Norm} = \|Y - \hat{Y}\|_F = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} (y_{ij} - \hat{y}_{ij})^2} \tag{5}$$

$\text{Tr}(\cdot)$ denotes the trace operator.

Table 3 presents the results of the proposed model (WCT) and the three baseline models across the evaluation metrics. Overall, the proposed model achieved the best performance in all metrics, demonstrating improvements of approximately 0.01 to 0.02 over the baselines. Notably, WCT recorded the lowest error values in MSE, MAE, RMSE; the highest Cosine Similarity; and the lowest Frobenius Norm, indicating superior predictive accuracy and matrix similarity.

Table 3. Post-processed model performance comparison (Cosine Sim. = Cosine Similarity).

Model	MSE	MAE	RMSE	Cosine Sim.	Frobenius Norm
WCT (Proposed)	0.06176	**0.17248**	0.24852	**0.94646**	**1.78960**
Wavelet-Decomposed CNN	0.06634	0.17909	0.25757	0.94252	1.87537
Wavelet-Decomposed Transformer	0.06500	0.17345	0.25494	0.94532	1.84685
CNN Transformer	**0.06132**	0.17983	**0.24763**	0.94624	1.83247
DCC-GARCH	0.06635	0.18725	0.25759	0.94332	1.92294

From a structural perspective, WCT integrates frequency decomposition, CNN, and Transformer components. This combination yielded more stable and accurate predictions compared to models using only a subset of these components. In particular, when compared to the Wavelet-decomposed Transformer, WCT exhibited the largest performance gains across all metrics, suggesting that CNN-based local pattern extraction from frequency-decomposed inputs effectively complements the Transformer's global pattern learning capabilities.

In particular, when comparing the WCT with the traditional econometric model DCC-GARCH(1,1) based on evaluation metrics, it is evident that the WCT demonstrates superior

performance. This finding indicates that the WCT proposed in this study possesses strong competitiveness relative to conventional econometric models.

Table 4 presents the per-currency evaluation results obtained from the WCT model. Among the eight test assets, NEO consistently achieved the best performance across all metrics (MSE, MAE, RMSE, and Cosine Similarity), indicating that its correlation structure was most accurately captured by the model. In contrast, XRP exhibited the weakest predictive performance, with higher error values and the lowest cosine similarity, suggesting that its correlation dynamics were relatively more challenging to model.

Table 4. Post-processed per-currency evaluation metrics.

Currency	MSE	MAE	RMSE	Cosine Similarity
BTC	0.05405	0.16291	0.23248	0.95725
ETH	0.05848	0.15899	0.24184	0.95268
BNB	0.06208	0.17898	0.24916	0.94430
NEO	0.04627	0.15144	0.21510	0.96522
LTC	0.07766	0.18676	0.27868	0.92751
QTUM	0.05871	0.17529	0.24231	0.95275
ADA	0.05586	0.16339	0.23634	0.95576
XRP	0.08098	0.20206	0.28457	0.92533

We conducted a statistical significance test to evaluate the predictive performance of the proposed WCT model using the Diebold–Mariano (DM) test and corresponding p-values. The DM test is a statistical method designed to compare the predictive accuracy of two competing time series models by analyzing the difference in their forecast errors. A negative DM statistic indicates that the proposed model achieves a lower average loss than the baseline, implying superior predictive accuracy. The p-value represents the probability of observing such a difference under the null hypothesis that both models have equal predictive performance; conventionally, a value below 0.05 is considered statistically significant.

Table 5 summarizes the DM statistics and p-values of WCT against each baseline model. The evaluation was performed on the off-diagonal (excluding the diagonal) elements of the correlation matrices.

Table 5. Diebold–Mariano test results between the WCT model and baselines.

Baseline	DM (MSE)	p-Value (MSE)	DM (MAE)	p-Value (MAE)
WC	−2.221	0.027	−2.445	0.015
WT	−2.114	0.036	−0.405	0.686
CT	0.605	0.546	−6.973	0.000
DCC-GARCH	−3.279	0.001	−6.979	0.000

As shown in Table 5, the WCT model exhibits statistically significant improvements in predictive accuracy for most comparisons—specifically, all metrics except for CT (MSE) and WT (MAE). These results confirm that the proposed WCT model consistently achieves significantly lower forecast errors than the baselines, demonstrating its superior capability in modeling and forecasting time-varying correlations.

Figure 9 illustrates an example of the predicted matrix, ground truth matrix, and absolute error matrix for a single sample. While differences in magnitude are present, the proportional correlation patterns are observed at similar positions, confirming the model's ability to capture the overall relational structure. Moreover, we illustrated the same example for the DCC-GARCH(1,1) (Figure 10). Through this comparison, as previously

argued, it can be reaffirmed that the proposed WCT model exhibits superior performance in predicting the correlation structure compared to the DCC-GARCH(1,1) model.

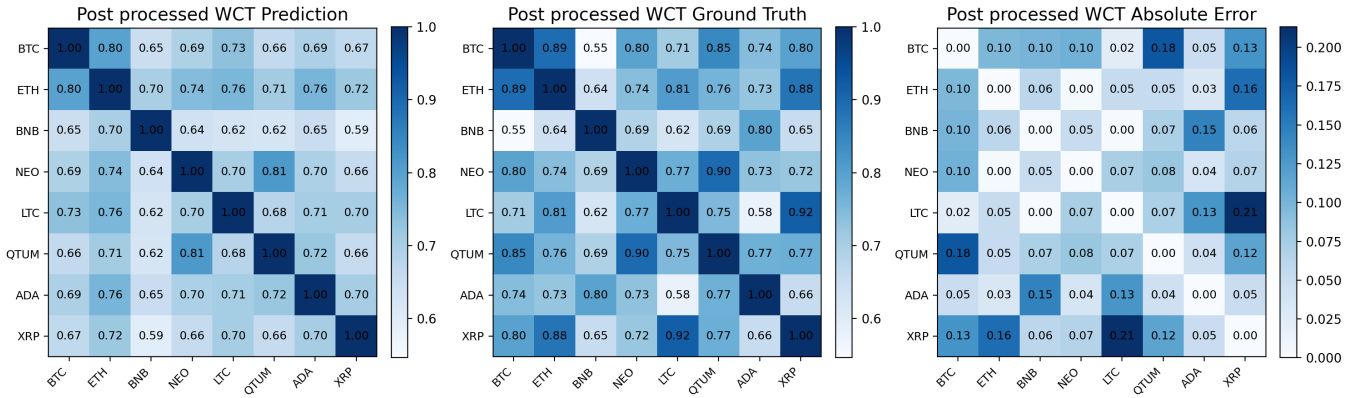


Figure 9. Post-processed WCT correlation matrix comparison.

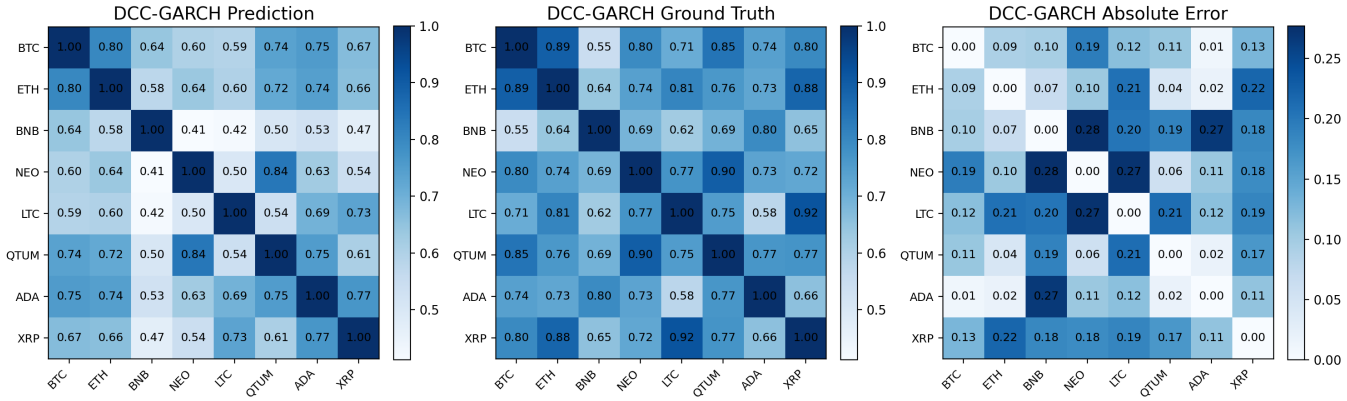


Figure 10. DCC-GARCH correlation matrix comparison.

Furthermore, Figure 11 shows the temporal evolution of each performance metric over the test dataset. In these plots, the red line represents MSE, the blue line represents MAE, the green line represents RMSE, the yellow line represents Cosine Similarity, and the purple line represents the Frobenius Norm. While all models display broadly similar temporal trends, WCT consistently maintains lower error metrics (MSE, MAE, RMSE) and higher Cosine Similarity compared to the baselines at most time points. In particular, for the Frobenius Norm, other models exhibit sharp increases at specific time intervals, whereas WCT remains relatively stable, demonstrating its strength in maintaining global prediction accuracy.

Moreover, to analyze the characteristics of prediction errors, we visualized the error distribution of the upper-triangular elements of the predicted correlation matrices, representing pairwise asset correlations. As shown in Figure 12, the error distribution is asymmetric and right-skewed, with the majority of errors concentrated around slightly negative values (approximately between –0.25 and 0.1).

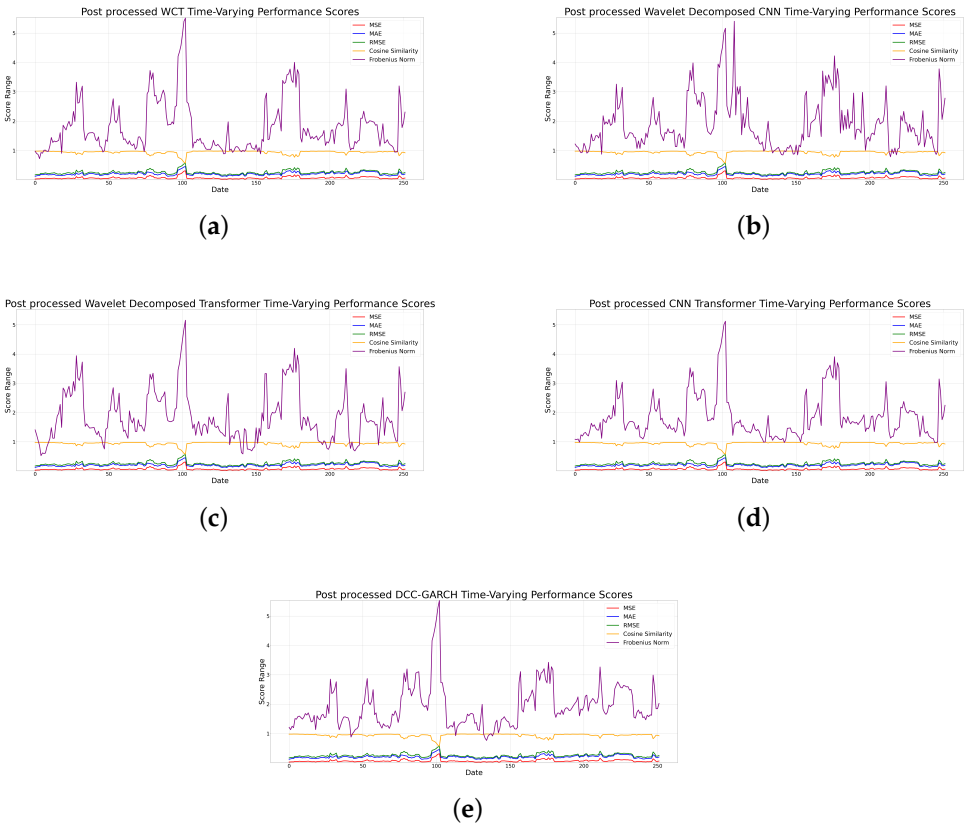


Figure 11. Comparison of post-processed model performance scores across all models. (**a**) WCT. (**b**) Wavelet CNN. (**c**) Wavelet Transformer. (**d**) CNN Transformer. (**e**) DCC-GARCH.

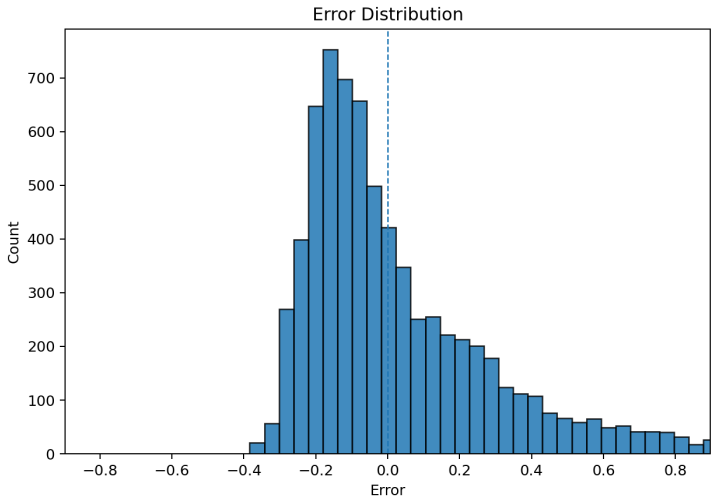


Figure 12. Error distribution of WCT.

This indicates that the model tends to slightly underestimate correlations rather than overestimate them. The long positive tail suggests that, in some instances, the model produces relatively higher predicted correlations compared to the actual ones, although such cases are less frequent. Overall, the distribution remains centered near zero, implying that the WCT model provides unbiased and stable predictions across different asset pairs, with most prediction errors falling within a narrow range.

These results confirm that the proposed model provides robust performance even under temporal variations in data distribution, and that the combination of frequency

decomposition with CNN and Transformer components offers superior generalization capability compared to single-component or partially combined architectures.

6. Discussion and Concluding Remarks

In this study, we proposed a hybrid model that integrates Wavelet frequency decomposition, CNN, and Transformer architectures (WCT) to predict the correlation structures among eight major cryptocurrencies. The Wavelet-based frequency decomposition module was employed to separate each asset's time series into long-, medium-, and short-term components, enabling multi-scale trend analysis. The CNN component was responsible for capturing spatial and local correlation patterns between assets, while the Transformer component modeled global temporal dependencies.

To verify the contribution of each component, we conducted ablation experiments by systematically altering the model architecture to form three baseline configurations: Wavelet-decomposed CNN, Wavelet-decomposed Transformer, and CNN Transformer. Furthermore, to benchmark the performance of the WCT model against traditional econometric approaches, we also employed the well-established DCC-GARCH model, which is widely recognized for modeling dynamic correlation structures. Furthermore, to obtain more stable and reliable prediction results, we applied Nick Higham's algorithm during the post-processing stage. Experimental results demonstrated that the proposed WCT consistently outperformed the baseline models across all evaluation metrics, including MSE, MAE, RMSE, Cosine Similarity, and Frobenius Norm. Notably, the performance gap between WCT and the Wavelet-decomposed Transformer model highlighted the crucial role of CNN in enhancing prediction accuracy when processing frequency-decomposed features. Furthermore, the WCT model maintained stable performance over time, indicating its robustness to temporal fluctuations in data distribution.

This study applies Wavelet decomposition to separate raw time series into short-, medium-, and long-term components, effectively extracting multi-scale information; A CNN is used to capture the local patterns of frequency band-specific correlation matrices, while the Transformer models both long-term dependencies and global correlations, thereby enhancing complex pattern recognition compared to single-structure approaches. The proposed model achieved either the best or the second-best performance across all evaluation metrics, with the exception of RMSE. Moreover, it maintained stable predictive accuracy even during periods of heightened market volatility.

Based on mean–variance theory, accurately forecasting correlations maximizes the benefits of portfolio diversification and enables the design of more sophisticated risk-hedging strategies than approaches relying solely on volatility. Furthermore, the proposed WCT model offers the potential for quantitative tracking of structural changes in the market, which are often difficult to detect using traditional price-prediction models. Although the proposed model does not exhibit dramatically superior performance compared to existing approaches, it can be regarded as a promising and practically useful model that demonstrates potential applicability.

Nevertheless, this study is subject to certain limitations. The primary challenge of our research is that the experiments were conducted on a fixed set of eight cryptocurrencies and daily time intervals, which may limit the model's generalizability to other asset classes or higher-frequency datasets. Moreover, exogenous market factors such as macroeconomic events, regulatory changes, and sentiment data were not explicitly incorporated into the model, potentially leaving unexplained variations in correlation dynamics.

Finally, we propose several directions for future research. First, the framework could be extended to a broader set of financial assets and different temporal resolutions, while also incorporating external information sources such as news and social media sentiment

to enhance the accuracy of correlation prediction. Second, future studies may employ multiple datasets across diverse markets and sample periods to examine the consistency and robustness of the findings. Third, exploring alternative temporal frequencies such as intraday or weekly horizons could provide deeper insights into the dynamics of spillovers across different time scales. Lastly, extending the analysis beyond cryptocurrencies to include traditional asset classes such as equities, bonds, and commodities would allow for a more comprehensive assessment of the generalisability and practical relevance of the results. Moreover, in future work, we plan to empirically support these claims through extensive case studies and financial backtesting, which will help evaluate the practical applicability and robustness of the proposed framework in real-world financial environments.

Author Contributions: Conceptualization, J.-W.K.; Data curation, J.-W.K.; Formal analysis, J.-W.K., D.K. and S.-Y.C.; Investigation, J.-W.K. and S.-Y.C.; Methodology, J.-W.K.; Software, J.-W.K.; Supervision, S.-Y.C.; Visualization, J.-W.K.; Writing—original draft, J.-W.K. and S.-Y.C.; Writing—review & editing, D.K. and S.-Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2024-00454493) and by Seoul R&BD Program(QR240016) through the Seoul Business Agency (SBA) funded by Seoul Metropolitan Government.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Conflicts of Interest: Author Daihyun Kwon was employed by the Quantum Intelligence Corp. The remaining authors declare that this research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. *Order Granting Approval of Proposed Rule Changes to List and Trade Shares of Spot Bitcoin Exchange-Traded Products*; Securities Exchange Act Release 34-99306; U.S. Securities and Exchange Commission: Washington, DC, USA, 2024.
2. Liu, Y.; Tsyvinski, A. Risks and Returns of Cryptocurrency. *Rev. Financ. Stud.* **2021**, *34*, 2689–2727. [CrossRef]
3. Borri, N. Conditional tail-risk in cryptocurrency markets. *J. Empir. Financ.* **2019**, *50*, 1–19. [CrossRef]
4. Bouri, E.; Kamal, E.; Kinateder, H. FTX Collapse and systemic risk spillovers from FTX Token to major cryptocurrencies. *Financ. Res. Lett.* **2023**, *56*, 104099. [CrossRef]
5. Sims, C.A. Macroeconomics and Reality. *Econometrica* **1980**, *48*, 1–48. [CrossRef]
6. De Livera, A.M.; Hyndman, R.J.; Snyder, R.D. Forecasting time series with complex seasonal patterns using exponential smoothing. *J. Am. Stat. Assoc.* **2011**, *106*, 1513–1527. [CrossRef]
7. Jung, G.; Choi, S.Y. Forecasting foreign exchange volatility using deep learning autoencoder-LSTM techniques. *Complexity* **2021**, *2021*, 6647534. [CrossRef]
8. Kim, J.; Kim, H.S.; Choi, S.Y. Forecasting the S&P 500 index using mathematical-based sentiment analysis and deep learning models: A FinBERT transformer model and LSTM. *Axioms* **2023**, *12*, 835.
9. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 6000–6010.
10. Li, S.; Jin, X.; Xuan, Y.; Zhou, X.; Chen, W.; Wang, Y.X.; Yan, X. Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 5243–5253.
11. Wu, H.; Xu, J.; Wang, J.; Long, M. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. In Proceedings of the Advances in Neural Information Processing Systems, Online, 6–14 December 2021; Curran Associates, Inc.: Red Hook, NY, USA, 2021; Volume 34, pp. 22419–22430.
12. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 11106–11115. [CrossRef]
13. Zhou, T.; Ma, Z.; Wen, Q.; Wang, X.; Sun, L.; Jin, R. FEDformer: Frequency Enhanced Decomposed Transformer for Long-term Series Forecasting. In Proceedings of the 39th International Conference on Machine Learning, Baltimore, MD, USA, 17–23 July 2022; Volume 162, pp. 27268–27286.
14. Zeng, A.; Chen, M.; Zhang, L.; Xu, Q. Are Transformers Effective for Time Series Forecasting? *Proc. AAAI Conf. Artif. Intell.* **2023**, *37*, 10913–10921. [CrossRef]

15. Xu, J.; Cao, L. Copula variational LSTM for high-dimensional cross-market multivariate dependence modeling. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *35*, 16233–16247. [CrossRef]
16. Widiputra, H.; Mailangkay, A.; Gautama, E. Multivariate CNN-LSTM model for multiple parallel financial time-series prediction. *Complexity* **2021**, *2021*, 9903518. [CrossRef]
17. Lu, W.; Li, J.; Li, Y.; Sun, A.; Wang, J. A CNN-LSTM-based model to forecast stock prices. *Complexity* **2020**, *2020*, 6622927. [CrossRef]
18. Zha, W.; Liu, Y.; Wan, Y.; Luo, R.; Li, D.; Yang, S.; Xu, Y. Forecasting monthly gas field production based on the CNN-LSTM model. *Energy* **2022**, *260*, 124889. [CrossRef]
19. Bai, S.; Kolter, J.Z.; Koltun, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv* **2018**, arXiv:1803.01271. [CrossRef]
20. Markowitz, H. Portfolio Selection. *J. Financ.* **1952**, *7*, 77–91. [CrossRef]
21. Rockafellar, R.T.; Uryasev, S. Optimization of Conditional Value-at-Risk. *J. Risk* **2000**, *2*, 21–41. [CrossRef]
22. Bollerslev, T. A conditionally heteroskedastic time series model for speculative prices and rates of return. *Rev. Econ. Stat.* **1987**, *69*, 542–547. [CrossRef]
23. Barbierato, E.; Gatti, A.; Incremona, A.; Pozzi, A.; Toti, D. Breaking Away From AI: The Ontological and Ethical Evolution of Machine Learning. *IEEE Access* **2025**, *13*, 55627–55647. [CrossRef]
24. Reinsel, G.C. *Elements of Multivariate Time Series Analysis*; Springer: New York, NY, USA, 1993.
25. Box, G.E.P.; Jenkins, G.M. *Time Series Analysis: Forecasting and Control*, revised ed.; Holden-Day: San Francisco, CA, USA, 1976.
26. Chen, W.; Wang, W.; Peng, B.; Wen, Q.; Zhou, T.; Sun, L. Learning to Rotate: Quaternion Transformer for Complicated Periodical Time Series Forecasting. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), Washington, DC, USA, 14–18 August 2022; pp. 146–156. [CrossRef]
27. Liu, S.; Yu, H.; Liao, C.; Li, J.; Lin, W.; Liu, A.X.; Dustdar, S. Pyraformer: Low-Complexity Pyramidal Attention for Long-Range Time Series Modeling and Forecasting. In Proceedings of the International Conference on Learning Representations (ICLR 2022), Virtual Event, 25–29 April 2022.
28. Zhang, Y.; Yan, J. Crossformer: Transformer Utilizing Cross-Dimension Dependency for Multivariate Time Series Forecasting. In Proceedings of the ICLR, Kigali, Rwanda, 1–5 May 2023.
29. Kim, J.; Park, S. A Convolutional Transformer Model for Multivariate Time Series Prediction. *IEEE Access* **2022**, *10*, 101319–101329. [CrossRef]
30. Allen, J. Short term spectral analysis, synthesis, and modification by discrete Fourier transform. *IEEE Trans. Acoust. Speech Signal Process.* **2003**, *25*, 235–238. [CrossRef]
31. Mallat, S.G. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *11*, 674–693. [CrossRef]
32. Bhowmick, A.; Chandra, M. Speech enhancement using voiced speech probability based wavelet decomposition. *Comput. Electr. Eng.* **2017**, *62*, 706–718. [CrossRef]
33. Lu, W.; Ghorbani, A.A. Network anomaly detection based on wavelet analysis. *EURASIP J. Adv. Signal Process.* **2008**, *2009*, 837601. [CrossRef]
34. Wang, J.; Wang, Z.; Li, J.; Wu, J. Multilevel wavelet decomposition network for interpretable time series analysis. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 2437–2446.
35. Berger, T. Forecasting based on decomposed financial return series: A wavelet analysis. *J. Forecast.* **2016**, *35*, 419–433. [CrossRef]
36. Tang, Q.; Shi, R.; Fan, T.; Ma, Y.; Huang, J. Prediction of financial time series based on LSTM using wavelet transform and singular spectrum analysis. *Math. Probl. Eng.* **2021**, *2021*, 9942410. [CrossRef]
37. Fernández-Macho, J. Wavelet multiple correlation and cross-correlation: A multiscale analysis of Eurozone stock markets. *Phys. A Stat. Mech. Its Appl.* **2012**, *391*, 1097–1104. [CrossRef]
38. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* **1989**, *1*, 541–551. [CrossRef]
39. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
40. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **2021**, *8*, 53. [CrossRef]
41. Li, Z.; Liu, F.; Yang, W.; Peng, S.; Zhou, J. A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 6999–7019. [CrossRef] [PubMed]
42. Alshingiti, Z.; Alaqel, R.; Al-Muhtadi, J.; Haq, Q.E.U.; Saleem, K.; Faheem, M.H. A deep learning-based phishing detection system using CNN, LSTM, and LSTM-CNN. *Electronics* **2023**, *12*, 232. [CrossRef]

43. Luo, A.; Zhong, L.; Wang, J.; Wang, Y.; Li, S.; Tai, W. Short-term stock correlation forecasting based on CNN-BiLSTM enhanced by attention mechanism. *IEEE Access* **2024**, *12*, 29617–29632. [CrossRef]
44. Han, K.; Wang, Y.; Chen, H.; Chen, X.; Guo, J.; Liu, Z.; Tang, Y.; Xiao, A.; Xu, C.; Xu, Y.; et al. A survey on vision transformer. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 87–110. [CrossRef]
45. Wen, Q.; Zhou, T.; Zhang, C.; Chen, W.; Ma, Z.; Yan, J.; Sun, L. Transformers in time series: A survey. *arXiv* **2022**, arXiv:2202.07125.
46. Lin, T.; Wang, Y.; Liu, X.; Qiu, X. A survey of transformers. *AI Open* **2022**, *3*, 111–132. [CrossRef]
47. Ahmed, S.; Nielsen, I.E.; Tripathi, A.; Siddiqui, S.; Ramachandran, R.P.; Rasool, G. Transformers in time-series analysis: A tutorial. *Circuits Syst. Signal Process.* **2023**, *42*, 7433–7466. [CrossRef]
48. Ghosh, S.; Manimaran, P.; Panigrahi, P.K. Characterizing multi-scale self-similar behavior and non-statistical properties of fluctuations in financial time series. *Phys. A Stat. Mech. Its Appl.* **2011**, *390*, 4304–4316. [CrossRef]
49. Gudelek, M.U.; Boluk, S.A.; Ozbayoglu, A.M. A deep learning based stock trading model with 2-D CNN trend detection. In Proceedings of the 2017 IEEE Symposium Series on Computational Intelligence (SSCI), Honolulu, HI, USA, 27 November–1 December 2017; pp. 1–8.
50. Livieris, I.E.; Pintelas, E.; Pintelas, P. A CNN–LSTM model for gold price time-series forecasting. *Neural Comput. Appl.* **2020**, *32*, 17351–17360. [CrossRef]
51. Chen, J.F.; Chen, W.L.; Huang, C.P.; Huang, S.H.; Chen, A.P. Financial time-series data analysis using deep convolutional neural networks. In Proceedings of the 2016 7th International Conference on Cloud Computing and Big Data (CCBD), Macau, China, 16–18 November 2016; pp. 87–92.
52. Zeng, Z.; Kaur, R.; Siddagangappa, S.; Rahimi, S.; Balch, T.; Veloso, M. Financial time series forecasting using CNN and transformer. *arXiv* **2023**, arXiv:2304.04912. [CrossRef]
53. Bui, N.K.H.; Chien, N.D.; Kovács, P.; Bognár, G. Transformer Encoder and Multi-features Time2Vec for Financial Prediction. *arXiv* **2025**, arXiv:2504.13801.
54. Izadi, M.A.; Hajizadeh, E. Time Series Prediction for Cryptocurrency Markets with Transformer and Parallel Convolutional Neural Networks. *Appl. Soft Comput.* **2025**, *177*, 113229. [CrossRef]
55. Higham, N.J. Computing the nearest correlation matrix—A problem from finance. *IMA J. Numer. Anal.* **2002**, *22*, 329–343. [CrossRef]
56. Celık, S. The more contagion effect on emerging markets: The evidence of DCC-GARCH model. *Econ. Model.* **2012**, *29*, 1946–1959. [CrossRef]
57. Shiferaw, Y.A. Time-varying correlation between agricultural commodity and energy price dynamics with Bayesian multivariate DCC-GARCH models. *Phys. A Stat. Mech. Its Appl.* **2019**, *526*, 120807. [CrossRef]
58. Ringim, S.H.; Alhassan, A.; Güngör, H.; Bekun, F.V. Economic policy uncertainty and energy prices: Empirical evidence from multivariate DCC-GARCH models. *Energies* **2022**, *15*, 3712. [CrossRef]

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