I.J. Information Engineering and Electronic Business, **2025, 5, 1-17**
Published Online on October 8, 2025 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijieeb.2025.05.01

A Transfer Learning–Enhanced Hybrid Deep Learning Framework for Bitcoin Price Forecasting Using Market Sentiment and Time Series Data

Rachid Bourday*
Computer Science Research Laboratory, Faculty of Sciences, IBN Tofail University, Kenitra, Morocco
Email: Rachid.bourday@uit.ac.ma
ORCID iD: https://orcid.org/0009-0003-2602-7180
*Corresponding Author

Issam Aattouchi
Computer Science Research Laboratory, Faculty of Sciences, IBN Tofail University, Kenitra, Morocco
Email: aattouchi.issam@uit.ac.ma
ORCID iD: https://orcid.org/0000-0001-6178-5567

Mounir Ait Kerroum
Computer Science Research Laboratory, Faculty of Sciences, IBN Tofail University, Kenitra, Morocco
Email: maitkerroum@gmail.com
ORCID iD: https://orcid.org/0000-0001-7519-4434

Received: 25 April, 2025; Revised: 23 June, 2025; Accepted: 10 September, 2025; Published: 08 October, 2025

Abstract: The extreme volatility of Bitcoin markets makes accurate price prediction notably difficult. This paper proposes a new hybrid deep learning model that incorporates a Gated Recurrent Unit (GRU), a Bidirectional Long Short-Term Memory (Bi LSTM) model, and a Multi Head Attention mechanism to permit the model to utilize both historical price data and sentiment information from Twitter. We constructed the model utilizing a two-stage transfer learning approach: we first pretrained the model on data from 2017−2019 to learn lower-level fluctuation behaviors, then we fine-tuned the model on data from 2021−2023 in order to be sensitive to recent market behaviors. The model performed exceptionally well against multiple state-of-the-art baselines using root mean square error (RMSE) and mean absolute error (MAE) metrics, reporting RMSE values of 679.61 and MAE of 452.95, achieving considerable improvement over the baseline models. Our experimental results show that leveraging Twitter sentiment greatly improved trend prediction. In addition, our benchmarks showed that our method performed better than the existing methods. Furthermore, our ablation studies illustrated how each particular feature performed. Overall, our results demonstrate that multi-scale temporal modeling combined with social media sentiment integration produces a scalable and resilient solution to combat the challenges of volatility to forecast cryptocurrency prices accurately and efficiently.

Index Terms: Bitcoin Price Prediction, Hybrid Deep Learning, Transfer Learning, GRU, Bi-LSTM, Multi-Head Attention, Sentiment Analysis, Time-Series Forecasting

1. Introduction

Bitcoin, an early decentralized alternative currency, does not operate through a central financial authority or individual customization. Bitcoin emerged in 2009 and became complicated after the roller coaster spiking in 2017, signaling either just a novelty item and or a true replacement of the economic system we were accustomed to. Bitcoin has been extremely volatile; volatility draws the attention of economists and investment companies wanting to determine price changes based on movement relative to other traditional asset classes on the market. A good price prediction is useful for existing and future investors as well as state policymakers to gain an understanding and possibly to lead to control of this new market [1]. There is currently an alternative economy, primarily built in a technological environment similar to the American economy and the world, which will continue to offer alternative developments of financial systems with further digitization and cryptographic technologies.

These technologies demonstrated to be deteriorating for economic models and potentially providing promise for an alternative economy. The alternative economy is expected to be worth $23 trillion, or approximately 25% of all economic activity, by 2025. The alternative economy has many forms; we need to understand the importance of market behavior and forecasting [2].

Initially, cryptocurrency fee prediction trusted classical statistical approaches, together with the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) [3] and Autoregressive Integrated Moving Average (ARIMA) [4]. However, those techniques, broadly speaking, become aware of linear styles and count on generally disbursed variables situations that not often preserve authentic in risky, non-Gaussian cryptocurrency markets [5]. Machine studying methodologies cope with those barriers with the aid of taking pictures of nonlinear relationships inside big datasets without strict assumptions approximately underlying statistics distributions. Nevertheless, conventional system studying algorithms like Support Vector Regression (SVR), Regression Tree (RT), and Bayesian Regularization Neural Networks (BRNN) [6,7] frequently be afflicted by drawbacks together with overfitting and trouble leveraging high-degree latent styles found in sequential statistics. To triumph over those challenges, current research has followed deep study primarily based on totally prediction models, always demonstrating advanced overall performance in comparison to conventional system studying techniques [8,9]. Deep learning approaches, which do not have the strong assumptions of stationarity, and have greater generalization ability, have shown their promises when using supervised learning for predicting financial time series. Systems based on Gated Recurrent Units (GRUs) or Long Short-Term Memory (LSTMs) without and with attention frameworks, can learn temporal dependencies in time-series data through GRU and LSTM systems and their ability to learn with gradient based optimization. All of these learning capabilities led to modelling complicated financial sequences, and accurate prediction [10]. In response to the identified limitations, this study proposes a composite deep learning architecture that combines historical price data and sentiment analysis to better predict Bitcoin prices. The framework consists of three neural network components: Gated Recurrent Unit (GRU) to capture short-term patterns, Bidirectional Long Short-Term Memory (Bi-LSTM) network to examine both past and future patterns, and Multi-Head Attention mechanism to detect relevant features in the volatile market.

The major contributions in this paper are described as follows:

- Novel Hybrid Deep Learning Approach: We propose a novel hybrid architecture with Gated Recurrent Units (GRU), Bidirectional Long Short-Term Memory (Bi-LSTM) and Multi-Head Attention functionalities. This approach leverages time series trends along with sentiment-informed effects to enable robust predictions of Bitcoin's nontraditional price trajectories. By extracting short-term dependencies, bidirectional patterns, and relevant features hierarchically, this model represents a dynamic approach for volatile cryptocurrency markets.
- Dual-Phase Transfer Learning for Temporal Flexibility: This approach is a novel transfer learning approach that integrates initial training on historical data (2017-2020) with a fine-tune on the new datasets (2021 and later). This is relevant so as to minimize overfitting to past trends while supporting the flexibility needed to adapt to incoming updates and continuously improve prediction accuracy in a rapidly shifting economic landscape.

This paper is structured as follows: Section 2 reviews related work on cryptocurrency price prediction. Section 3 outlines the proposed methodology. Section 4 describes the implementation and testing procedures. Section 5 provides the ablation study. Section 6 discusses the limitations and future work. Finally, Section 7 concludes the paper.

2. Related Work

Predictions of the price of Bitcoin has become an increasingly popular research phenomenon in the past several years since bitcoin is volatile and has an impact on global economic markets. Several approaches have been employed by researchers, ranging from traditional statistical models to machine learning and deep learning approaches to analyze and predict the price fluctuations of Bitcoin. This section surveys the literature and identifies gaps and challenges while considering how this research study adds value to the broader literature. Machine learning and deep learning approaches have been used to predict cryptocurrency prices, with a particular emphasis on Bitcoin. In their study on cryptocurrencies, the authors [11] used logistic regression, support vector machines (SVM) and random forest models to predict bitcoin prices. In their work, logistic regression had the highest accuracy (66%). Their findings were illustrative of the independence of bitcoin returns with respect to other cryptocurrency returns and macro factors.

Authors in [12] conducted a thorough comparison of forecasting methods, including ARIMA, k-nearest neighbors. (kNN), SVR, Random Forest, LSTM, GRU, and Temporal Fusion Transformer (TFT). These authors highlighted the superior performance of LSTM models, and they highlighted their ability to capture temporal dependencies and achieve the lowest root mean square error (RMSE). Similarly, authors in [13] examined deep learning models that included LSTM, GRU, and Bi-Directional LSTM (Bi-LSTM), and they found that Bi-LSTM provided the best forecasts, as it effectively used sequential information going forward and using information from the past. On a related note, authors in [14] presented reinforcement learning models that outperformed traditional models (e.g., ARIMA and neural networks), while also showing more flexibility in volatile market environments. Authors in [15] compared random forest regression and LSTM, and here concluded random forest regression produced better predictive performance, especially with decisive signals such as stock market indices and Ethereum prices. Similarly, authors in [16] confirmed the predictive powers of random forests and bagging methods, identifying technical signals as important predictors of Bitcoin price movements.

Authors in [17] examined the role of information articles in Bitcoin price changes using text mining techniques, namely N-Gram, TF-IDF, Doc2vec, and the more experimental SentiGraph technique. They found that more advanced text sentiment analysis greatly improved forecasting accuracy. Authors in [18] conducted a follow-up study which found that multisource economic sentiment analysis significantly improved the forecasting capabilities of deep learning architecture. In high-frequency trading, there is an author [19] who focused on direct feed-forward neural networks (DFFNNs) trained with a variety of optimization algorithms and concluded that the Levenberg-Marquardt method was the most effective. Authors in [20] proposed a recurrent reinforcement learning solution that effectively modeled market dynamics and yielded substantial trading gains.

Hybrid modeling techniques have also gained prominence. Authors in [21] introduced the VMD-AGRU-RESVMD-LSTM hybrid framework, effectively managing market volatility and surpassing traditional models in forecasting accuracy. Authors in [22] integrated technical indicators with Performer neural networks and BiLSTM, significantly improving predictive accuracy and computational efficiency. Similarly, authors in [23] employed LSTM-Attention mechanisms with gradient-specific optimization, demonstrating reliable forecasting results.

Authors in [24] expanded such work by examining high-frequency Bitcoin and Ethereum price data during the COVID-19 period, confirming LSTM's resilience under crisis conditions. More recent studies by authors in [25] explored advanced architectures such as Transformers and hybrid models, highlighting their superior ability to model intricate temporal and cross-feature interdependencies compared to conventional methodologies.

Social signals have also been shown to influence cryptocurrency forecasts. Authors in [26] incorporated data from GitHub and Reddit, achieving notable improvements in predictive performance. Authors in [27] introduced ChainNet, which relies on blockchain graph features to significantly enhance forecast accuracy. In parallel, authors in [28] applied BERT classifiers and weak supervision, substantially elevating cryptocurrency return forecasts. Meanwhile, authors in [29] analyzed high-frequency markets through deep learning, demonstrating remarkable predictive accuracy.

Lastly, authors in [30] improved Bitcoin price predictions by leveraging social media sentiment through fine-tuned BERT models, reflecting continuing efforts to advance accuracy and robustness. Collectively, these works illustrate notable strides in cryptocurrency price prediction, rooted in advanced machine learning methods, hybrid models, sentiment analysis, and reinforcement learning. Nevertheless, enduring challenges—such as mitigating data noise, strengthening model robustness, and optimizing computational efficiency persist as critical focal points for future research. The summary of reviewed methodologies is displayed in Table 1.

Table 1. Summary of Reviewed Approaches

Reference	Methods	Objectives	Limitations
[11]	Logistic Regression, SVM, Random Forest	Predict Bitcoin prices using classical ML techniques	Limited accuracy (max 66%); ignores temporal dependencies
[12]	ARIMA, kNN, SVR, RF, LSTM, GRU, TFT	Compare time-series models for Bitcoin prediction	Traditional models underperform; high computational cost for TFT
[13]	LSTM, GRU, Bi-LSTM	Evaluate deep sequential models	Limited external data integration; lacks real-time adaptability
[14]	Reinforcement Learning	Model volatile market conditions	Complexity in training; sensitive to reward shaping
[15]	Random Forest, LSTM	Compare performance with technical/market signals	Model performance varies by signal quality
[16]	Random Forest, Bagging	Analyze technical signals' predictive power	Prone to overfitting; limited interpretability
[17]	TF-IDF, N-Gram, Doc2Vec, SentiGraph	Use textual sentiment to enhance predictions	Model performance varies by text source quality
[18]	Deep Learning + Multi-source Sentiment	Integrate diverse economic sentiment data	Sentiment signals may introduce noise if not filtered
[19]	DFFNN + Optimization Algorithms	Optimize network training for high-frequency trading	Scalability issues; lacks temporal feedback modeling
[20]	Recurrent Reinforcement Learning	Forecast trends under dynamic market conditions	Difficult to generalize across markets
[21]	VMD-AGRU-RESVMD-LSTM Hybrid	Address volatility and increase forecast accuracy	High model complexity: parameter tuning is challenging
[22]	Performer, Bi-LSTM, Technical Indicators	Improve prediction and efficiency	Limited validation in live trading environments
[23]	LSTM + Attention + Gradient Tuning	Enhance focus on key signals	Attention mechanisms can be computationally intensive
[24]	LSTM on High-Frequency COVID-19 Data	Assess robustness during crisis	Lacks transferability to post-crisis market conditions
[25]	Transformers, Hybrid Models	Capture deep cross-feature dependencies	Requires large datasets and high computing power

[26]	LSTM with GitHub and Reddit Sentiment	Improve forecasts using developer activity and public sentiment	Social and activity metrics are context-dependent and may not generalize
[27]	ChainNet (Blockchain Graph Learning)	Leverage on-chain graph data	Requires rich blockchain feature engineering
[28]	BERT, Weak Supervision	Improve return predictions using NLP	Fine-tuning BERT models are resource-intensive
[29]	Deep Learning on High-Frequency Data	Maximize predictive accuracy in fast markets	Sensitive to noise and requires continuous retraining
[30]	Fine-tuned BERT with social media	Use social media sentiment for prediction	Data labeling and interpretation remain challenging

3. Methodology

This section explains the end-to-end architecture for Bitcoin price prediction, covering data collection, preprocessing, sentiment analysis, and model design. To enhance generalization across varying market regimes, our pipeline includes a two-phase transfer-learning strategy. In Phase 1 (Pretraining), the GRU → Multi-Head Attention → BiLSTM model is trained on the 2017–2019 dataset to learn stable long-term patterns. In Phase 2 (Fine-tuning), the pretrained model is further trained on the 2021–2023 dataset, allowing it to adapt efficiently to recent market dynamics without retraining from scratch. The overall workflow is illustrated in Fig. 1:

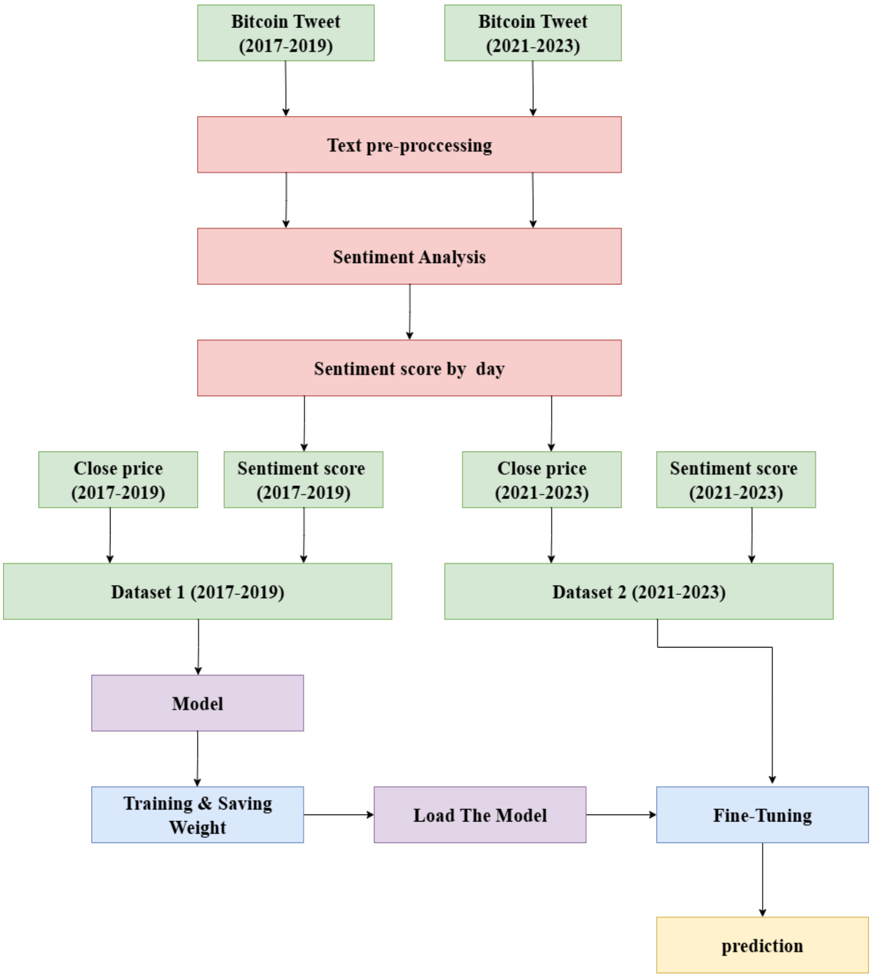

Fig. 1. An overview of the overall workflow of the proposed architecture.

3.1 Data Collection

This study used two datasets (spanning 2017–2019 and 2021–2023) to investigate the relationship between Bitcoin price movements and Twitter sentiment. Both datasets were obtained from publicly available Kaggle repositories containing Bitcoin price information and Bitcoin-related tweets. The historical price data set provides daily closing prices, trading volumes, and other key financial metrics, all aligned with the same temporal resolution as the sentiment data. Fig. 2 illustrates daily Bitcoin price trends for the 2021–2023 period.

For sentiment analysis, a pre-collected dataset of Bitcoin-related tweets was retrieved from Kaggle, featuring hashtags such as #Bitcoin, #BTCPrice, #Crypto, and #BitcoinNews. More than 20 million tweets were gathered, yielding valuable insights into public sentiment and enabling an exploration of the correlation between social media sentiment and Bitcoin price dynamics. Fig. 3 provides an example of raw tweets from 2023.

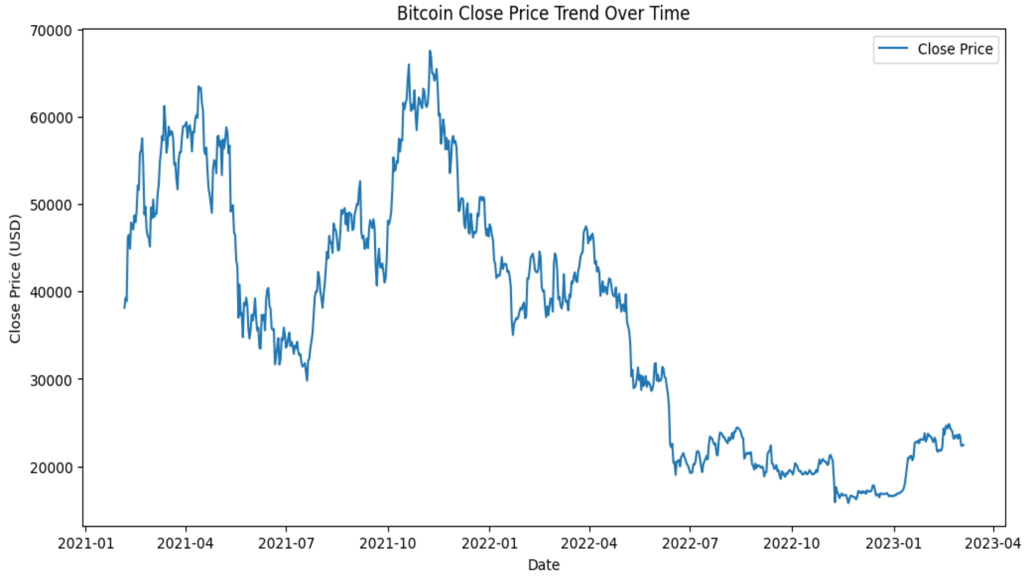

Fig. 2. Bitcoin closing prices recorded over the time period from 5 February2021 to 5 March 2023.

	timestamp	text
0	2023-03-01 23:59:59	Which #bitcoin books should I think about read...
1	2023-03-01 23:59:47	@ThankGodForBTC I appreciate the message, but ...
2	2023-03-01 23:59:42	#Ethereum price update: \n\n#ETH $1664.02 USD\...
3	2023-03-01 23:59:36	CoinDashboard v3.0 is here\nAvailable on ios a...
4	2023-03-01 23:59:32	#Bitcoin Short Term Fractal (4H) 💥 \n\nIn lower ...

Fig. 3. Sample of raw tweets from 2023, illustrating the state of the data before preprocessing.

3.2 Feature engineering and data preprocessing

To identify relevant input features for the prediction model, we conducted a Pearson correlation analysis among all available numerical variables in the dataset. The correlation matrix is summarized in Fig. 4. As observed, the Sentiment feature shows a moderate positive correlation with the Close price (0.43), suggesting that public sentiment has a non-trivial relationship with the market's closing value. While other features such as Open, High, and Low also exhibit very high correlations with Close (above 0.99), these variables are collinear and may introduce redundancy. To minimize overfitting and improve generalizability, we selected Sentiment and Close as the primary features—capturing both external (news sentiment) and internal (price-based) market dynamics. This selection balances predictive strength with unique features. To enhance the quality and usability of the close price and tweet datasets (2017–2019 and 2021–2023), several preprocessing steps were implemented. First, duplicate rows were eliminated using the drop_duplicates method. Next, the datetime function was used to convert the 'date' column into a standardized datetime format. The re library was then utilized to remove extraneous hashtags, newline characters, hyperlinks (URLs), alphanumeric text, and mentions from the tweet data. Fig. 5 illustrates examples of cleaned tweets.

Additionally, 20-day and 50-day moving averages for the Bitcoin closing price were calculated to capture short-term fluctuations and illuminate longer-term trends (see Fig. 6). This approach provides clearer insights into Bitcoin price movements over time.

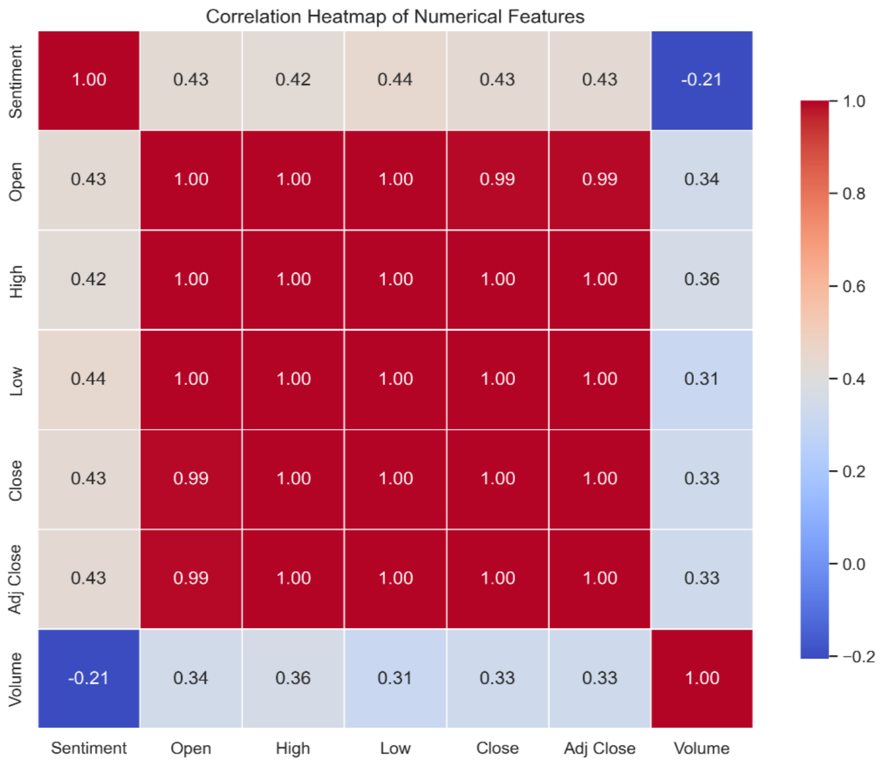

Fig. 4. Correlation matrix displaying the Pearson correlation coefficients between all numerical variables in the dataset from 2021 to 2023.

	timestamp	text	CleanTwt
0	2023-03-01 23:59:59	Which #bitcoin books should I think about read...	Which bitcoin books should I think about readi...
1	2023-03-01 23:59:47	@ThankGodForBTC I appreciate the message, but ...	I appreciate the message, but not a fan of the...
2	2023-03-01 23:59:42	#Ethereum price update: \n\n#ETH $1664.02 USD\...	price update: $1664.02 USDbitcoin 0.070428 b...
3	2023-03-01 23:59:36	CoinDashboard v3.0 is here\nAvailable on ios a...	CoinDashboard v3.0 is hereAvailable on ios and...
4	2023-03-01 23:59:32	#Bitcoin Short Term Fractal (4H) 💥 \n\nIn lower ...	bitcoin Short Term Fractal (4H) 💥 In lower timef...

Fig. 5. Sample processed and cleaned tweets after applying preprocessing techniques.

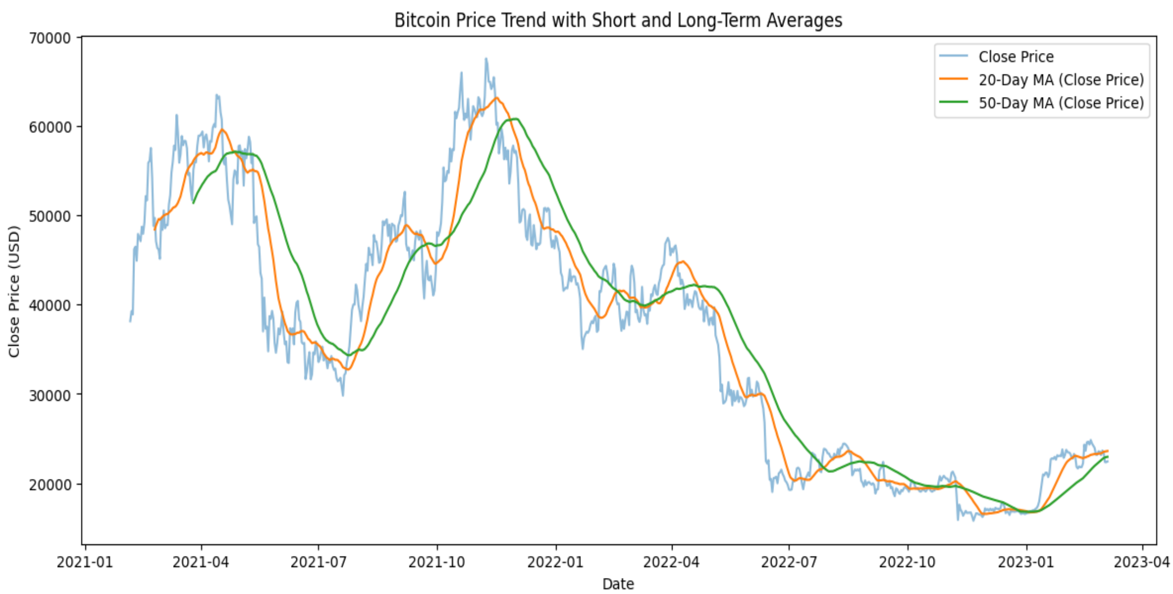

Fig. 6. Visualization of Bitcoin Close Price Trends observed during the period of 2021 to 2023.

3.3 Sentiment Analysis for Both Time Durations of Bitcoin Tweets

Sentiment analysis, which identifies the emotions (positive, negative, or neutral) conveyed in text, benefits greatly from the real-time and diverse nature of social media posts such as tweets. Since they are user-generated, tweets provide particularly valuable data for sentiment analysis. In this study, the cleaned tweets from 2017–2019 and 2021–2023 were analyzed using the VADER Sentiment Analysis tool—an approach specifically designed for social media text. VADER calculates a compound score for each tweet, indicating overall sentiment on a scale from -1 (most negative) to +1 (most positive), with scores near zero suggesting no dominant sentiment. By leveraging predefined weights to gauge each word's contribution, VADER facilitates accurate sentiment detection in short, informal texts like tweets. Once the sentiment score is computed, each tweet is classified as positive, negative, or neutral according to defined thresholds.

Although VADER was chosen for sentiment analysis due to its computational efficiency and suitability for real-time analysis, it has certain limitations. As a lexicon and rule-based model, VADER may struggle with capturing sarcasm, contextual ambiguity, and domain-specific jargon commonly found in financial tweets. It can also be sensitive to noisy or non-representative data. However, VADER is specifically optimized for social media language—handling emojis, slang, and punctuation-based emphasis (e.g., capitalization)—which makes it highly effective for Twitter-based sentiment analysis. Its lightweight design and ease of deployment enabled efficient processing of large-scale datasets without requiring labeled data or fine-tuning. While modern transformer-based models like BERT or RoBERTa are capable of capturing deeper contextual meaning, they are computationally intensive and typically require substantial labeled datasets for fine-tuning. In contrast, VADER's out-of-the-box usability and real-time performance made it a practical and robust choice for this study. The high accuracy achieved by our forecasting model further proves that the sentiment signals extracted by VADER were sufficiently effective for the task.

3.4 Sentiment Score Group-by-Day

The daily sentiment scores are initially calculated as a mean value for each day (see Fig. 7), illustrating how Bitcoin-related discussions shift over time. Subsequently, the dataset is augmented with positive and negative ratios, representing the percentages of tweets labeled positive or negative on a given day. To capture both short-term and long-term sentiment patterns, 20-day and 50-day moving averages of the daily sentiment scores are calculated (see Fig. 8). This smooth approach provides deeper insight into overall sentiment trends and their potential impact on Bitcoin price movements.

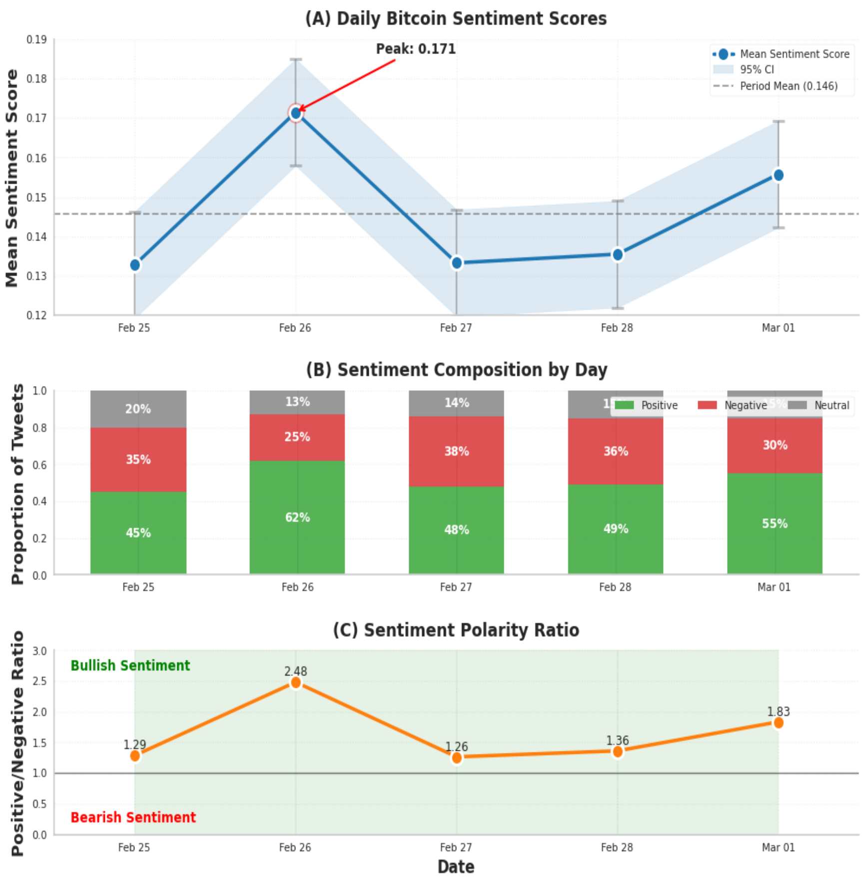

Fig. 7. Sentiment Trends in Bitcoin-Related Tweets: Scores, Composition, and Polarity Ratios (Feb–Mar 2023).

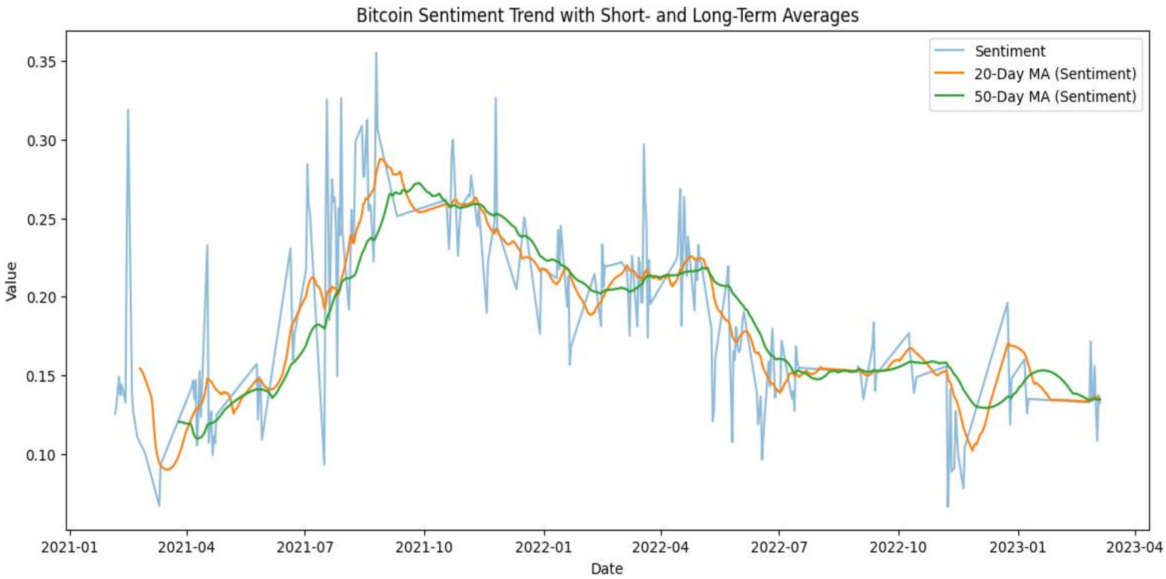

Fig. 8. Visualization of Bitcoin sentiment trends (2021–2023), illustrating sentiment fluctuations over time.

3.5 Merging Sentiment Scores with Bitcoin Closing Prices Across Two Periods

To investigate sentiment alongside price fluctuations, this study combines daily sentiment scores from tweets with Bitcoin's closing prices in a single CSV file for two distinct periods: 01/01/2017–23/11/2019 and 05/02/2021–05/03/2023. The datasets are merged using the timestamp as the common key, producing two integrated datasets—Dataset 1 for 2017–2019 and Dataset 2 for 2021–2023. Each dataset includes columns for both sentiment scores and closing prices, creating a comprehensive resource for subsequent model training. Additionally, Fig. 9 illustrates correlation graphs between sentiment and closing prices.

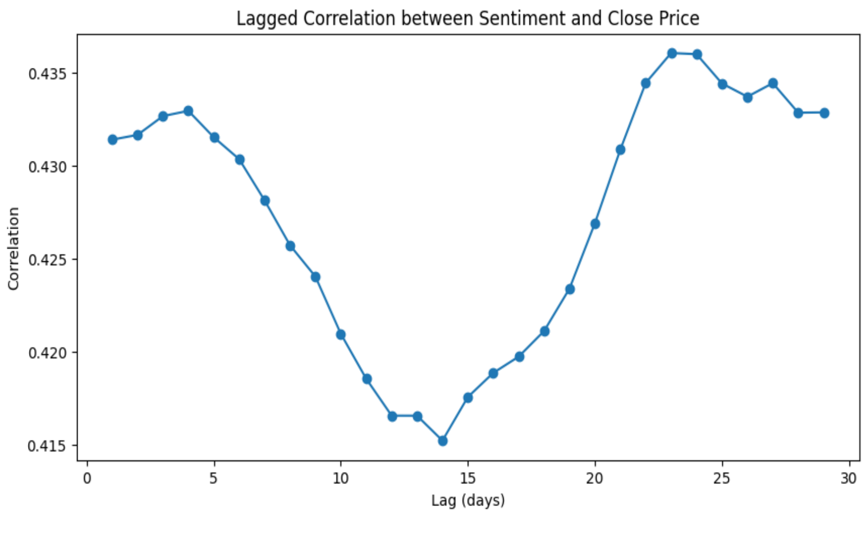

Fig. 9. Correlation analysis between sentiment scores and Bitcoin closing prices from 2021 to 2023.

3.6 Model Architecture

A hybrid model was developed to predict Bitcoin price trends by integrating two distinct feature sets: historical Bitcoin closing prices and sentiment scores extracted from relevant tweets. The architecture, illustrated in Fig. 10, operates as follows:

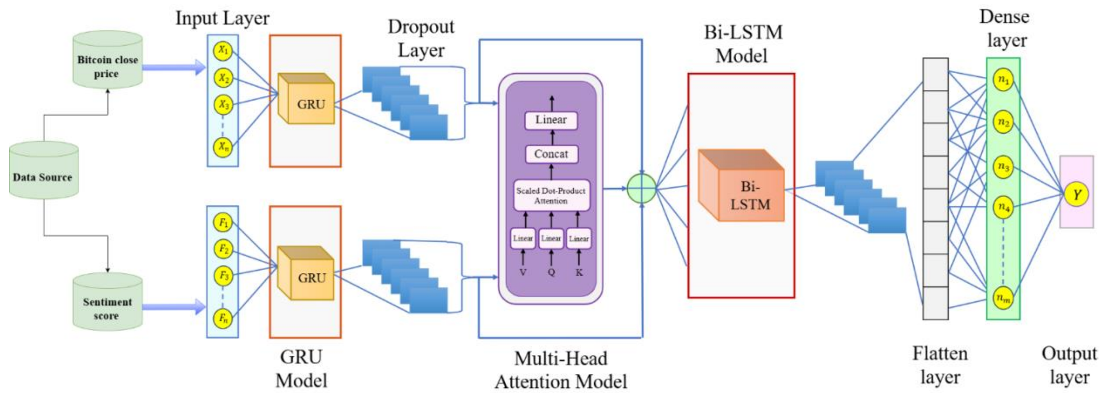

Fig. 10. A comprehensive overview of the proposed model architecture.

- **Step 1: Load the Datasets**

The proposed architecture processes two distinct sets of input features in parallel: historical Bitcoin closing prices and sentiment scores. Handling both feature sets concurrently enables the model to capture the interaction between time-series price dynamics and sentiment-driven market behavior. Let the two input feature sets be denoted as follows:

$$D_2 = \{X_1, t_1\} \tag{1}$$

Where X_1 is the Bitcoin Close price data and t_1 is the corresponding timestamp

$$D_2 = \{X_2, t_2\} \tag{2}$$

Where X_2 is the sentiment data, and t_2 is the corresponding timestamp.

$$X_1 = \{x_1, x_2 \ldots x_m\} \tag{3}$$

$$X_2 = \{x_1, x_2 \ldots x_n\} \tag{4}$$

Where $x_i \in R^d$ is a feature vector at time t_i for both sets.

- **Step 2: Extract Temporal Dependencies Using GRU**

Each feature set is passed to its corresponding Gated Recurrent Unit (GRU) layer, where temporal dependencies are captured, and sequential representations are inferred. One GRU focuses on the dynamic aspects of Bitcoin closing price movements, while the other extracts sentiment patterns, ensuring that the unique characteristics of each feature set are effectively learned.

$$G_1 = GRU\{X_1\} \tag{5}$$

$$G_2 = GRU\{X_2\} \tag{6}$$

- **Step 3: Regularization with Dropout Layers**

To reduce overfitting and enhance generalization, in each GRU output, we apply a dropout layer. Dropout uses a random selection of input units to turn off the training instances, which provides the ability to reduce overfitting as the training outcome is less reliant on specific features (i.e., the learned representation has to adapt).

$$D_1^{drop} = Dropout\{G_1\} \tag{7}$$

$$D_2^{drop} = Dropout\{G_2\} \tag{8}$$

- **Step 4: Attention Mechanism for Feature Integration**

The GRU outputs, after regularization, are fed into a multi-head attention layer that highlights significant features from both sets, emphasizing key aspects of Bitcoin prices and sentiment scores. This method provides a mechanism for the model to learn long-term dependencies and fine-grained interactions. By attending and prioritizing particular aspects of Bitcoin price and sentiment score in each feature set, the multi-head attention layer allows the model to compute the contextual understanding of the interrelations.

$$A = MultiHeadAttention\left(D_1^{drop}, D_2^{drop}\right) \tag{9}$$

- **Step 5: Feature Fusion and Sequential Learning**

The outputs from the dropout and multi-head attention layers are concatenated, integrating the learned features from each pathway. Once combined, these features were input into a Bi-Directional Long Short-Term Memory (Bi-LSTM) layer, which learns both forward and backward dependencies within the concatenated input feature set. By learning dependencies in both directions, bi-LSTM will be able to learn more complex relationships and more interactions.

$$C = Concatenate\left(A, D_1^{drop}, D_2^{drop}\right) \tag{10}$$

$$BL = BiLSTM(C) \tag{11}$$

- **Step 6: Final Regularization and Prediction**

To reduce overfitting further, I utilized a dropout layer on the output of the Bi-LSTM which will enhance the robustness of learned features. The output of the Bi-LSTM features is then flattened and passed through a dense layer, aggregating all the information to produce the final output of the predicted trend of Bitcoin prices.

$$D^{drop} = Dropout(BL) \tag{12}$$

$$F = Flatten(D^{drop}) \tag{13}$$

$$Y = Dense(F) \tag{14}$$

4. Implementation and Testing

4.1 Model Training

This study adopts a two-stage training strategy that leverages the benefits of transfer learning:

- Pretraining with Dataset 1:

The model is *initially trained* on Dataset 1, which *includes* sentiment ratings and Bitcoin last *test nets* from 01/01/2017 to 23/11/2019. This *phase develops* the *model'*s foundational *ability* to *identify dispersed patterns* in *historical data* and *entangled relationships between* sentiment and price. *The model* learns long-*term* temporal dependencies as well as market fundamentals through pretraining and creates a base of generalized knowledge for future fine-tuning.

- Fine-Tuning with Dataset 2:

Building at the pretrained foundation, the model is then fine-tuned the use of Dataset 2, which incorporates sentiment rankings and Bitcoin remaining fees from 05/02/2021 to 18/08/2022. Fine-tuning is critical, because it adjusts the version to evolving marketplace situations and sentiment trends, updating formerly found out styles to mirror the traits of an extra latest and doubtlessly unstable dataset. By refining its know-how of nearby versions and short-time period trends, the version can provide robust, context-particular insights that efficaciously reply to contemporary marketplace nuances.

4.2 Model Testing

To evaluate the efficacy of the proposed architecture, Dataset 2 is split into training and testing sets with an 80:20 ratio. The training set is used to fine-tune the previously profiled model. The testing set describes sentiment scores and daily Bitcoin closing prices between 19/08/2022 and 05/03/2023 and is used to evaluate the model's performance once fine-tuning is complete. This testing set will allow the comparison of performance before and after fine-tuning. Initially, after training on Dataset 1, the model is tested on data from 19/08/2022 to 05/03/2023. This step reveals how well the model applies learned patterns to a markedly different period. The same testing data is used again after fine-tuning on Dataset 2, enabling a clear evaluation of how the additional fine-tuning step enhances performance.

4.3 Performance Evaluation

This paper evaluates the proposed model using both quantitative performance metrics and visual representations to assess its predictive accuracy. Specifically, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are employed to summarize the model's performance, with each metric calculated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\rho_i - \hat{\rho}_i)^2} \tag{15}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\rho_i - \hat{\rho}_i\right| \tag{16}$$

where ρ is the actual Bitcoin price at index i and $\hat{\rho}$ is the predicted Bitcoin price at index i.

The daily Bitcoin closing price data exhibits a complex interplay of upward trends and nonlinear patterns, reflecting the inherent volatility of the cryptocurrency market. In this case, the nonlinear behavior of our model captures this. The training and validation loss curves in Fig. 11, show decreasing error and convergence during the training phase exercise. On the test dataset, the model performance on RMSE and MAE is good; with an RMSE of 679.6131 and MAE of 452.9485, which outperforms the state-of-the-art techniques used. The results demonstrate that the model effectively minimizes the errors between the predicted and true values, summarized in Fig. 12. Furthermore, Fig. 13 provides a visual comparison between historical Bitcoin prices and the predicted values across the complete dataset, highlighting the precision of the model's forecasts. Collectively, the consistently low error metrics and reliable predictive performance affirm the robustness of the proposed model within a highly volatile market context.

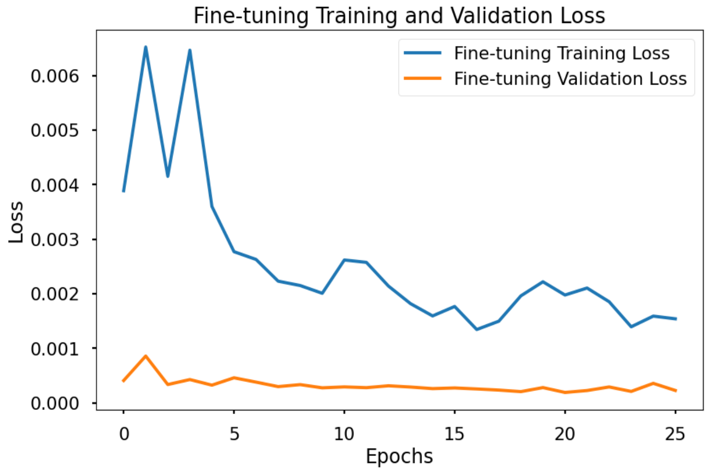

Fig. 11. Training and validation loss curves during the fine-tuning process, which emphasizes that the model's performance improved over time

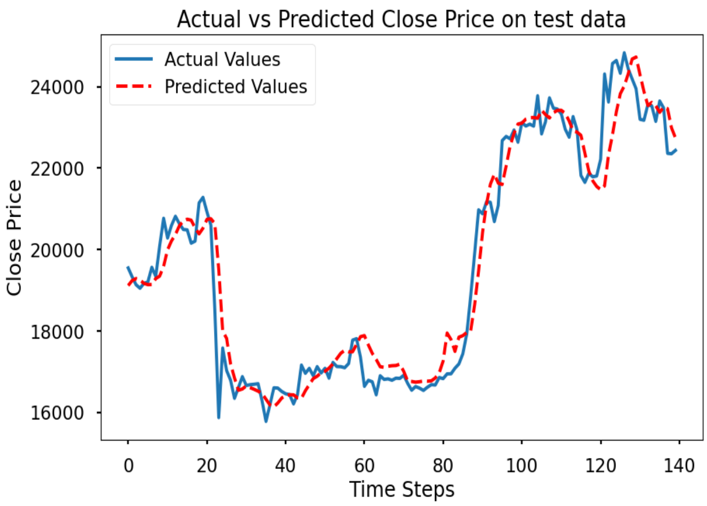

Fig. 12. Graphical representation of the predicted Bitcoin closing price curve on the test dataset, illustrating the model's forecasts over a given time period.

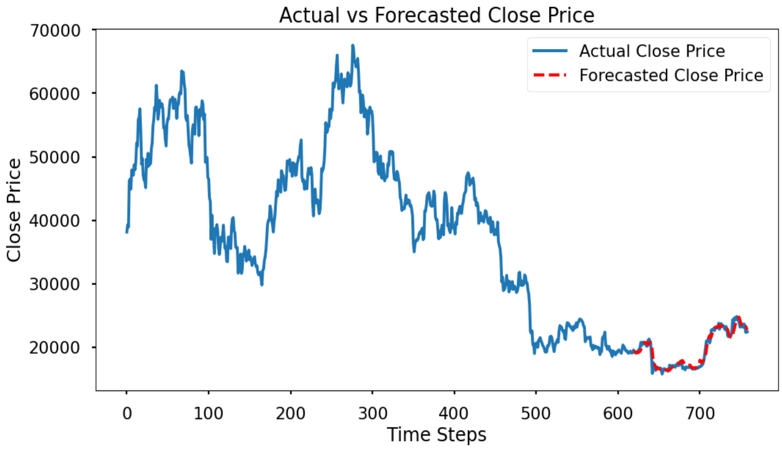

Fig. 13. Graphical comparison of predicted and actual Bitcoin closing prices across the full dataset, illustrating the model's forecasting accuracy.

4.4 Nested Cross-Validation for Transfer Learning

To address concerns about overfitting and enhance the robustness of model evaluation, we implemented a 5×5 nested cross-validation framework. In this setup, the 2017 dataset was exclusively used for pretraining and partitioning into outer folds, while the 2023 dataset was utilized for fine-tuning and inner fold evaluation. For each of the five outer folds, the model was pretrained on a subset of the 2017 data and validated on the corresponding holdout set. Subsequently, a five-fold inner cross-validation was conducted on the 2023 data to fine-tune the pretrained model and assess its generalization across multiple temporal splits. This process resulted in 25 model evaluations, providing a robust estimate of performance variability. Table 2 presents the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) for each outer fold, demonstrating consistent performance across folds. Fig. 14 displays the predicted vs. actual values for the best-performing outer-inner fold combination, highlighting the model's ability to closely follow real-world trends.

These results confirm the reliability and stability of the proposed transfer learning approach, validating its applicability to time-series forecasting tasks with domain shift.

Table 2. Performance metrics (MAE and RMSE) across 5 outer folds during nested cross-validation.

Outer Fold	MAE	RMSE
1	3260.60	4065.42
2	3089.10	3714.84
3	2796.30	3422.60
4	2497.12	3082.43
5	**2178.35**	**2761.05**

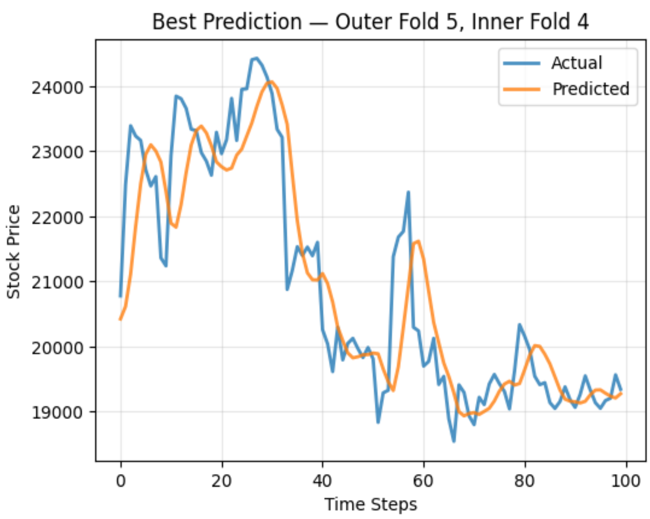

Fig. 14. Predicted vs. actual values for the best-performing model from Outer Fold 5 and Inner Fold 4, showing accurate forecasting alignment.

4.5 Comparison with Other Models

Table 3 shows a comparative evaluation of the suggested model's performance vs a few other architectures. Individual models (such as Bi-LSTM, LSTM, and GRU) have been previously applied as baseline methods for temporal dependencies and directional patterns. While these individual architectures exhibit commendable predictive capabilities, their effectiveness remains limited compared to more sophisticated approaches. Conversely, integrating GRU with Bi-LSTM significantly improves predictive accuracy by simultaneously capturing short-term dependencies and leveraging the bidirectional context of temporal data. Furthermore, recent studies indicate that the inclusion of Attention mechanisms within GRU-based models enhances performance through refined emphasis on sequential data, surpassing configurations like Stacked LSTM with Attention and CNN-LSTM.

However, the proposed architecture, which combines GRU, Bi-LSTM, and Multi-Head Attention, achieves superior overall performance, attaining the lowest RMSE (679.6131) and MAE (452.9485). This integrated framework not only exceeds current state-of-the-art models but also offers compelling evidence for its enhanced accuracy and reliability in predicting Bitcoin prices. Consequently, the proposed model distinctly outperforms previously studied methods within existing literature.

Table 3. Performance comparison of individual and Hybrid Deep learning models.

Components	**RMSE**	**MAE**
Bi-LSTM	789.8469	548.9025
LSTM	868.0900	580.9269
GRU	851.0728	659.1544
Stacked LSTM	1245.1245	930.2034
GRU + BiLSTM	760.2150	511.2294
GRU + Attention	860.6273	661.5795
Stacked LSTM + Attention	1365.3421	1033.8769
CNN + LSTM	1694.5250	1493.8619
Our Model	**679.6131**	**452.9485**

5. Ablation Study

5.1 Use of Dataset 2 only (2021–2023)

An experiment was conducted where the model was trained exclusively on Dataset 2 (2021–2023) and then evaluated using the same test set as the proposed architecture. The results revealed poor performance, confirming that Dataset 2 alone is inadequate for this task. Notably, the model recorded high error metrics, with a Root Mean Squared Error (RMSE) of 1225.5140 and a Mean Absolute Error (MAE) of 940.5186. Table 4 and Fig. 15 illustrate this performance decline, highlighting a clear gap related to the more robust and fine-tuned model.

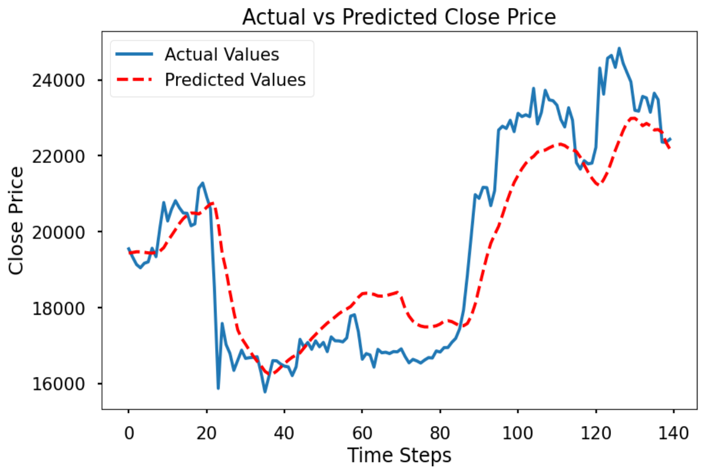

Fig. 15. Prediction results using only Dataset 2.

Table 4. Performance comparison of the proposed hybrid model under various ablation settings.

Dataset	RMSE	MAE
Only Dataset 2	1225.5140	940.5186
Only Dataset 1	1042.838	833.0995
Both Datasets (Close Price only)	774.3941	483.4996
Our model using both datasets and both features (price and sentiment)	**679.613**	**452.9485**

5.2 Use of Dataset 1 only (2017–2019 & 2021–2023)

When trained exclusively on Dataset 1 (2017–2019), the model achieved better results than the model trained solely on Dataset 2; however, its performance remained noticeably lower than that of the fully proposed methodology. Although the RMSE (1042.8382) and MAE (833.0995) were lower than the metrics obtained with Dataset 2, they still fell short of optimal performance. These findings underscore that restricting the model to Dataset 1 diminishes its predictive capacity, reinforcing the value of integrating multiple datasets to boost generalization and accuracy. Table 3 and Fig. 16 provide further evidence of these performance differences.

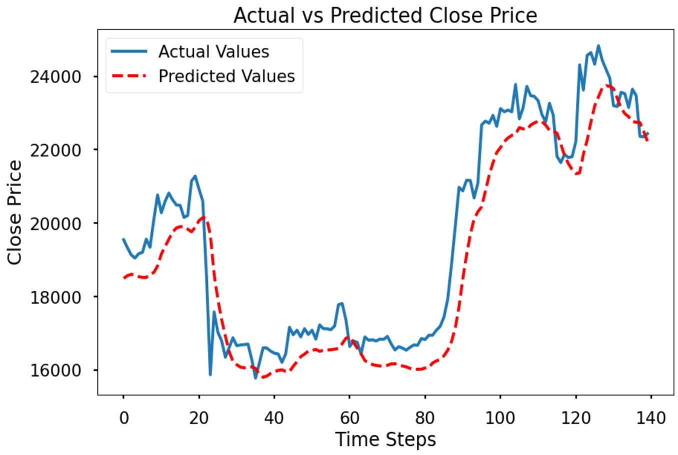

Fig. 16. Prediction results using only Dataset 1.

5.3 Use of Only Bitcoin Close Price Data (2017–2019 & 2021–2023)

Utilizing only Bitcoin close price data from both periods (2017–2019 and 2021–2023) resulted in reduced model performance, as summarized in Table 3. Although transfer learning effectively transferred knowledge from the earlier period to the later one, relying exclusively on numerical price trends proved insufficient, limiting the model's predictive ability. Fig. 17 further illustrates these limitations. By integrating sentiment data alongside historical price trends, the model achieves a more comprehensive understanding of influencing factors, thus improving both accuracy and robustness.

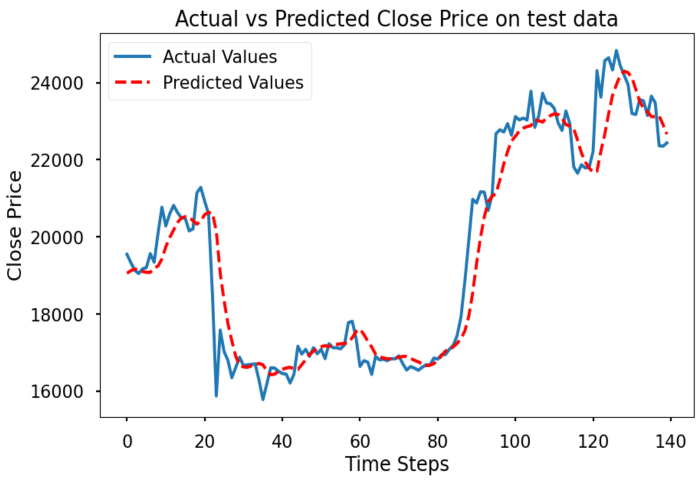

Fig. 17. Predictions using only Bitcoin price data from both periods.

6. Limitations and Future Work

This section outlines critical limitations of the proposed hybrid model and offers potential directions for further enhancement. While our architecture performs robustly under various market regimes, a deeper look into real-world challenges and future opportunities is essential.

6.1 Model Generalization in Stable or Crash Periods

During extended periods of market stability or prolonged crashes, reduced volatility may limit the diversity of learnable patterns, potentially affecting the model's generalization capabilities. However, the combined use of GRU, Multi-Head Attention, and Bi-LSTM enhances the model's ability to capture subtle temporal and contextual dependencies, helping mitigate this issue.

6.2 Real-Time Adaptability and Deployment

The proposed model is inherently well-suited for real-time forecasting applications. Using a sliding window mechanism and periodic fine-tuning on the latest data, the model can be adapted for near real-time predictions. Technologies like Kafka for live data ingestion (e.g., Twitter and market feeds), and deployment tools such as TensorFlow Serving or PyTorch Serve, offer a pathway for seamless integration into financial dashboards or trading platforms.

6.3 Short-Term vs Long-Term Forecasting Horizons

While the current model focuses on short-term (daily) forecasting—which is useful for high-frequency traders—longer-term predictions (weekly or monthly) may be more valuable to institutional investors and policymakers. Future work could explore multi-horizon learning frameworks or hierarchical temporal models to capture both short- and long-term market dynamics within a unified architecture.

6.4 Model Interpretability

Interpretability remains a key concern in deep learning-based financial systems. While attention mechanisms offer some transparency by highlighting influential time steps and sentiment signals, components like Bi-LSTM and GRU are still considered black-box models. Future research may leverage post-hoc explanation techniques such as SHAP, LIME, or attention weight visualizations to improve transparency and build user trust—especially in high-stakes environments like finance.

6.5 Ethical Considerations

The use of social media sentiment in financial forecasting brings ethical challenges. Tweets can contain biased, noisy, or manipulated content, which may mislead models. As sentiment-based prediction tools gain traction in algorithmic trading, future implementations must include safeguards against misinformation, promote responsible data sourcing, and ensure transparency in automated decision-making to prevent unintended market consequences.

7. Conclusion

In summary, this research highlights the ability of sentiment analysis to enhance daily Bitcoin price prediction methods. The hybrid method developed utilized Gated Recurrent Unit (GRU), Bidirectional LSTM, and Multi-Head Attention and it outperformed the benchmark model. The hybrid model demonstrates transfer learning where the same model trained on two different time-series data is able to generalize across time-periods, capturing temporal and sentiment differences under differing market conditions. With evaluation results showing RMSE of 679.6131 and MAE of 452.9485, the hybrid model is more effective than the state-of-the-art models, including RNN-based methods. The visual assessment of the daily predictive graphs suggests that the predicted values were very close to the actual values. By using Bitcoin sentiment data, combined with advanced neural architectures, we have shown that the prediction of complex relationships which connect the public sentiment on social media with Bitcoin price changes is highly predictable. Overall, these results suggest that hybrid architecture and the use of transfer learning are valuable approaches to improve both accuracy and robustness in cryptocurrency price forecasting models.

References

[1] Kaushal, R. (2016). Bitcoin: First Decentralized Payment System. International Journal of Engineering and Computer Science, 5(5), 16514–16517. doi:10.18535/ijecs/v5i5.30.

[2] Oliver, J. (2013). Mastering blockchain distributed ledger technology, and smart contracts explained. Packt Publishing, Birmingham, United Kingdom.

[3] Dudek, Grzegorz, Fiszeder, Piotr, Kobus, Pawe, and Orzeszko, Witold. 2023. "Forecasting cryptocurrencies volatility using statistical and machine learning methods: A comparative study". Elsevier BV. https://doi.org/10.1016/j.asoc.2023.111132.

[4] Ibrahim, A., Kashef, R., & Corrigan, L. (2021). Predicting market movement direction for bitcoin: A comparison of time series modeling methods. Computers and Electrical Engineering, 89, Article 106905.

[5] Kervanci, I. and Akay, F.. 2020. "Review on Bitcoin Price Prediction Using Machine Learning and Statistical Methods". Sakarya University Journal of Computer and Information Sciences. https://doi.org/10.35377/saucis.03.03.774276.
[6] R. K. Jana, I. Ghosh, and D. Das, "A differential evolution-based regression framework for forecasting Bitcoin price," Annals of Operations Research, vol. 306, no. 1, pp. 295–320, Nov. 2021,https://doi.org/10.1007/s10479-021-04000-8.
[7] S. Lahmiri and S. Bekiros, "Cryptocurrency forecasting with deeplearning chaotic neural networks," Chaos, Solitons & Fractals, vol. 118,pp. 35–40, Jan. 2019, https://doi.org/10.1016/j.chaos.2018.11.014.
[8] Cui, T., Ding, S., Jin, H., & Zhang, Y. (2023). Portfolio constructions in cryptocurrency market: A CVaR-based deep reinforcement learning approach.EconomicModelling,119. doi:10.1016/j.econmod.2022.106078.
[9] Ortu, M., Uras, N., Conversano, C., Bartolucci, S., & Destefanis, G. (2022). On technical trading and social media indicators for cryptocurrency price classification through deep learning. Expert Systems with Applications, 198, 116804. doi:10.1016/j.eswa.2022.116804.
[10] Murray, Kate, Rossi, Andrea, Carraro, Diego, and Visentin, Andrea. 2023. "On Forecasting Cryptocurrency Prices: A Comparison of Machine Learning, Deep Learning, and Ensembles". Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/forecast5010010.
[11] Dimitriadou, A., & Gregoriou,A. (2023). Predicting Bitcoin Prices Using Machine Learning.Entropy, 25(777).https://doi.org/10.3390/e25050777.
[12] Murray, K., Rossi, A., Carraro,D., & Visentin, A. (2023). OnForecasting Cryptocurrency Prices: A Comparison ofMachine Learning, DeepLearning, and Ensembles.Forecasting, 5, 196–209.https://doi.org/10.3390/forecast5010010.
[13] Seabe, P.L., Moutsinga, C.R.B., & Pindza, E. (2023). Forecasting Cryptocurrency Prices Using LSTM, GRU, and Bi-Directional LSTM: A Deep Learning Approach. Fractal and Fractional, 7(203).https://doi.org/10.3390/fractalfract7020203.
[14] Mariappan, L.T., Pandian, J.A., Kumar, V.D., Geman, O., Chiuchisan, I., & N˘astase, C. (2023). AForecasting Approach to Cryptocurrency Price Index Using Reinforcement Learning. Applied Sciences, 13(2692). https://doi.org/10.3390/app13042692.
[15] Chen, J. (2023). Analysis of Bitcoin Price Prediction Using Machine Learning. Journal of Risk and Financial Management, 16(51). https://doi.org/10.3390/jrfm16010051.
[16] Basher, S. A., & Sadorsky, P. (2022). Forecasting Bitcoin price direction with random forests: How important are interest rates, inflation, and market volatility? Machine Learning with Applications, 9, 100355. https://doi.org/10.1016/j.mlwa.2022.100355.
[17] Yao, W., Xu, K., & Li, Q. (2019). Exploring the Influence of News Articles on Bitcoin Price with Machine Learning. Proceedings of the 2019 International Conference on Computer Science and Communication Engineering (ICSC), Shenzhen, China.
[18] Passalis, N., Avramelou, L., Seficha, S., Tsantekidis, A., Doropoulos, S., Makris, G., & Tefas, A. (2022).Multisource financial sentiment analysis for detecting Bitcoin price change indications using deep learning. Neural ComputingandApplications, 34, 19441-19452. https://doi.org/10.1007/s00521-022-07509-6.
[19] Lahmiri, S., & Bekiros, S. (2021). Deep Learning Forecasting in Cryptocurrency High-FrequencyTrading. Cognitive Computation, 13, 485–487. https://doi.org/10.1007/s12559-021-09841-w.
[20] Borrageiro, G., Firoozye, N., & Barucca, P. (2022). The Recurrent Reinforcement Learning Crypto Agent. arXiv preprint arXiv:2201.04699.
[21] Jin, C., & Li, Y. (2023). Cryptocurrency Price Prediction Using Frequency Decomposition and Deep Learning. Fractal and Fractional.ttps://doi.org/10.3390/fractalfract7100708.
[22] Khaniki, M.A., & Manthouri, M. (2024). Enhancing Price Prediction in Cryptocurrency Using Transformer Neural Network and Technical Indicators. ArXiv, abs/2403.03606.
[23] Ladhari, A., & Boubaker, H. (2024). Deep Learning Models for Bitcoin Prediction Using Hybrid Approaches with Gradient-Specific Optimization. Forecasting.
[24] Bouslimi, J., Boubaker, S., & Tissaoui, K. (2024). Forecasting of Cryptocurrency Price and Financial Stability: Fresh Insights based on Big Data Analytics and Deep Learning Artificial Intelligence Techniques. Engineering, Technology & Applied Science Research, 14(3), 14162–14169. https://doi.org/10.48084/etasr.7096.
[25] R.Bourday, I.Aatouchi, M.AitKerroum & A.Zaaouat. (2024),"Cryptocurrency Forecasting UsingDeep Learning Models:A Comparative Analysis,"HighTech and Innovation Journal Vol. 5, No.4, December, 2024.Doi: 10.28991/HIJ-2024-05-04-013.
[26] Glenski, M., Weninger, T., & Volkova, S. (2019). Improved Forecasting of Cryptocurrency Price using Social Signals. arXiv preprint arXiv:1907.00558.
[27] Nazmiye Ceren Abay, Cuneyt Gurcan Akcora, Yulia R. Gel, Umar D. Islambekov, Murat Kantarcioglu, Yahui Tian, Bhavani Thuraisingham. "ChainNet: Learning on Blockchain Graphs with Topological Features." To Appear in the 2019 IEEE International Conference on Data Mining (ICDM).
[28] Ider, D., & Lessmann, S. (2022). Forecasting Cryptocurrency Returns from Sentiment Signals: An Analysis of BERT Classifiers and Weak Supervision. arXiv preprint arXiv:2204.05781.
[29] Fan Fang, Waichung Chung, Carmine Ventre, Michail Basios, Leslie Kanthan, Lingbo Li, Fan Wu."Ascertaining price formation in cryptocurrency markets with Deep Learning." arXiv preprint arXiv:2003.00803.
[30] Frohmann, M., Karner, M., Khudoyan, S., Wagner, R., & Schedl, M. (2023). Predicting the Price of Bitcoin Using Sentiment-Enriched Time Series Forecasting. Big Data and Cognitive Computing, 7(3), 137. https://doi.org/10.3390/bdcc7030137.

Authors' Profiles

Rachid Bourday is a doctoral student in the Department of Computer Engineering, Faculty of Science, Ibn Tufail University, Morocco. He holds a bachelor's degree in Statistics and Econometrics from Mohamed V University, Morocco. He is currently serving as a mathematics teacher at the middle school level. His research interests include artificial intelligence, machine learning, deep learning, and the application of machine learning in finance.

Prof. Dr. Issam Aattouchi is an assistant Professor in National School of Business and Management, Ibn Tofail University, Morocco. He obtained his state engineer diploma in Computer Engineering from Cadi Ayyad University, Morocco, and he gained his Ph.D. in computer science from Ibn Tofail University. His research interests include artificial intelligence, big data analytics, machine learning, human machine interfacing, and application of machine learning in finance.

Prof. Dr. Mounir Ait Kerroum is a Higher education Professor in National School of Business and Management, Ibn Tofail University, Morocco. He obtained his M.Eng. in Computer Science and Telecommunications, and he gained his Ph.D. in IT and Telecommunications from Mohammed V University of Rabat, Morocco. His research interests include artificial intelligence, pattern recognition, deep learning, classification of hyperspectral images, classification of medical images, recognition of Arabic manuscript text.

How to cite this paper: Rachid Bourday, Issam Aattouchi, Mounir Ait Kerroum, "A Transfer Learning–Enhanced Hybrid Deep Learning Framework for Bitcoin Price Forecasting Using Market Sentiment and Time Series Data", International Journal of Information Engineering and Electronic Business(IJIEEB), Vol.17, No.5, pp. 1-17, 2025. DOI:10.5815/ijieeb.2025.05.01