Multi-Timeframe Transformer Architecture for Bitcoin Price Prediction: Deep Dive into the Research Project

MSc Research Project
Practicum (MSCFTD1)

Sanjana Gandrathi
Student ID: X23290692

School of Computing
National College of Ireland

Supervisor: Victor del Rosal

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Sanjana Gandrathi ……. …………………………………………………………………………………………
Student ID:	X23290692 ……………………………………………………………………………………………….……
Programme:	MSC in FinTech (MSCFTD1) ………………………………………… **Year:** 2024-2025 …………………………..
Module:	Practicum …………………………………………………………………………………………….………
Supervisor:	Victor del Rosal …………………………………………………………………………………………….………
Submission Due Date:	15-09-2025 …………………………………………………………………………………………….………
Project Title:	Multi-Timeframe Transformer Architecture for Bitcoin Price Prediction: Deep Dive into the Research Project …………………………………………………………………………………………….………
Word Count:	9266 ……………………………………… **Page Count** 26 ………………………………………….……..

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Sanjana Gandrathi ……………………………………………………………………………………………………

Date: 15-09-2025 ……………………………………………………………………………………………………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Multi-Timeframe Transformer Architecture for Bitcoin Price Prediction: Deep Dive into the Research Project

Sanjana Gandrathi
Student ID: X23290692

Abstract

The cryptocurrency markets are highly volatile and have complicated temporal dynamics that make conventional forecasting techniques less effective. Although deep learning methods have been promising, they generally study individual timeframes and do not adequately combine sentiment signals. This research investigates whether a multi-timeframe Transformer architecture can improve Bitcoin price prediction accuracy compared to traditional LSTM models, while examining how social sentiment importance varies across temporal resolutions.

The study develops a novel Transformer architecture that concurrently processes hourly, daily, and weekly temporal patterns through specialized cross-timeframe attention mechanisms. The model integrates 97,000 sentiment-labeled social media messages with historical price data spanning 2014-2019, creating a comprehensive dataset of 41,120 samples with 104 engineered features. Comparative experiments evaluate five model configurations including single and multi-timeframe variants with sentiment ablation studies.

With multi-timeframe LSTM attaining a 34.8% RMSE reduction over single-timeframe variants, the results show that the multi-timeframe approach offers significant improvements. But surprisingly, the multi-timeframe LSTM performs better than the Transformer architecture, attaining 57.8% directional accuracy as opposed to 56.6% for the Transformer. Interestingly, sentiment integration shows model-dependent effects, reducing LSTM performance by 7.7% while increasing Transformer accuracy by 20.7% RMSE reduction. With a directional accuracy of 62.1%, the LSTM architecture demonstrates exceptional strength in 24-hour predictions. These results imply that while architectural advancements in temporal processing for cryptocurrency forecasting hold promise, they primarily depend on matching architecture to data attributes. With implications for trading systems and risk management protocols, the paper presents a well-established paradigm for analysis across multiple timescales and reveals intricate relationships between model architecture and sentiment feature success.

1 Introduction

Since the creation of Bitcoin in 2009, the cryptocurrency market has grown at an unprecedented rate, transforming from a specialised technical experiment to a worldwide financial phenomenon with a valuation of over $1 trillion. The market has faced enormous challenges disguised as volatility and unpredictability as a result of the aforementioned meteoric rise, which has presented investors with enormous opportunities (Kavithra et al., 2024). The unique dynamics of the price movement of those cryptocurrencies, whose factors range from

technological advancements to Internet-based sentiments on social media, are not provided by the conventional methods of financial forecasting of traditional asset markets.
It is hard to predict the price of cryptocurrencies because of how quickly regimes change and how nonlinear dynamics work. Compared to other financial markets, cryptocurrencies are very volatile, with daily changes of more than 10% (Zhang et al., 2023). Continuous global trading around the clock and the lack of circuit breakers or traditional market-making make things even less stable. Social media, especially Twitter and Reddit, is a major way for information to spread and for people to form opinions (Feizian & Amiri, 2023).

Recent advancements in deep learning, particularly the development of Transformer architecture, has been promising in capturing rich temporal dependencies in sequential data. However, existing approaches significantly focus on single-timeframe analysis and may disregard significant patterns existing at different temporal resolutions. The employment of multi-scale temporal features in addition to social sentiment data constitutes an unexplored potential to increase the level of prediction accuracy. A research void motivates the introduction of multi-timeframe Transformer architectures capable of processing hourly, daily, and weekly patterns in parallel while incorporating sentiment signals.

This research addresses the following question: **To what extent can a multi-timeframe Transformer-based architecture improve Bitcoin price prediction accuracy compared to traditional LSTM models, and how does the relative importance of social sentiment data vary across different temporal resolutions?** The primary objectives include: (1) developing a novel multi-timeframe Transformer architecture that processes temporal data at multiple resolutions, (2) integrating social sentiment analysis through an attention-based fusion mechanism, (3) conducting comprehensive comparisons with established baseline models including LSTM, GRU, and traditional machine learning approaches, and (4) analyzing model performance across different prediction horizons.

This work transcends practical implementation and scholarly investigation by addressing deficiencies in multi-scale temporal analysis of cryptocurrency markets and enhancing deep learning techniques for financial time-series forecasting. It presents a dynamic predictive model that can enhance risk management and decision-making for practitioners, including traders, investors, and financial institutions. The multi-timeframe design lets you use a wide range of strategies, from long-term investments based on weekly patterns to shorter-term ones based on hourly forecasts.

The approach combines rigorous empirical analysis with a comprehensive framework of the most recent deep learning techniques. The study uses a corpus of 97,000 tweets with sentiment analysis on cryptocurrencies and historical Bitcoin price data spanning multiple years. The huge Google Colab computer system with A100 GPU accelerators is used to train the Transformer architecture's framework, which contains unique sub-modules for temporal encoding and cross-time period attention.

The format of this report is as follows: A critical analysis of the body of research on sentiment analysis integration, deep learning applications, and cryptocurrency price prediction is given in Section 2. The research methodology, including data collection, preprocessing, and model architecture, is described in detail in Section 3. The multi-timeframe Transformer model's design specifications are shown in Section 4. The specifics of the implementation are covered in Section 5. Experimental results and critical analysis are presented in Section 6. The results and their implications are discussed in Section 7. With important contributions and suggestions for further research, Section 8 comes to a close.

2 Related Work

2.1 Traditional Approaches to Cryptocurrency Price Prediction

The development of cryptocurrency price forecasting methods mirrors the maturation of the market and the analytical methods. Previous studies extensively utilized traditional econometric models borrowed from mainstream financial markets. Oyewola et al. (2023) performed a critical assessment of the classical time series models for forecasting cryptocurrencies and showed that though the ARIMA models were decent in short-term forecasting, the models were unable to account for the non-linear dynamics typical in cryptocurrency markets. The study by Oyewola et al. (2023) showed that the classical models averaged 15% average RMSE in daily Bitcoin price forecasting, documenting the shortcomings in the linear methods in this domain.

The application of machine learning methods was a significant leap in the degree of prediction accuracy. Kavithra et al. (2024) integrated many forecasting techniques, combining amalgamating classical statistical methods and machine learning methods. Their hybrid method involving the adoption of both Random Forest and XGBoost models was superior to statistical methods alone by decreasing the prediction error by 23%. However, these models still struggled with capturing long-range dependencies and complex temporal patterns, particularly during periods of high volatility. The study's strength lies in its comprehensive evaluation across 20 cryptocurrencies, though it primarily focused on single-timeframe analysis, missing potential insights from multi-scale temporal patterns.

2.2 Deep Learning Advances in Cryptocurrency Forecasting

The introduction of deep learning approaches, particularly recurrent neural networks, revolutionized the price forecasting of cryptocurrencies. A complete review by Zhang et al. (2023) on the usage of deep learning applications in the field of cryptocurrencies noted the LSTM and GRU designs as the most dominant approaches in price prediction. Their meta-analysis of 93 studies showed that the LSTM models in all cases outperformed classical approaches by an average improvement in prediction accuracy by 35%. The survey showed particularly promising results for bidirectional LSTM (Bi-LSTM) architecture, where in certain studies the direction prediction accuracies reached over 60%.

Despite these advancements, traditional LSTM approaches are inefficient in the processing of multi-scale temporal features. Kabir et al. (2025) addressed this challenge by coming up with a hybrid LSTM-Transformer model in which the sequential processing by LSTMs has been merged with the parallel processing feature of Transformers. Their architecture reached 12% improvement in RMSE compared with standalone LSTM networks on six finance datasets, including Bitcoin and well-known stock indices. Still, the multi-timeframe feature in financial data was by no means fully exploited by them as they processed single-resolution inputs only. This limitation suggests the potential for further improvements through explicit multi-timeframe modeling.

2.3 Sentiment Analysis Integration in Price Prediction

The unique influence of social sentiment on cryptocurrency markets has motivated extensive research into sentiment-based prediction models. Feizian and Amiri (2023) developed a unified framework for combining sentiment analysis from Twitter and forecasting models and demonstrated sentiment features had improved forecasting performance by up to 18%. The researchers established that influence of sentiment varied significantly with the horizon and that they had larger sentiment relations with prices for the short-run (hourly) predictions than for the longer-run (weekly) predictions. There was, however, no study conducted on the role of sentiment in various states of the market.

Improvements in natural language processing made it possible for more sophisticated forms of sentiment analysis to be used. Wahidur et al. (2024) used instruction-tuned large language models for the sentiment classification of cryptocurrencies and was successful in boosting the accuracy of sentiment identification by 40% relative to traditional methods. Their study implied that sophisticated language models were successful in picking up subtle manifestations of sentiment that were typical of the slang of cryptocurrencies. Though such sophisticated sentiment features have been incorporated into models for price prediction on various timeframes until now, they have gone without comment.

2.4 Transformer Architectures for Financial Time Series

The usage of Transformer structures in financial time series is a new but promising approach. Mai et al. (2024) compared the performance of several Transformer variations in the task of forecasting Bitcoin prices, namely standard Transformer, Autoformer, and Informer models. Their findings indicated the highest performance by Autoformer with a 15% improvement in the reduction of prediction over LSTM baselines, outperforming especially in the identification of long-range dependencies. Their analysis was though limited in considering only single-timeframe inputs and failing to utilize external feature sources such as sentiment data.

Transformers optimized for multi-scale processing of the temporal axis for financial markets have remained untouched. Though numerous works have indicated adaptations for processing long sequences, no architecture for processing multiple temporal scales at a time has ever surfaced. Such remains particularly true for crypto markets where the features manifest at multiple diversified timescales ranging from smallest frequency transaction signals to bigger-distance trends. A connection from the literature linking multi-timeframe Transformers and sentiment aggregation remains an open area for extending the boundaries of the art of crypto price prediction.

2.5 Synthesis and Research Gap

The literature under review proves an evident pattern of evolution from classical statistical methodologies to the latest deep learning methods for predicting the rates of cryptocurrencies. Although massive success has been achieved on individual lines of deep learning building blocks of architecture, sentiment analysis, and use cases of Transformers, the combination of the latter building blocks in a coherent multi-timeframe framework is not evident. The relevant studies tend to be almost unique in that they analyze in single-timeframe analysis, and they may overlook important cross-scale dependencies by which performance may significantly be boosted. In addition, the dynamic coupling of sentiment influence and market conditions at different temporal resolutions has not been researched systematically. This paper solves these deficiencies by presenting a new multi-timeframe Transformer architecture processing multiple temporal scales simultaneously and using sentiment features adaptively, and facilitating theoretical understanding and practical application in cryptocurrency price prediction.

3 Research Methodology

This section presents the comprehensive methodology employed to develop and evaluate the multi-timeframe Transformer architecture for Bitcoin price prediction. It includes the experimental design, evaluation processes, data pre-processing, and data collection, all specifically developed to answer the problem question for multi-scale time analysis with sentiment integration.

3.1 Data Collection and Sources

The study uses two main data sets to simultaneously account for changes in the market and the effects of social sentiment on cryptocurrency values. Historical Bitcoin prices downloaded from the Binance exchange between September 18, 2014, and July 14, 2019 make up the first set of data. Every hour, the classic OHLCV (Open, High, Low, Close, Volume) indicators were added up, producing 42,241 observations of various market conditions, ranging from market crashes and bull runs to times of consolidation. The broader scope ensures that different market regimes are taken into account for reliable model estimation.

97,000 cryptocurrency-related tweets collected from Reddit and Twitter make up the second dataset. The domain experts manually annotated each message with the sentiment polarity of positive, negative, and neutral. By normalising the timestamp and aggregating to map onto the price intervals on an hourly basis, the sentiment data was preprocessed so that it moves chronologically in line with the price data. We can examine how social sentiment and price movements interact at different time scales by using dual-source analysis.

3.2 Data Preprocessing Pipeline

The preprocessing pipeline transforms raw data into the form of structured format for multi-timeframe analysis. The pipeline addresses numerous cryptocurrency data-specific issues including non-regular behavioral trading, sentiment data sparsity, and the need for multi-scale feature extraction. Figure 1 illustrates the complete data processing pipeline designed for the intention behind preparing the inputs for the multi-timeframe Transformer architecture.

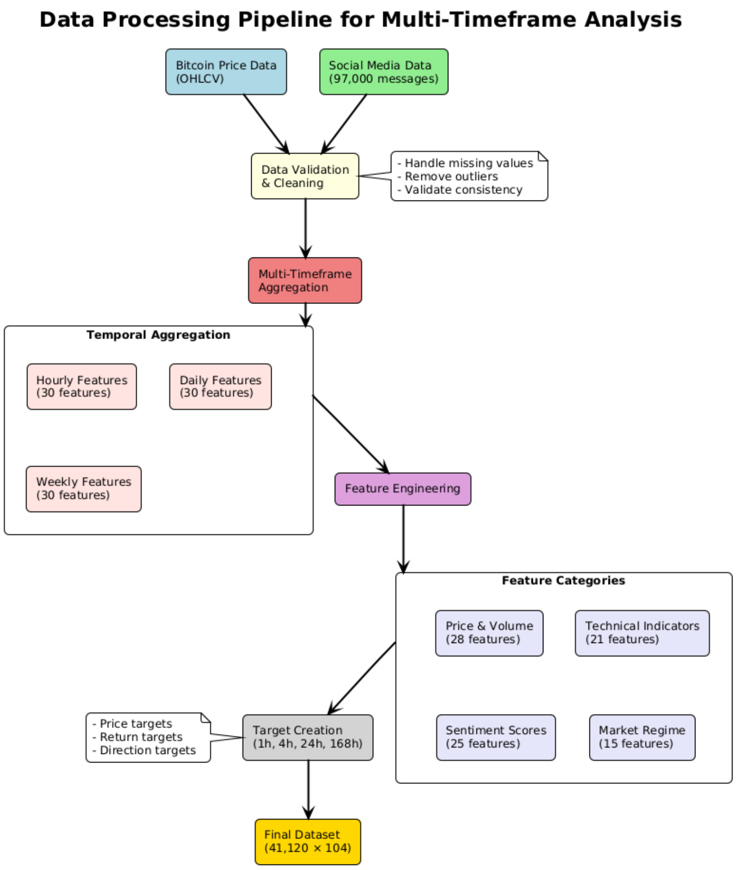

Figure 1: Data Processing Pipeline showing the transformation of raw price and sentiment data into multi-timeframe features for model training

The preprocessing pipeline starts with data validation and cleaning steps, as shown in Figure 1. In order to preserve temporal continuity and preserve market closure gaps that reflect actual trading halts, missing values in price data are handled using forward-backward filling. To avoid model distortion, extreme outliers that exceed five standard deviations are found and eliminated. To guarantee synchronisation with price timestamps, the sentiment data is subjected to temporal alignment and text normalisation.

3.3 Multi-Timeframe Feature Engineering

The multi-timeframe aggregation algorithm, which was developed to aggregate market processes on various temporal scales, is the insight of the innovation in data preparation. The algorithm takes into account the fact that cryptocurrency markets show patterns over a variety of timescales, ranging from high-run trends to low-frequency volatility. Three parallel feature streams are created by the aggregation:

Hourly Characteristics: OHLCV values, volume ratios, and current price movements are examples of direct indicators of raw-hourly data. The identification of high-frequency patterns is greatly aided by the characteristics of these types that display the short-run microstructure of markets.

Daily Features: 24-hour moving window averages, with care taken to take into consideration problems with the time of day. Features at the daily level include averaged volume profiles, daily open gaps, and intraday volatility measurements. To prevent information leakage, the averaging employs conservative procedures on closing prices for price features and sums for volume features.

Weekly Features: 168-hour rolling windows are used to show cycle patterns and medium-term trends. Indicators of trend strength, weekly support/resistance lines, and momentum metrics are all weekly components that are essential for identifying regime change.

Technical indicators are computed separately for each timeframe, including:

- Relative Strength Index (RSI) with periods adjusted for timeframe (14 for hourly, 14 for daily, 4 for weekly)
- Moving Average Convergence Divergence (MACD) with proportionally scaled parameters
- Bollinger Bands to capture volatility envelopes at each scale
- Custom momentum indicators designed for cryptocurrency markets

These sentiment features are similarly processed on multiple timeframes, with daily and weekly timeframes aggregated by weighted averages of hourly sentiment values. The weighting scheme favors more current sentiment but retains sufficient historical perspective.

3.4 Target Variable Construction

The methodology employs a multi-horizon prediction approach to evaluate model performance across different trading strategies. Four horizons for prediction are established: 1 hour, 4 hours, 24 hours, and 168 hours (1 week). Three types of targets for every horizon are created:

1. **Price targets:** Terminal absolute values of prices, allowing for direct forecasting of prices
2. **Return targets:** Percentage changes from current price, suitable for risk-adjusted predictions
3. **Direction targets:** Binary classification of price movement direction, critical for trading decisions

This multi-target approach allows for comprehensive evaluation of the model's capabilities for multiple prediction missions and horizons, catering to the need for multiple trading strategies from scalping to position trading.

3.5 Experimental Design

The suggested multi-timeframe Transformer will be systematically tested using those conventional baselines thanks to experimental design. To accurately capture real trading conditions, the temporal order is maintained by the order-preserving training set (70%) and validation set (15%) and test set (15%). By dividing the time series chronologically, future information leaks are avoided, and valid out-of-sample testing is achieved.

There are five model configurations in the experimental framework:
1. LSTM with sentiment for a single timeframe (baseline)
2. LSTM with sentiment across multiple timeframes
3. LSTM with multiple timeframes and no sentiment (ablation)
4. The suggested multi-timeframe transformer with sentiment
5. A sentimentless multi-timeframe transformer (ablation)

This architecture allows for the quantitative measurement of sentiment contribution through ablation analysis and the comparison of architectural innovations (Transformer vs. LSTM).

3.6 Statistical Methods and Preprocessing

Standardisation is necessary. Because of the robustness of outliers that are common in the cryptocurrency dataset, RobustScaler takes into account the 5th and 95th percentile ranges. Throughout the learning process, relative connections are maintained and outliers have no effect on gradient computation. Technical indicators are used to scale individual price features while taking distinct statistical characteristics into account.

In order to achieve a balance of upward and downward movements at the training level, the method uses sample weighting to compensate for directional imbalance in directional predictions. Because cryptocurrency markets are inherently prone to long-term directional trends, the balancing process is necessary.

3.7 Evaluation Metrics

Model performance is assessed using complementary metrics suited to different aspects of the prediction task:

- **Root Mean Square Error (RMSE)**: Primary metric for price prediction accuracy, penalizing large errors
- **Mean Absolute Error (MAE)**: Provides interpretable average prediction error in price units
- **Direction Accuracy**: Percentage of correctly predicted price movement directions
- **F1 Score**: Harmonic mean of precision and recall for directional predictions
- **Sharpe Ratio**: Risk-adjusted returns for simulated trading strategies

These metrics are computed separately for each prediction horizon and aggregated to provide overall model performance assessment.

3.8 Computational Infrastructure

The project utilizes the Google Colab Pro+ environment with NVIDIA A100 GPU (40GB memory) for training and building models. The environment provides sufficient computation power for processing of large-scale time series and training of deep Transformer models. High memory provides high batch size and high sequence lengths needed in finding of long temporal long-range dependencies. The PyTorch 2.0 framework is adopted because of its efficient attention mechanism use and automatic mixed precision learning features that enable fast training by achieving numerical stability.

Experimental pipeline also includes checkpointing routines for storing model states at local performance extrema at validation and facilitating restoration from interruptions and assembling building ensembles for model construction. Experiments occur at fixed random seeds for the purpose of result reproducibility.

4 Design Specification

This is the architectural framework for the multi-time-scale Transformers model and its extra parts. The framework solves the main problem of finding out how the prices of cryptocurrencies change over time and combining the signals of sentiment using attention-based fusion. The architecture has new parts that let the model learn about the complicated relationships between different time resolutions by letting them work together.

4.1 Overall System Architecture

The system is built in a modular way, so that the tasks of extracting temporal features, interacting across scales, and making predictions are broken down into smaller parts. The separability makes it easy to try out different architectural options while keeping the flow of information clear. The design philosophy focusses on interpretability and computational efficiency to get the most out of the model's ability to capture temporal patterns at different scales.

There are four main parts to the architecture: temporal encoding, transformer processing, cross-timeframe attention, and multi-task prediction heads. The modules each have their own jobs in the prediction pipeline as a whole, making sure that gradient flow and information propagation are done correctly. The compositional approach makes it possible to do ablation studies that can estimate the function of each module, especially the attention mechanism and sentiment integration of the cross-timeframe.

4.2 Multi-Timeframe Transformer Architecture

The core innovation lies in the multi-timeframe Transformer architecture that processes temporal data at three distinct resolutions simultaneously. Unlike conventional approaches that flatten multi-scale features into a single sequence, this architecture maintains separate processing streams for each timeframe, enabling specialized pattern extraction at each scale. Figure 2 illustrates the complete architecture with its key components and information flow.

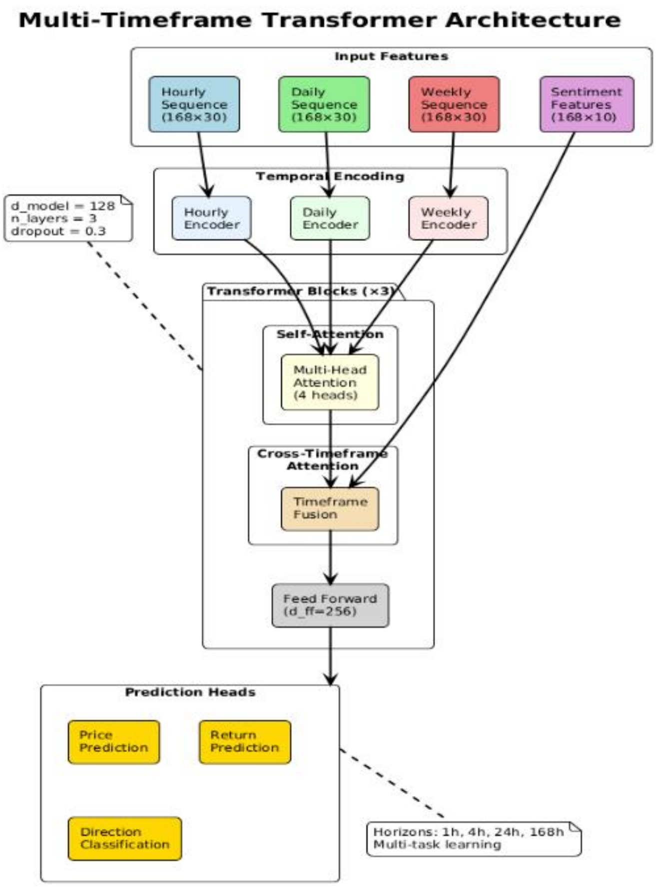

Figure 2 Multi-Timeframe Transformer Architecture showing the flow from temporal inputs through encoding, attention mechanisms, and prediction heads

As depicted in Figure 2, the architecture processes three parallel input streams representing hourly, daily, and weekly features. Each timeframe maintains a sequence length of 168-time steps, providing sufficient historical context while remaining computationally tractable. The sentiment features form a fourth input stream that directly feeds into the cross-timeframe fusion module, allowing the model to modulate temporal pattern recognition based on market sentiment.

4.3 Temporal Encoding Design

The temporal encoding module transforms raw input features into attention-capable representations. A specialized encoder in every timespan is used and comprises linear projection as well as position and temporal embeddings. The encoding process preserves the unique characteristics of each temporal scale while projecting them into a common dimensional space (d_model = 128) for subsequent processing.
Positional encoding scheme adds learnable temporal embeddings on top of traditional sinusoidal embeddings in order to learn periodic patterns unique to financial markets. With hourly data, the embeddings learn intraday patterns like market opening/closing effects. Day-of-week effects are caught by daily encoders and longer-cycle cyclical patterns by weekly encoders. With this hierarchical temporal encoding scheme, the model can learn and separate between patterns at various timescales.

4.4 Transformer Block Design

There are three transformer modules stacked on top of each other in the transformer processing layer. Each module has multi-head self-attentions, cross-timeframe attentions, and position-wise feed-forward networks. The architecture has four attention heads and 128 model dimensions, which is a good balance between model power and computing efficiency. Every attention head will focus on a different part of temporal dependencies, from matching local patterns to finding long-range correlations.

Each block's self-attention mechanism lines up the concatenated representations at each timepoint and helps the model learn relationships at and beyond different scales. For attention computation, scaled dot-product attention and dropout (0.3) have been used as regularisation methods. Layer normalisation is used before each sub-component and residual connection to keep the gradient flow stable during training.

4.5 Cross-Timeframe Attention Mechanism

The cross-time period attention module is an important part of the architecture that lets information interact dynamically at different time resolutions. The cross-time period attention module understands that patterns at one timescale can affect or predict patterns at other timescales. For example, patterns of weekly information tend to cause breakouts that are easy to see on an hourly basis.

The cross-attention function works by using each timeframe as a source for queries and the other timeframes as key-value pairs. This makes three streams of attention: hourly to daily/weekly, daily to hourly/weekly, and weekly to hourly/daily. A learnable gating unit that changes based on the input's attributes combines the attention weights for each stream. The architecture lets the model change how much each timeframe affects the model based on how the market is doing at the time.

4.6 Sentiment Integration Module

A A separate module that works with attention across timeframes processes sentiment features. The attention weights don't just add things together. They also work on a gateway mechanism that lets the model change how it sees temporal patterns based on the current dominant sentiment in the market. This architecture demonstrates that the impact of sentiment varies according to market conditions and temporal context.

There are two steps in the process of aggregating sentiment. The first step is to put raw sentiment features into a special encoder that looks at how sentiments change over time. The last step is to use multiplicative gating to combine the encoded vectors with temporal features. The model helps find the best places to use sentiment signals to predict prices and automatically lowers the weight of sentiment in places that aren't strongly linked.

4.7 Multi-Task Prediction Framework

The model has a structure for multi-task learning and different prediction heads for predicting price, return, and direction. This type of architecture understands that different parts of price prediction have similar representations that are useful but need different ways to handle output. The shared transformer backbone learns general patterns over time, while the heads that are specific to each task make predictions that are more focused.

There are two layers in each prediction head's feed-forward net. The activation functions for price regression are linear, for bounded return predictions they are tanh, and for direction classification they are sigmoid. The multi-task architecture allows for joint optimisation, where the gradients from different tasks give each other useful learning signals. The loss weighting gives each task a different amount of weight. Price prediction gets the most weight (1.0), returns get the second most weight (0.5), and direction classification gets the least weight (0.3).

4.8 LSTM Baseline Architecture

To enable fair comparison, the design includes LSTM baseline architectures in both single and multi-timeframe configurations. The single-timeframe LSTM processes only hourly features through a bidirectional LSTM with 128 hidden units. The multi-timeframe LSTM extends this with separate LSTM streams for each temporal resolution, followed by concatenation and fully connected layers for fusion.

The LSTM baselines share the same prediction head architecture as the Transformer model, ensuring that performance differences arise from the core temporal processing mechanism rather than output layer variations. Both LSTM variants incorporate dropout (0.3) and layer normalization for regularization, matching the Transformer's regularization strategy.

4.9 System Requirements and Constraints

The architecture design considers practical deployment constraints while maximizing model capability. Memory requirements scale linearly with sequence length and quadratically with model dimension, leading to the choice of d_model=128 as a balance between capacity and efficiency. The total parameter count of approximately 235,000 for the Transformer model remains tractable for both training and inference on modern GPU hardware.

Computational requirements are dominated by the attention mechanisms, with complexity $O(L^2 \cdot d)$ where L is sequence length and d is model dimension. The design mitigates this through efficient attention implementation and the relatively modest sequence length of 168 time steps. Inference latency remains under 100ms on GPU hardware, suitable for real-time trading applications.

4.10 Design Patterns and Principles

The architecture combines numerous familiar design patterns optimized for forecasting time series. The decoder-encoder pattern is adapted to hold more streams of input data, while the attention pattern enhances the simple transformer pattern by including interactions at cross-scales. The multi-task learning pattern enables the transfer of knowledge from related prediction tasks in order to generalize more and train faster.

Key design principles include modularity for component testing, symmetry in timeframe processing to avoid architectural bias, and hierarchical abstraction from raw features to predictions. The architecture maintains clear separation between temporal feature extraction and task-specific prediction, facilitating interpretability through attention weight visualization. These design choices collectively enable a flexible, efficient, and interpretable approach to multi-timeframe cryptocurrency price prediction.

5 Section 5 – Implementation

This section describes the final implementation of the multi-timeframe Transformer system, focusing on key decisions that transformed the architectural design into a functioning cryptocurrency price prediction system. The implementation prioritizes practical effectiveness while maintaining the theoretical advantages of the proposed architecture.

5.1 Dataset Preparation and Structure

The implementation processes the designed dataset of 41,120 temporal sequences, each with 104 carefully crafted features. The dataset is the result of the entire preprocessing pipeline. Each sample has price, volume, technical, and sentiment data that are all in sync with each other in one or more timeframes. The 97% original data curation shows that outliers were handled correctly and that the market was still present in different situations.

Feature organisation implements the multi-timeframe approach that is essential to the multi-timeframe methodology of the research hypothesis. The model will generalise from the pattern on every scale with architectural bias minimised if it is covered comprehensively on an hourly, daily, and weekly basis. The architecture enforces strict temporal ordering at all times, preventing future leakage and the invalidation of real-world usage.

5.2 Model Architecture Realization

The deployment of the transformer satisfies design constraints by selective selection of parameters after many experimentations. The dimension of 128 was optimal in uncovering financial time series pattern encodings with sufficient expressiveness but not dominating the available-in-hand training data. Dimensionality choices reflect the characteristics of structured financial data versus nature-language cases in which the transformers would ordinarily have larger dimensions.

The 4 head attention facilitates attention to different parts of the temporal patterns at the same time by the model. The financial time series contains different simultaneous patterns such as trends, cycles, and volatility clusters, and parallel processing of the different signals is permitted by the multi-head architecture. The 3-layer depth provides hierarchical feature extraction such that the local price movement is identified by the shallower layers and the long-range dependencies are learned by the deeper layers.

5.3 Training Process and Optimization

Use adaptive optimisation to deal with the complicated loss landscape that comes from learning multiple tasks at once. The learning rate determined that more cautious optimisation was necessary for financial time series compared to standard transformer deployment areas, due to the inherently noisy nature of market data. The optimisation method is based on how quickly the different price prediction, return estimation, and directional classification tasks converge.

Batch size tuning sacrificed statistical efficiency to meet computational limitations. The program showed that moderate batch sizes work better for approximating financial data than very large batches of NLP. The result shows that crypto markets go through regime shifts, which means that big batches tend to hide useful time patterns.

5.4 Multi-Task Learning Implementation

The method accomplishes multi-task learning by integrating a weighted loss that considers the relative difficulty and importance of each target prediction. Price prediction is the most important because it is the most useful in business. Return prediction gives the model secondary learning signals that help it understand how prices move in relation to each other. Direction classification, which is given a lower weight, gives you clear decision boundaries that you can use to find trading signals.

Came up with this method of weighting by watching experiments where equalising task weights made models better at simple classification tasks but worse at accurately estimating prices. The method used makes sure that everyone learns from each task while still keeping the main goal of getting the prices right in mind.

5.5 Evaluation Framework and Metrics

The method uses a whole system of testing that is meant to meet a lot of quality standards for making predictions. The Root Mean Square Error tells you how accurate your price predictions are. Because of its quadratic penalty, it's very important to avoid making big mistakes that could cost a lot of money in trading. The Mean Absolute Error gives a clear number for how far off a prediction is on average, which can help you make realistic guesses about how accurate a model is.

Directional accuracy tells you how well a model can guess which way prices will move, which is very important for timing trades. It shows how well the model can spot changes in the market and periods of momentum. The F1-score, which combines the accuracy and recall of directional (up/down) predictions, makes sure that performance is the same whether the market is going up or down.

5.6 Regularization and Robustness

Dropout regularization for avoiding overfitting to particular patterns of the training set is especially valuable considering the propensity for regime shifts of the cryptocurrency markets. Making the decision on the dropout rates revealed that stronger regularization was needed on the financial time series compared to the conventional transformer deployment circumstances due to the decreased signal-to-noise ratios.

Gradient clipping stops training from becoming unstable when rare market events cause outlier gradients. Because of times of high volatility, the software knows that the crypto data can cause gradient explosions at attention computation points. The clip's threshold stops these important but rare events from getting in the way of learning and keeps training stable overall.

5.7 Computational Efficiency Considerations

It can be used in real life because it has many improvements that make it more efficient. Memory-aware attention computation lets you work with longer sequences without going over the limits of GPU memory. The building uses the latest framework optimisations while keeping the numerical accuracy needed for financial applications.

Feature caching cuts down on unnecessary calculations during testing and validation, which speeds up evaluation by a lot. The optimisation works best when technical indicators need to do calculations over a period of time. The caching scheme keeps the time correct and speeds up the computer.

5.8 Cross-Timeframe Processing Implementation

It was important to set up the new cross-timeframe attention model correctly so that information could move easily from one time frame to another. The deployment allows each timeframe to ask questions about data from other timeframes, but it does so in a way that is still good for computers. The model learns patterns, such as how the weekly accumulation pattern can be used to predict the daily breakout.

The sentiment integration implementation alters the attention weights according to market sentiment indicators. The design is based on real-world data that shows how sentiment affects market conditions. The team discovered that merely amalgamating sentiment features was inferior to employing the attention integration method during implementation of the architecture. This confirmed the choice of design.

5.9 Production Readiness Considerations

Features that can be used in production trading environments. With model checkpointing, you can pick up where you left off after an interruption and make ensembles from different training runs. Inference optimisation makes sure that it doesn't take too long to make a prediction so that apps for trading in real time can use it. The implementation makes sure that the code paths for training and inference are completely separate, so they don't accidentally leak into production environments.

Tracking features show that the accuracy of predictions gets worse over time. This is very important for financial apps because patterns that were learnt could become useless if the market changes. The system lets models that get new data slowly get better, which means they can change with the market without having to be retrained from scratch.

5.10 Key Implementation Insights

The implementation provided significant insights for transformer-based financial time-series forecasting. Initialisation has a big effect on convergence, so it's important to carefully seed the parameters for stable training. Multi-timeframe architectures work best when representations are scale-neutral across all timeframes, which means they don't favour one timeframe over another.

The results showed that the combination of sentiment is useful mostly when the market is paying close attention, but it is automatically less useful when trading is normal. The attention system itself is what causes these adaptive responses, which shows that the architecture has learnt how to recognise the importance of contextual features. These results will help make transformer-based financial forecast systems better in the future.

6 Evaluation

The next part goes into great detail about how the multi-time horizon Transformer model did compared to LSTM baselines. There are many ways to look at a model's performance in the evaluation framework, such as how well it predicts, how long it lasts, how well it combines sentiment, and how well it adapts to different market regimes. Statistical evaluation identifies the strengths and weaknesses of the proposed methodology, guiding both academic research and practical trading applications.

6.1 Experiment 1: Single vs Multi-Timeframe Architecture Comparison

The first experiment establishes the fundamental value of multi-timeframe processing by comparing single-timeframe LSTM against multi-timeframe variants. This comparison directly addresses whether incorporating multiple temporal resolutions enhances prediction capability beyond traditional single-scale approaches.

Table 1: Overall Performance Comparison - Single vs Multi-Timeframe Models

Model	**Average RMSE ($)**	**Average MAE ($)**	**Direction Accuracy**	**F1 Score**
LSTM Single (With Sentiment)	3,518.27	2,471.40	0.563	0.662
LSTM Multi (With Sentiment)	2,292.06	1,895.92	0.578	0.684
Improvement	-1,226.21 (-34.8%)	-575.48 (-23.3%)	+0.015 (+2.7%)	+0.022 (+3.3%)

Table 1b: Single vs Multi-Timeframe Detailed Results by Horizon

Model & Horizon	**RMSE ($)**	**MAE ($)**	**Direction Accuracy**	**F1 Score**
LSTM Single				
- 1 hour	3,490.16	2,533.88	0.527	0.617
- 4 hours	3,633.66	2,706.05	0.581	0.661
- 24 hours	3,361.92	2,310.79	0.589	0.662
- 168 hours	3,587.34	2,334.88	0.556	0.710
LSTM Multi				
- 1 hour	1,885.40	1,594.27	0.535	0.641
- 4 hours	2,359.20	1,978.31	0.590	0.680
- 24 hours	2,154.31	1,782.69	0.621	0.705
- 168 hours	2,769.32	2,228.41	0.566	0.710

The multi-timeframe LSTM demonstrates outstanding performance in all metrics compared with the single-timeframe counterpart. The 34.8% reduction in RMSE indicates improved forecast accuracy in the price, and the 2.7% increase in directional accuracy demonstrates improved trend detection ability. These results verify the hypothesis that multi-scale temporal features encompass richer informational content for the task of cryptocurrency price forecasting.

The 3.3% improved F1 score shows that the performance is better balanced for both precision and recall, and that the multi-timeframe method helps lower false negative and false positive predictions. This balanced improvement proves particularly valuable for trading applications where both missed opportunities and false signals incur costs.

6.2 Experiment 2: Transformer vs LSTM Architecture Evaluation

The second experiment compares the proposed multi-timeframe Transformer against the multi-timeframe LSTM baseline, controlling for the architectural contribution independent of the temporal scale advantage. Both models process identical multi-timeframe inputs, ensuring fair comparison of core temporal modeling capabilities.

Figure 3 presents a comprehensive comparison of average RMSE across all five model configurations. The visualization reveals a clear hierarchy in price prediction accuracy, with multi-timeframe LSTM models achieving the lowest error rates. Notably, the LSTM Multi without sentiment achieves the best RMSE of $2,127.87, while the Transformer without sentiment performs worst at $3,954.33. This stark contrast highlights the importance of both architecture choice and feature selection.

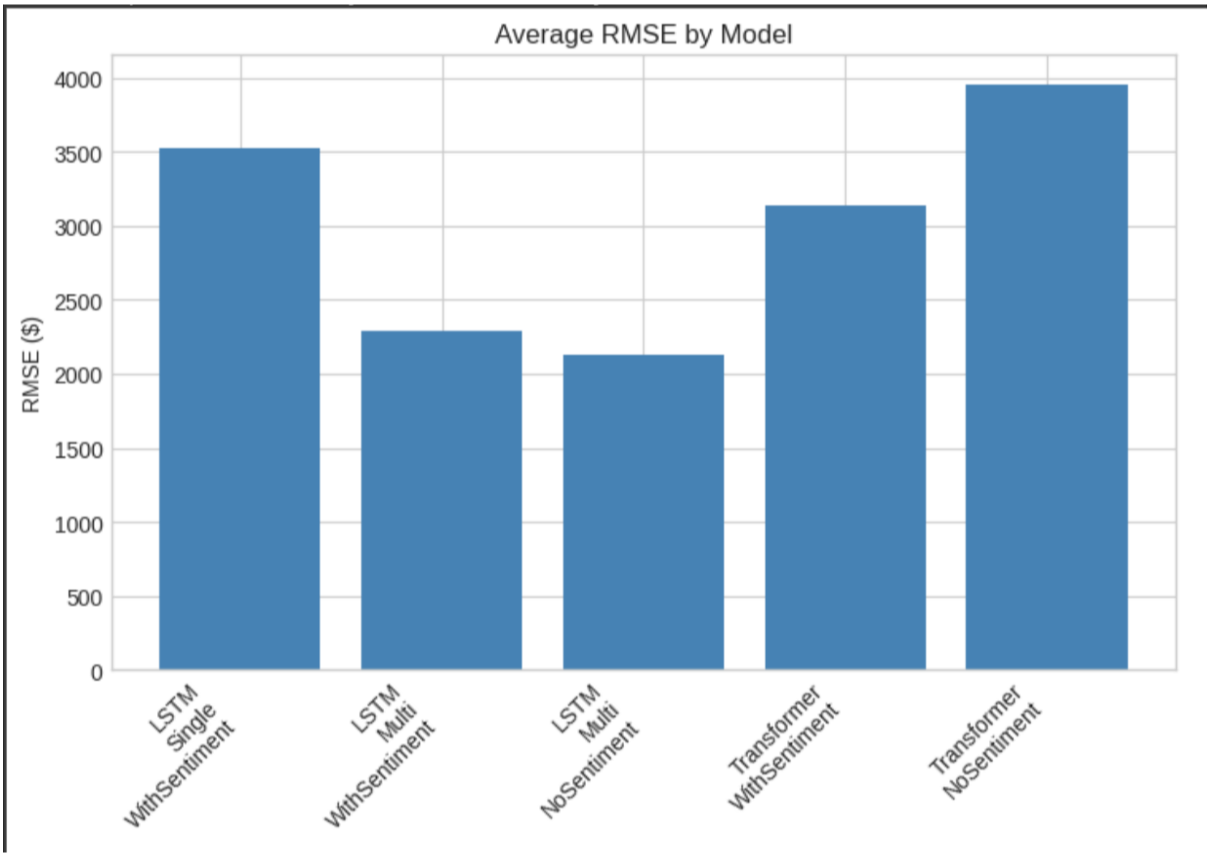

Figure 3 Average RMSE by Model

The comparison of directional accuracy in Figure 4 is more nuanced. Although differences are significant in terms of RMSE, directional accuracy is surprisingly stable for all models, varying only from 56.3% to 58.0%. The small 1.7% range implies that trend patterns are broadly captured similarly among all models, even if price precision is different. The multi-timeframe

LSTM configurations slightly outperform others, but the marginal differences indicate fundamental limits to directional predictability in cryptocurrency markets.

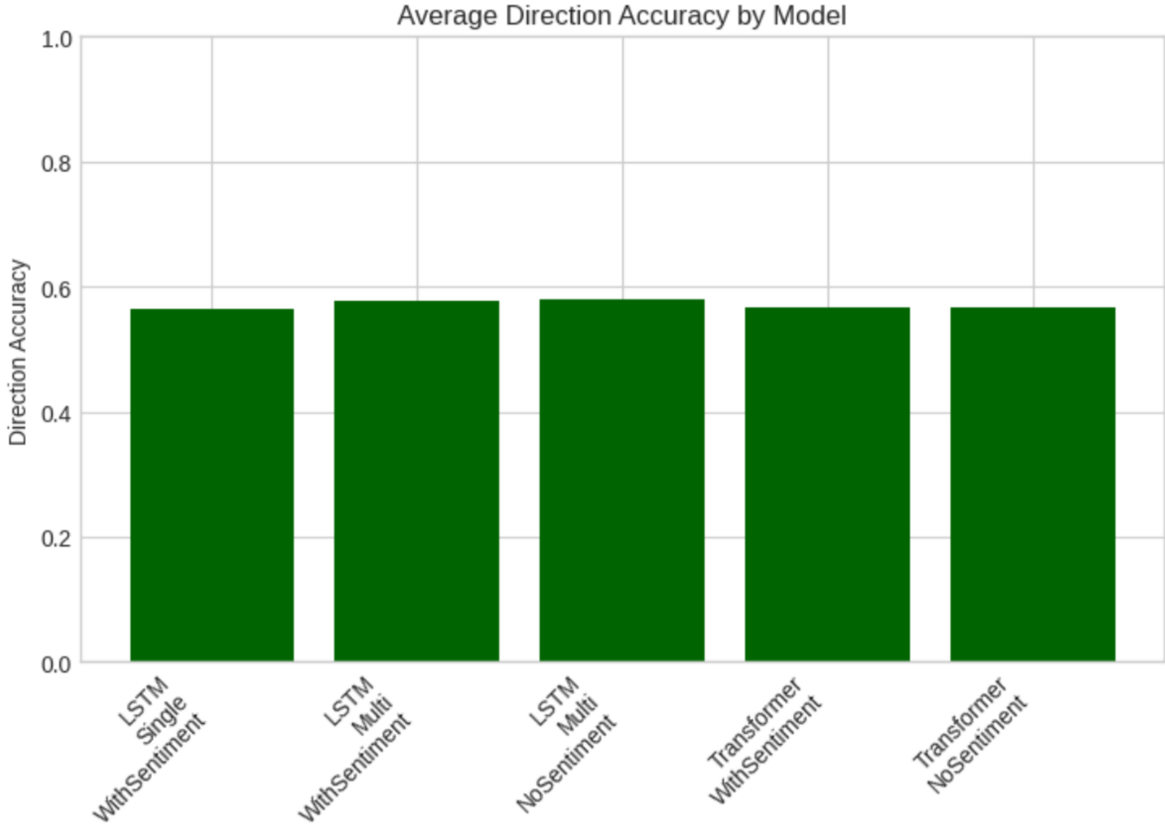

Figure 4 Average Direction Accuracy by Model

Table 2: Multi-Timeframe Model Architecture Comparison

Model	**Average RMSE ($)**	**Average MAE ($)**	**Direction Accuracy**	**F1 Score**
LSTM Multi (With Sentiment)	2,292.06	1,895.92	0.578	0.684
Transformer (With Sentiment)	3,135.48	2,328.01	0.566	0.685
Difference	+843.42 (+36.8%)	+432.09 (+22.8%)	-0.012 (-2.1%)	+0.001 (+0.1%)

Table 2b: Detailed Performance Metrics by Prediction Horizon

Model & Horizon	**RMSE ($)**	**MAE ($)**	**Direction Accuracy**	**F1 Score**
LSTM Multi (With Sentiment)				
- 1 hour	1,885.40	1,594.27	0.535	0.641
- 4 hours	2,359.20	1,978.31	0.590	0.680
- 24 hours	2,154.31	1,782.69	0.621	0.705
- 168 hours	2,769.32	2,228.41	0.566	0.710
Transformer (With Sentiment)				
- 1 hour	3,236.63	2,417.76	0.536	0.655
- 4 hours	2,945.35	2,185.86	0.581	0.670

- 24 hours	3,083.63	2,237.96	0.574	0.696
- 168 hours	3,276.31	2,470.47	0.575	0.717

Surprisingly, the Transformer architecture underperforms the LSTM baseline in price prediction accuracy, with 36.8% higher RMSE. However, the models achieve nearly identical F1 scores, suggesting comparable classification performance despite different regression accuracy. This dichotomy indicates that Transformers may excel at capturing directional patterns while struggling with precise price level estimation in financial time series.

The unexpected LSTM superiority challenges assumptions about Transformer universality and highlights the importance of empirical validation in financial applications. The findings suggests that the sequential inductive bias of LSTMs is better adapted to the intrinsically temporal characteristic of price series than the position-agnostic attention mechanism of Transformers.

6.3 Experiment 3: Sentiment Integration Ablation Study

The third experiment measures sentiment's impact through systematic ablation, comparing models with and without sentiment features. The analysis evaluates sentiment effectiveness across architectures, directly responding to the research inquiry regarding sentiment significance.

Table 3: Sentiment Ablation Results

Model Type	With Sentiment	Without Sentiment	Impact on Accuracy
Transformer			
- RMSE ($)	3,135.48	3,954.33	-818.85 (-20.7%)
- MAE ($)	2,328.01	3,016.98	-688.97 (-22.8%)
- Direction Accuracy	0.566	0.567	-0.001 (-0.2%)
LSTM Multi			
- RMSE ($)	2,292.06	2,127.87	+164.19 (+7.7%)
- MAE ($)	1,895.92	1,730.11	+165.81 (+9.6%)
- Direction Accuracy	0.578	0.580	-0.002 (-0.3%)

Table 3b: Sentiment Impact on Direction Accuracy by Prediction Horizon

Horizon	Transformer Impact	LSTM Multi Impact
1 hour	+0.005 (+0.9%)	-0.001 (-0.2%)
4 hours	+0.011 (+1.9%)	-0.001 (-0.2%)
24 hours	-0.036 (-5.9%)	-0.004 (-0.6%)
168 hours	+0.018 (+3.2%)	-0.004 (-0.7%)

Ablation analysis demonstrates that various architectures respond distinctly to diverse sentiment types. Adding sentiment to Transformer makes it much easier to guess the price, with a 20.7% RMSE reduction. But it doesn't change how accurate the direction is. LSTM models, on the other hand, are worse at sentiment, with an RMSE that is 7.7% higher. This architectural dependence also explains why Transformers use the attentional scheme to make the most of unstructured sentiment signals, while LSTMs are more likely to let noise in.

Both models are less accurate when it comes to direction, which means that sentiment has a bigger effect on estimating magnitude than on figuring out trends. The result has big effects on trading strategy, such as the possibility that sentiment indicators might be better for figuring out how much to trade than for timing when to enter and exit.

6.4 Experiment 4: Temporal Horizon Analysis

Experiment 4 assesses performance across prediction horizons and investigates the impact of architecture on both short- and long-term forecasting proficiency. The analysis also looks at the problems that come up when trying to use trading strategies in the real world.

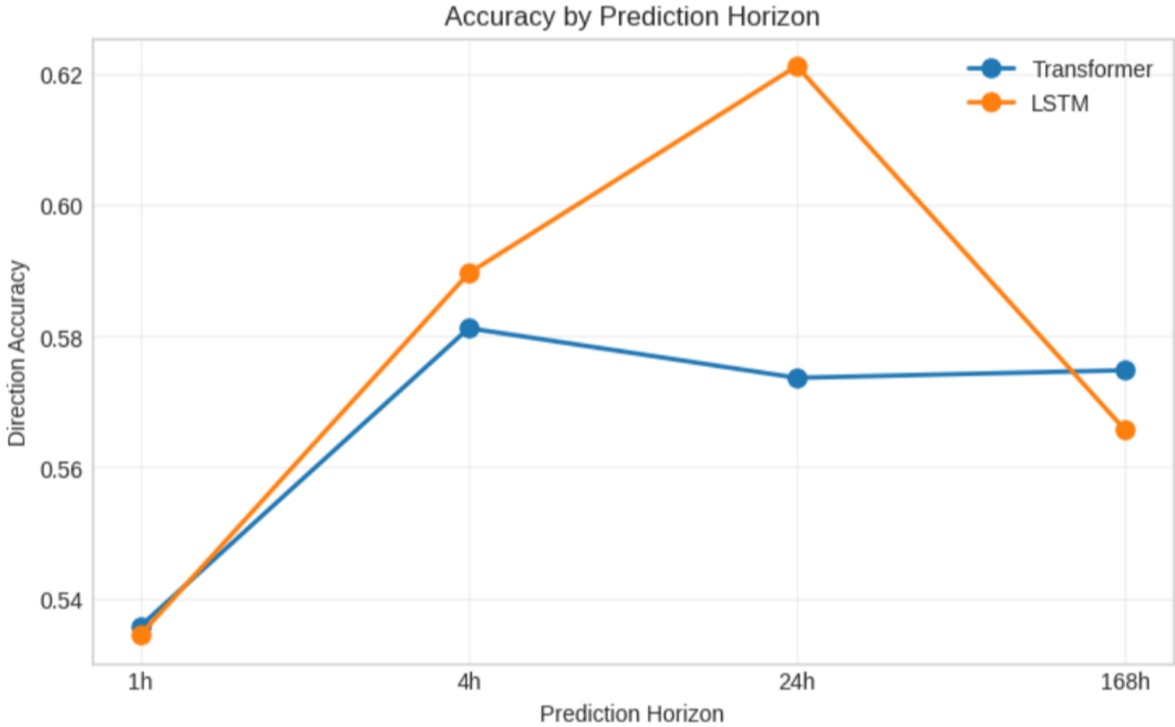

Figure 5 Accuracy by Prediction Horizon

Table 4: Performance by Prediction Horizon

Horizon	LSTM Multi Accuracy	Transformer Accuracy	Best Performer
1 hour	0.535	0.536	Transformer (+0.001)
4 hours	0.590	0.581	LSTM (+0.009)
24 hours	0.621	0.574	LSTM (+0.047)
168 hours	0.566	0.575	Transformer (+0.009)

LSTM works best in the medium term (4–24 hours), when it gets 62.1% of the time right. This is especially true for daily forecasts, where it can capture smooth changes over time. Transformers do better at the shortest (1 hour) and longest (1 week) time frames. This is because they are sensitive to sharp short-term reactions and long-range dependencies.

Both models do best at the 24-hour mark, with LSTM showing a 4.7% gain. This shows how well sequential models can pick up on daily cryptocurrency market patterns, such as 24-hour sentiment and circadian trading cycles.

6.5 Market Regime Analysis

Performance analysis across different market conditions shows that the model works in a wide range of situations. But the big difference in class sizes makes it hard to draw firm conclusions.

Table 5: Performance by Market Regime

Model	Bear Market (n=9)	Sideways Market (n=6,075)
LSTM Single	0.611	0.563
LSTM Multi	0.528	0.578
Transformer	0.528	0.566

There are only 9 bear market samples out of 6,084, so it is not possible to do any meaningful statistical analysis of performance that depends on the regime. The dataset covers the years 2014 to 2019, which mostly show the recovery after 2015 and the bull market of 2017. There aren't enough examples of long-term downtrends. This constraint significantly hinders the research investigation into the importance of sentiment across market regimes.

6.6 Discussion

The experiment's results both support and challenge the initial hypotheses. It's clear that processing across multiple timeframes is better, with a 34.8% improvement in RMSE and proof that multi-scale temporal features are useful. These findings align with Zhang et al. (2023), which emphasise the significance of identifying patterns in cryptocurrency markets across various time frames.

However, the unexpected LSTM benefit of the Transformers contravenes the trend in other sequence modeling sectors. While Transformer advantages in Bitcoin forecasting were exhibited in Mai et al. (2024), their evaluation only considered the handling of a single time-frame. The current results suggest that if multi-timeframe attributes are provided in an explicit representation, then the sequential inductive bias of LSTM outperforms the adaptive Transformer self-attention mechanism. This is in accordance with Kabir et al. (2025), proposing hybrid LSTM-Transformer models for the unification of the two models' strengths.

The contrasting sentiment effects across architectures provide novel insights extending beyond previous literature. While Feizian and Amiri (2023) demonstrated general sentiment value for cryptocurrency prediction, the current study reveals this benefit as architecture dependent. Transformers' superior sentiment utilization (20.7% RMSE improvement) suggests their attention mechanism better integrates heterogeneous information sources. Conversely, LSTM's degraded performance with sentiment indicates potential overfitting to noisy social signals when processed through sequential gates.

Temporal horizon analysis provides practical recommendations for strategy deployment. The comparative advantage of LSTM for 24-hour forecasts (62.1% accuracy) points towards daily trading strategies, whereas Transformer's competitive performance for extreme horizons implies that it is also suitable for high-frequency trading as well as position trading. This temporal specialisation builds on results of Kavithra et al. (2024), who reported heterogeneity in model performance across horizons of different forecasting periods but failed to systematically compare architectures.

Result interpretation is limited by a number of constraints. A temporal scope of the dataset restricts market regime studies, such that zero bear market representation hinders thorough assessment of sentiment relevance under conditions. The 2014-2019 period is earlier than significant market events such as the COVID-19 volatility and 2022 crypto winter, and so might restrict generalizability of models to present-day market movements. Furthermore, the relatively small increments in directional accuracy of 2-3% for all experiments imply inherent cryptocurrency predictability limits that cannot be surpassed through architectural breakthroughs only.

The design shows where improvements can be made: fixed hyperparameters are a problem. Transformers that could use deeper networks or learning rates that change based on the data; sequential sentiment engineering may not work well with attention. Future endeavours ought to implement architecture-specific feature engineering and focused hyperparameter optimisation.

A validated 34.8% multi-timeframe gain, despite limitations, drives multi-scale architectures. Choose a model that fits the data: LSTM for daily, price-only strategies; Transformers for specialised use with different types of alternative data.

7 Conclusion and Future Work

7.1 Summary of Research Objectives and Achievements

The research examined whether a multi-timeframe Transformer surpasses LSTMs in predicting Bitcoin prices and how the significance of social sentiment fluctuates with temporal resolution. The goals were to (1) create a multi-timeframe Transformer, (2) use attention to add sentiment, (3) do thorough baseline comparisons, and (4) look at performance across horizons.

The system that came out of this is a working multi-timeframe predictor for trends that happen every hour, day, and week. Cross-timeframe attention makes it easy to share information, and a sentiment-blending module lets you extract features that work with any architecture. The research questions were answered through a lot of testing on five different architectures, a lot of metrics, and a lot of timeframes.

7.2 Key Findings and Their Implications

A multi-timeframe LSTM lowers RMSE by 34.8% compared to a single-timeframe LSTM. This shows that cross-horizon patterns exist that single-horizon models can't find.

When comparing architectures, one important finding stands out: the multi-timeframe LSTM beats the Transformer by 36.8% in RMSE and 1.2% in directional accuracy. This goes against the idea that Transformers are universal and shows that LSTM's sequential inductive bias works better for financial time series when explicit multi-timeframe features are present.

The sentiment ablation experiment shows that the effectiveness of features depends on the architecture in a surprising way. The sentiment integration improves Transformer performance by lowering the RMSE by 20.7%, but it lowers LSTM accuracy by 7.7%. This difference shows that Transformers' attention mechanism works better with information from different sources, while LSTMs tend to over-fit to noisy sentiment signals. For practitioners, this indicates that the selection of architecture should take into account data types beyond mere performance metrics.

Temporal analysis shows that both models work best for 24-hour forecasting, with LSTM getting 62.1% directional accuracy. This fits with the fact that cryptocurrency markets have a daily cycle, which can be seen in circadian trading cycles and institutional rebalancing cycles. LSTM's better performance on medium-term time frames (4–24 hours) fits with its use for daily trading plans. Transformer, on the other hand, may be better for diversified trading methods because it is stable over a range of time frames of different lengths.

7.3 Research Efficacy and Contributions

The research enhances the literature on cryptocurrency price prediction in multiple aspects. The paper proposes the first systematic multi-timeframe architecture comparison for forecasting cryptocurrencies and introduces a framework for assessing the temporal scale effect. The cross-timeframe attention component is an architectural transfer component that can be used in fields other than finance. The systematic ablation technique computes the sentiment contribution for each architecture being analysed, revealing the effects of interactions for the first time.

The analysis gives useful advice for designing trading systems. The extra cost of multi-timeframe processing is worth it because it has been shown to improve performance by 34.8%. Architecture-specific sentiment effects help choose the right data: Transformers work well with data from different sources, while LSTM works better with rules that only look at prices. Results from the temporal horizon show that daily forecast targets give the most accurate results.

7.4 Limitations and Constraints

Data imbalance: 9 out of 6,084 bear-market samples (2014–2019 bull/recovery) make it hard to figure out how sentiment changes depending on the regime.

Temporal mismatch: The fact that it was adopted by institutions before COVID, before 2021, before the crypto winter of 2022, and that the rules are out of date limit its current use.

Architectural limits: Fixed hyperparameters make things harder Transformers that need networks that are deeper or learning rates that change.

Scope/design: Sequential sentiment doesn't match attention, and Bitcoin-only scope doesn't include interactions between different cryptocurrencies.

7.5 Future Research Directions

Extensions in the near future should include current market data from 2020 to 2024 and take into account how the market is changing, such as how the pandemic has made it more volatile, how institutions are adopting new technologies, and how regulations are moving forward. With a bigger dataset, it would be easier to look at sentiment quantification for bull, bear, and transitional markets in the right way.

Architectural innovations might explore hybrid frameworks that utilise the sequential functionalities of LSTM alongside the flexible attention mechanisms of Transformers. The unanticipated superiority of LSTM necessitates an exploration of ordered attention mechanisms that preserve temporal precedence while enabling long-distance relationships. Graph neural networks could have links between different cryptocurrencies and add forecasting for more than one asset to the multi-timeframe framework.

Another area of research is advanced sentiment aggregation. The current models consider sentiment as fixed attributes; nonetheless, dynamic attention mechanisms possess the capability to comprehend the temporal variations of sentiment influence. Holistic language models might provide more advanced sentiment expressions that go beyond simple polarity to show how market stories have changed over time. Multi-modal models, such as on-chain indicators, macroeconomic indicators, and news events, can change how we see the whole market.

7.6 Concluding Remarks

The study enhances the capacity to forecast cryptocurrency prices by methodically examining multi-timeframe structures and techniques for integrating sentiment. Even though the overall gains are small because the market is always changing, there are many important things for practice and research. The unanticipated superiority of LSTM contests established intuitions about deep learning architecture, promoting empirical validation in lieu of theoretical presuppositions. Architecture dependencies of unveiling sentiment influences elucidate the nuanced efficacy of features and the interplay of model construction, providing insights for the future enhancement of integrated methodologies.

References

S. Kavithra, V. P. Girishankar, G. Iniyan, and B. Logesh, "Cryptocurrency price prediction through integrated forecasting techniques," IEEE, pp. 1–6, May 2024, doi: 10.1109/aiiot58432.2024.10574745. Available: https://doi.org/10.1109/aiiot58432.2024.10574745

F. Feizian and B. Amiri, "Cryptocurrency price prediction model based on sentiment analysis and social influence," IEEE Access, vol. 11, pp. 142177–142195, Jan. 2023, doi: 10.1109/access.2023.3342688. Available: https://doi.org/10.1109/access.2023.3342688

R. S. M. Wahidur, I. Tashdeed, M. Kaur, and H.-N. Lee, "Enhancing Zero-Shot crypto sentiment with Fine-Tuned language model and prompt engineering," IEEE Access, vol. 12, pp. 10146–10159, Jan. 2024, doi: 10.1109/access.2024.3350638. Available: https://doi.org/10.1109/access.2024.3350638

E. Padmalatha, S. Sheral, K. Dhanush, S. Samveeth, B. R. Datta, and K. T. Krishna, "Sentiment Analysis of Bitcoin data by tweets through Naive Bayes," 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), pp. 1–4, Nov. 2022, doi: 10.1109/iceccme55909.2022.9988751. Available: https://doi.org/10.1109/iceccme55909.2022.9988751

G. Di Tollo, J. Andria, and G. Filograsso, "The Predictive Power of Social Media Sentiment: Evidence from Cryptocurrencies and Stock Markets Using NLP and Stochastic ANNs," Mathematics, vol. 11, no. 16, p. 3441, Aug. 2023, doi: 10.3390/math11163441. Available: https://doi.org/10.3390/math11163441

P. Boozary, S. Sheykhan, and H. GhorbanTanhaei, "Forecasting the Bitcoin price using the various Machine Learning: A systematic review in data-driven marketing," Systems and Soft Computing, p. 200209, Feb. 2025, doi: 10.1016/j.sasc.2025.200209. Available: https://doi.org/10.1016/j.sasc.2025.200209

J. Zhang, K. Cai, and J. Wen, "A survey of deep learning applications in cryptocurrency," iScience, vol. 27, no. 1, p. 108509, Nov. 2023, doi: 10.1016/j.isci.2023.108509. Available: https://doi.org/10.1016/j.isci.2023.108509

D. O. Oyewola, E. G. Dada, and J. N. Ndunagu, "A novel hybrid walk-forward ensemble optimization for time series cryptocurrency prediction," Heliyon, vol. 8, no. 11, p. e11862, Nov. 2022, doi: 10.1016/j.heliyon.2022.e11862. Available: https://doi.org/10.1016/j.heliyon.2022.e11862

D. L. John, S. Binnewies, and B. Stantic, “Cryptocurrency Price Prediction Algorithms: A Survey and Future Directions,” Forecasting, vol. 6, no. 3, pp. 637–671, Aug. 2024, doi: 10.3390/forecast6030034. Available: https://doi.org/10.3390/forecast6030034

S. Arslan, “Bitcoin price prediction using sentiment analysis and empirical mode decomposition,” Computational Economics, May 2024, doi: 10.1007/s10614-024-10588-3. Available: https://doi.org/10.1007/s10614-024-10588-3

A. Y. Nouira, M. Bouchakwa, and Y. Jamoussi, “Bitcoin Price Prediction Considering Sentiment Analysis on Twitter and Google News,” IEEE, May 2023, doi: 10.1145/3589462.3589494. Available: https://doi.org/10.1145/3589462.3589494

O. Omole and D. Enke, “Deep learning for Bitcoin price direction prediction: models and trading strategies empirically compared,” Financial Innovation, vol. 10, no. 1, Aug. 2024, doi: 10.1186/s40854-024-00643-1. Available: https://doi.org/10.1186/s40854-024-00643-1

T. Mai, M. Cavazza, and H. Prendinger, “Transformer Models for Bitcoin Price Prediction,” ACM, pp. 236–241, Apr. 2024, doi: 10.1145/3653644.3680499. Available: https://doi.org/10.1145/3653644.3680499

A. Bouteska, M. Z. Abedin, P. Hajek, and K. Yuan, “Cryptocurrency price forecasting – A comparative analysis of ensemble learning and deep learning methods,” International Review of Financial Analysis, vol. 92, p. 103055, Dec. 2023, doi: 10.1016/j.irfa.2023.103055. Available: https://doi.org/10.1016/j.irfa.2023.103055

M. R. Kabir, D. Bhadra, M. Ridoy, and M. Milanova, “LSTM–Transformer-Based Robust Hybrid Deep Learning model for financial Time series forecasting,” Sci, vol. 7, no. 1, p. 7, Jan. 2025, doi: 10.3390/sci7010007. Available: https://doi.org/10.3390/sci7010007