Integrating Blockchain and Federated Learning for Cryptocurrency Market Prediction: Major Exchanges as Nodes

Zijie Wang
School of Advanced Technology
Xi'an Jiaotong-Liverpool University
Suzhou, China
zijie.wang23@student.xjtlu.edu.cn

Ziyi Guo
Business School
University of Sydney
Sydney, Australia
zguo0742@uni.sydney.edu.au

Wanxin Li*
School of Advanced Technology
Xi'an Jiaotong-Liverpool University
Suzhou, China
wanxin.li@xjtlu.edu.cn
*Corresponding author

Jie Zhang*
School of Advanced Technology
Xi'an Jiaotong-Liverpool University
Suzhou, China
jie.zhang01@xjtlu.edu.cn
*Corresponding author

Hao Guo*
School of Software
Northwestern Polytechnical University
Xi'an, China
haoguo@nwpu.edu.cn
*Corresponding author

2025 8th International Conference on Artificial Intelligence and Big Data (ICAIBD) | 979-8-3315-1936-0/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICAIBD64986.2025.11081867

***Abstract*—This study introduces a framework that integrates the Hyperledger Fabric blockchain with federated learning to improve cryptocurrency market prediction. Using three major exchanges as distributed nodes, the platform processes trading data and sentiment analysis locally, training machine learning models on each node. The results show that the federated model achieves a prediction deviation of 0. 65% from the actual prices, exceeding the deviation of the centralized LSTM model of 3. 35%. The Hyperledger Fabric network also handles up to 298.7 TPS with zero transaction failures and low latency (0.01s), highlighting the model's effectiveness for secure and accurate market prediction in the fintech sector.**

***Index Terms*—Hyperledger Fabric, Federated Learning, Cryptocurrency Market Prediction, Sentiment Analysis, Data Security**

I. Introduction

Blockchain technology enhances data security and transparency, with Hyperledger Fabric being a leading enterprise blockchain supporting complex workflows [1]. Blockchain applications extend beyond traditional use cases to decentralized systems like traffic signal control and secure event recording [2], [3]. Hierarchical and location-aware consensus protocols further improve blockchain efficiency, especially in IoT applications [4].

Cryptocurrency markets are highly volatile due to factors like economic indicators and sentiment. Traditional predictive models often fail to capture this complexity, leading to suboptimal predictions. Sentiment analysis, leveraging data from social media and news, has demonstrated potential in improving prediction accuracy by incorporating real-time market sentiments [5], [6].

This study proposes a novel solution combining federated learning, blockchain, and Long Short-Term Memory (LSTM)-based sentiment analysis to address these challenges. Federated learning ensures decentralized model training while preserving privacy, and blockchain enhances data security and transparency. LSTM networks effectively capture temporal dependencies in financial data, while sentiment analysis enriches the model with real-time contextual information. This paper makes the following contributions:

- A novel framework integrating Hyperledger Fabric, federated learning, and LSTM-based sentiment analysis for cryptocurrency market prediction, addressing privacy and security concerns.
- An innovative federated learning approach that aggregates local LSTM models, improving global model accuracy while accommodating non-IID financial data.
- Implementation and experimental evaluation using real-world data from major cryptocurrency exchanges, demonstrating the impact of sentiment analysis on predictive accuracy and blockchain performance metrics.

The paper is structured as follows: Section II reviews related work, Section III defines system architecture, Section IV describes the methodology and results, Section V presents block chain platform stress test results, and Section VI concludes the study.

II. Literature Review

Blockchain technology has been extensively studied for enhancing data security, transparency, and accountability. Hyperledger Fabric, a leading enterprise blockchain, supports modular workflows and is widely used in industries requiring stringent data protection. Blockchain applications extend beyond traditional use cases to decentralized systems such as traffic signal control and secure event recording [2], [3]. Hybrid blockchain architectures further address scalability and privacy concerns, especially in healthcare data management and adaptive traffic prediction [7], [8].

Federated learning has emerged as a key framework for collaborative model training without sharing raw data, ensuring privacy in domains like finance and healthcare. Its integration with blockchain enhances data security and ensures trust, with applications in adaptive traffic prediction and healthcare systems [9], [8]. However, federated learning faces challenges such as handling non-independent and identically distributed (non-IID) data and managing high communication overhead, which limit scalability [10], [11].

Sentiment analysis plays a pivotal role in financial predictions, leveraging data from social media and news to capture market trends and investor behavior [12], [5]. In the context of NFTs, pricing mechanisms driven by rarity and market dynamics have been proposed to better reflect intrinsic and extrinsic value [13], [14]. Studies show that sentiment analysis can significantly improve prediction accuracy, especially when combined with machine learning models [6].

LSTM networks are particularly effective for time series predictions in volatile financial markets, as they excel at modeling sequential data and capturing long-term dependencies [15], [16]. By incorporating sentiment analysis, LSTM-based models combine historical trends with real-time contextual insights, further enhancing predictive capabilities [17].

The integration of blockchain, federated learning, and LSTM networks presents a promising approach for financial market prediction. Blockchain ensures secure data sharing, federated learning preserves privacy while enabling collaborative model training, and LSTM networks capture complex temporal patterns. However, challenges remain, including data synchronization, managing non-IID data, and balancing communication efficiency with model accuracy [10], [11]. Addressing these issues is critical for developing robust and scalable predictive frameworks.

III. System Architecture

In this section, we present the architecture design of our proposed framework, which integrates Hyperledger Fabric with federated learning and LSTM-based sentiment analysis for cryptocurrency market prediction. The architecture ensures data security, privacy, and scalability while providing accurate market predictions. The proposed system, as illustrated in Figure 1, consists of a Hyperledger Fabric network with four nodes representing three major cryptocurrency exchanges and a council node.

A. System Overview

1) Hyperledger Fabric Network: The Hyperledger Fabric network is designed with the following nodes:

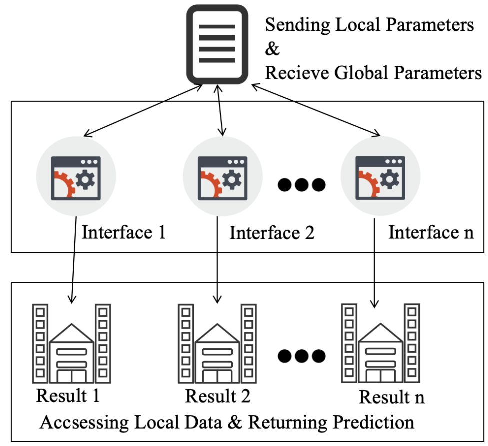

Fig. 1. System Architecture for Integrating Hyperledger Fabric with Federated Learning.

- **Council Node**: Responsible for maintaining the blockchain and managing node registration and consensus. It utilizes the default Raft consensus mechanism provided by Hyperledger Fabric.
- **Exchange Nodes**: Each node represents a major cryptocurrency exchange and acts as a client in the federated learning framework. These nodes are responsible for local model training and uploading local parameters for aggregation.

All nodes run the approved chaincode, which facilitates the federated learning process by aggregating and broadcasting global parameters. Quantitative team from the exchanges can use these parameters to gain predicted price in order to assist trading. Figure 1 illustrates the proposed system architecture. As shown in the figure, the system integrates Hyperledger Fabric with federated learning, where each node represents a major cryptocurrency exchange. The global model parameters are distributed across these nodes and the local parameters are securely uploaded and aggregated to update the global model.

2) Federated Learning Process: The federated learning process is as follows.

- **Upload Parameters**: Users upload the necessary parameters for model training and prediction.
- **Parameter Verification**: The system checks if the parameters are complete. If incomplete, it requests the missing parameters from the user.
- **Chaincode Execution**: The chaincode randomly selects a node and broadcasts a global notification.
- **Model Aggregation**: The selected node initiates the model aggregation process and writes the results to the ledger, which is then globally broadcasted.
- **Training Simulation**: Each node retrains using the updated global parameters and compares the training performance with the termination conditions. If the conditions are not met, the system reverts to the parameter upload step.
- **Training Termination**: If the conditions are met, the

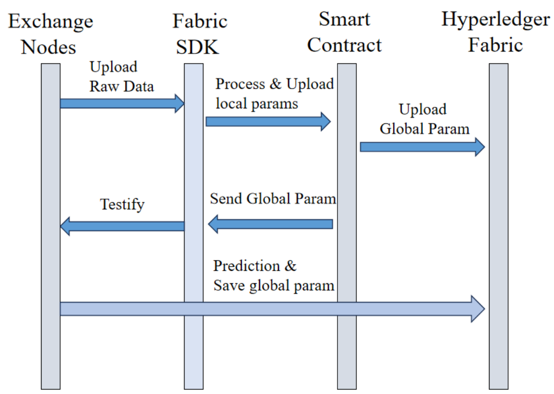

Fig. 2. Workflow of Federated Learning with Hyperledger Fabric for Cryptocurrency Prediction.

system terminates the training process.

B. Federated Learning with Hyperledger Fabric

To enhance security and prevent parameter leakage, our system employs random node selection for aggregation. In each training round, the current model parameters are broadcast to all nodes, which train on local data and upload updated parameters to the blockchain. A randomly selected node aggregates these parameters to form the global model, which is shared with all nodes for the next round. Validation occurs locally at each node, with metrics sent to the blockchain. The process terminates upon convergence; otherwise, training continues. Security features include data encryption, privacy-preserving computation, and blockchain verification for an immutable audit trail. Figure 2 illustrates the workflow, and Algorithm 1 outlines the steps.

Algorithm 1 Federated Learning Process with Hyperledger Fabric

Require: Global model parameters w_0

for each round $t = 1, 2, \dots, T$ **do**
  Broadcast w_{t-1} to all nodes
  for each node i in parallel **do**
    Node i updates local parameters $w_{t,i}$ and uploads them to the blockchain
  end for
  Randomly select a node for aggregation
  Aggregate local parameters into global model: $w_t = \frac{1}{N}\sum_{i=1}^{N} w_{t,i}$
  Update global model on the blockchain
  if termination conditions are met **then**
    break
  end if
end for

C. LSTM Model Construction with Sentiment Analysis

We utilize a Long Short-Term Memory (LSTM) model to predict asset closing prices by capturing temporal dependencies in sequential data. The simplified form is defined below, and model architecture includes:

- **Input Layer**: Sequences of length T, each containing six features—high, low, close, amplitude, sentiment score, and nasdaq_close.
- **LSTM Layer**: A single layer with 32 hidden units to model temporal patterns.
- **Fully Connected Layer**: Maps the LSTM output to a single predicted closing price.

$$h_0 = \mathbf{0}, \quad c_0 = \mathbf{0} \tag{1}$$

$$\text{out}, (h_n, c_n) = \text{LSTM}(x, (h_0, c_0)) \tag{2}$$

$$\hat{y} = \text{Linear}(\text{out}_{-1}) \tag{3}$$

1) Data Preprocessing: Features are normalized to the [0, 1] range using `MinMaxScaler`. Steps include:

- Loading the dataset and selecting required features.
- Normalizing features for stable training.
- Creating sequences of specified length.
- Splitting data into training (80%) and validation (20%) sets.

2) Training and Evaluation: The model is trained over 10 epochs using Mean Squared Error (MSE) loss and the Adam optimizer (learning rate 0.001).

3) Results Visualization and Model Export: Training and validation losses are plotted to assess learning progress. The initialized model parameters are exported in JSON format for use in federated learning across different clients (e.g., COINBASE, BINANCE, OKX).

4) Integration of Sentiment Analysis: Acknowledging the impact of public sentiment on cryptocurrency prices, we incorporate sentiment analysis into the LSTM model:

- **Data Collection**: Daily Bitcoin-related tweets are gathered from Twitter.
- **Sentiment Scoring**: Randomly select 20 tweets per day and analyze sentiment using TextBlob.
- **Data Processing**: Calculate average daily sentiment scores; interpolate missing data.
- **Feature Integration**: Add sentiment scores as an additional feature in the LSTM model.

For dates lacking sufficient data, linear interpolation between neighboring dates estimates sentiment scores.

Algorithm 2 Daily Sentiment Analysis with Interpolation

function ANALYZEANDINTERPOLATESENTIMENT(date, sentimentData)
  tweets ← FetchTweets(date, keyword="Bitcoin", count=20)
  sentiments ← [TextBlob(tweet).polarity for tweet in tweets]
  averageSentiment ← Mean(sentiments)
  if averageSentiment is None **then**
    averageSentiment ← Interpolate(sentimentData, date)
  end if
  sentimentData[date] ← averageSentiment
  return sentimentData
end function

By integrating sentiment analysis, the model captures the influence of public perception on Bitcoin price movements, potentially enhancing predictive accuracy. This approach, combined with federated learning on Hyperledger Fabric, offers a secure and scalable solution for cryptocurrency market prediction. Future work will focus on optimizing model training and improving system scalability and robustness.

IV. Experiment and Evaluation

A. Experimental Setup and Procedure

The federated learning system was implemented on a Hyperledger Fabric network with three nodes, each representing a client. The experiments were carried out on a local machine equipped with an Intel® Core™ i5 CPU, 7.4 GiB of memory, and integrated Intel® UHD Graphics.

The federated learning process proceeded as follows:

1) Initialize the global model parameters.
2) Repeat until the termination criterion is met:
    - Distribute the global model parameters to all clients.
    - Each client trains the model on its local dataset.
    - Clients upload their updated parameters to the network.
    - Aggregate the client parameters to update the global model.
    - Evaluate the global model against the termination criterion.

The process terminated when the mean squared error (MSE) on the validation set fell below 0.002.

B. Federated Learning Implementation

The federated learning algorithm was implemented using PyTorch for local training and Hyperledger Fabric for secure parameter sharing. The following pseudo code illustrates the main logic of the federated learning process:

Algorithm 3 Federated Learning Process

Require: Initialize global model parameters θ_0

for each round $r = 1, 2, \ldots, T$ **do**
    Broadcast current model θ_{r-1} to all clients
    for each client i in parallel **do**
        Update local model θ_i^r using local data
        Upload θ_i^r to Fabric network
    end for
    Aggregate updates to obtain global model θ_r
    if Termination criteria met **then**
        break
    end if
end for
return Final global model θ_T

This implementation demonstrates the recursive nature of federated learning, where global parameters are iteratively refined through repeated local training and aggregation steps until the desired performance is achieved.

1) Feature Correlation Analysis: First, we conducted a correlation analysis of data features across all exchanges. From Figure 3, we can observe that strong positive correlations between open, high, low, and close prices, which is expected as they all reflect price information. Moderate positive correlations between volume and both price change and volume change, indicating that changes in trading volume may signal price fluctuations. Positive correlations between amplitude and both price change and volume change, suggesting that market volatility is related to changes in price and trading volume. These correlations provide valuable input features for our LSTM model.

2) Prediction Performance Evaluation: Next, we present the prediction results for three clients (corresponding to three different exchanges). Figures 4, 5, and 6 demonstrate the model's prediction performance for each client. We can observe that:

The model captures the overall trend of price movements across all three clients. Predictions closely follow the true values, especially during periods of relative stability. The model shows some lag in capturing sudden price spikes or drops, which is common in time series prediction models. Client_1's data spans a longer time period, allowing the model to learn from a more diverse range of market conditions. The prediction accuracy appears to be consistent across all three clients, suggesting that the federated learning approach has successfully leveraged data from multiple exchanges without compromising individual privacy. These results indicate that our federated learning approach, combined with LSTM and sentiment analysis, provides robust and accurate cryptocurrency price predictions across different exchanges. The model's ability to capture both long-term trends and short-term fluctuations demonstrates its potential for real-world application in cryptocurrency trading and market analysis.

To compare the results of the centralized LSTM model, we also implemented a Federated Learning model using three clients, each trained separately, and predicted the price for the last day. The predictions from the three clients are as follows:

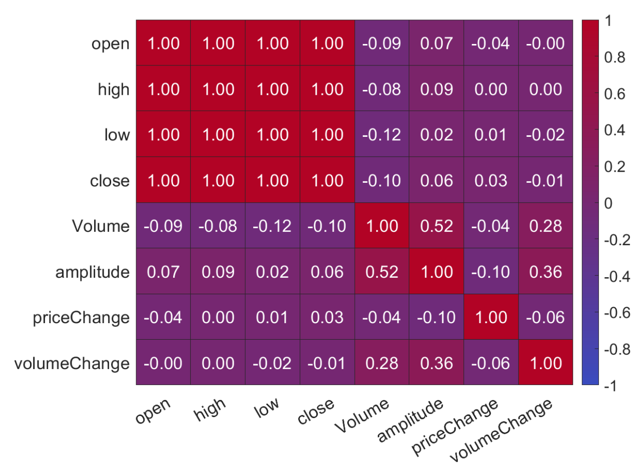

Fig. 3. Average Correlation Heat-map of Features Across All Exchanges

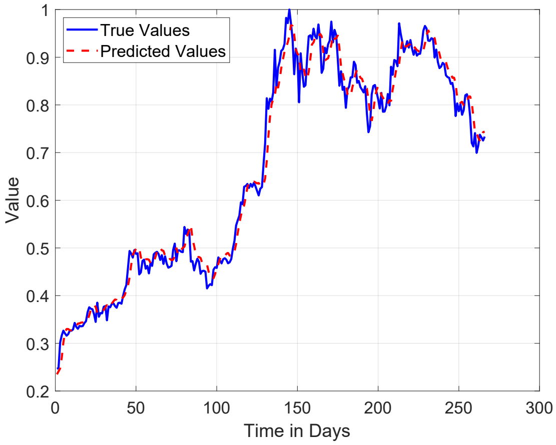

Fig. 4. True vs Predicted Prices for client_3

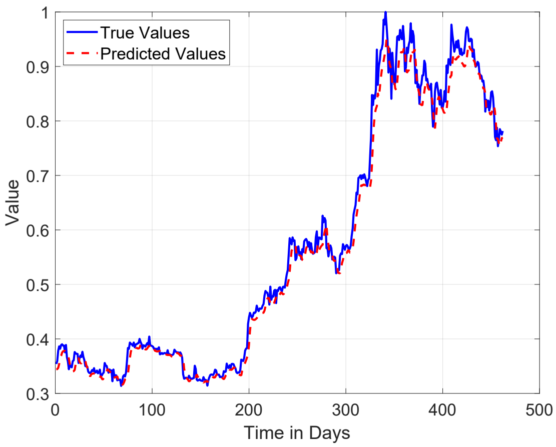

Fig. 5. True vs Predicted Prices for client_2

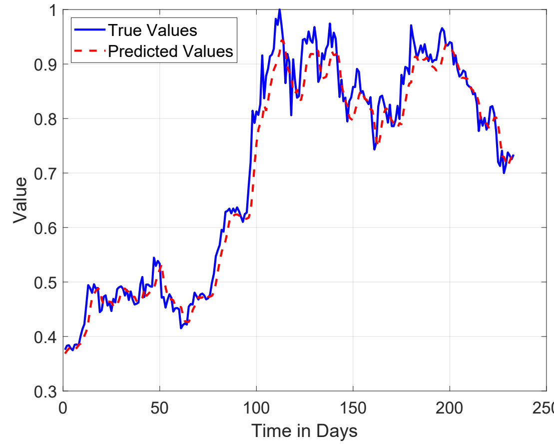

Fig. 6. True vs Predicted Prices for client_1

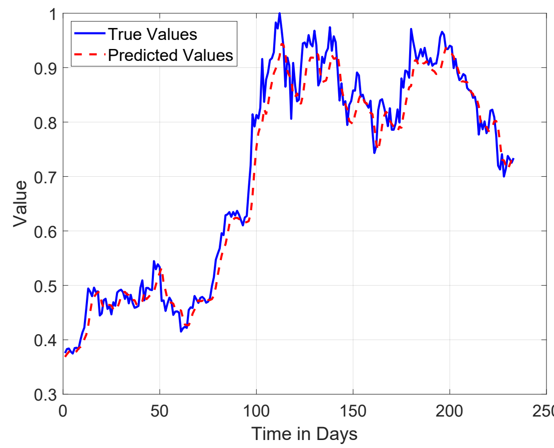

Fig. 7. Comparison of Bitcoin prices predicted by the LSTM model with actual prices

Last Predicted Price for client 1: **59965.57**, Last Predicted Price for client 2: **58887.02**, Last Predicted Price for client 3: **55727.42**.

C. LSTM Model Implementation

In this section, we implemented an LSTM model to predict Bitcoin prices. After training the model with the training set, the time interval is from **2018** to **2024**, we used it to make predictions on the test data, and the results are shown below:

As shown in Figure 7, the blue line represents the actual Bitcoin prices, while the orange line represents the prices predicted by the LSTM model. Although the predicted results generally follow the trend of the actual prices, there are some discrepancies during periods of high volatility. This figure shows that the LSTM model predicted the price for the last day to be **$59,750.05** .

D. Analysis and Summary

In the comparison between the LSTM model and the Federated Learning model, the centralized LSTM model predicted the last day's price to be approximately 59,750.05 USD, while the average prediction from the three clients in the Federated Learning setup was approximately 58,193.34 USD. The difference between the LSTM model's prediction and the average prediction from the Federated Learning model was about 2.61%.

When comparing these predictions with the actual price (which is the average of three exchanges, approximately **$57,814.70**), the Federated Learning model provided a more accurate prediction, with a difference of about 0.65%. In contrast, the LSTM model's prediction deviated more significantly, with a difference of approximately 3.35%.

These results suggest that the Federated Learning model, which aggregates knowledge from decentralized sources, provided a closer prediction to the actual market prices compared to the centralized LSTM model. This highlights the potential benefits of incorporating Federated Learning, especially in scenarios where data from multiple decentralized sources can lead to more accurate and robust predictions.

V. Performance Analysis of Hyperledger Fabric Network

This section presents a comprehensive performance analysis of our Hyperledger Fabric network under various workload scenarios. The tests were designed to evaluate the network's performance, scalability, and stability under different conditions, including steady-state, burst traffic, stress, and long-running scenarios. Results are shown in Tables I and II.

TABLE I
SUMMARY OF PERFORMANCE METRICS (PART 1)

Scenario	Succ	Fail	Send Rate (TPS)
steady-state-240tps	9008	3	44.2
burst-traffic	8964	0	298.7
stress-test	15766	3	111.2
long-running-test	55262	0	184.2

TABLE II
SUMMARY OF PERFORMANCE METRICS (PART 2)

Scenario	Max Lat. (s)	Min Lat. (s)	Avg Lat. (s)	Throughput (TPS)
steady-state-240tps	160.82	0.00	0.19	44.1
burst-traffic	0.28	0.00	0.01	298.7
stress-test	0.40	0.00	0.01	111.2
long-running-test	0.60	0.00	0.01	184.1

A. Test Scenarios and Result Analysis

Four distinct test scenarios were executed to assess the network's performance:

1) Steady-state test at 240 TPS
2) Burst traffic test
3) Stress test with increasing load
4) Long-running test at moderate TPS

The burst traffic test demonstrated the network's capability to handle high transaction rates, achieving an impressive throughput of 298.7 TPS with no failed transactions. This performance showcases the network's ability to efficiently process sudden spikes in transaction volume, which is crucial for applications with variable workloads. In the long-running test, the network exhibited remarkable stability, processing 55,262 successful transactions over a 300-second period with no failures. The sustained throughput of 184.1 TPS with an average latency of just 0.01 seconds underscores the network's reliability for extended operations under moderate load. Detail information is shown in Table I. The stress test revealed the network's resilience to gradually increasing load. Despite not reaching the target TPS range, the system maintained a respectable throughput of 111.2 TPS while keeping the average latency at a low 0.01 seconds. This performance indicates the network's ability to gracefully handle increasing demands without significant degradation in response times. While the steady-state test at 240 TPS presented some challenges, it provided valuable insights into the network's behavior under sustained high load, identifying areas for potential optimization. Detail information is shown in Table II.

VI. Conclusion

In this study, we presented an innovative framework that integrates Hyperledger Fabric blockchain technology with federated learning to enhance cryptocurrency market prediction. Our architecture, comprising three cryptocurrency exchanges as distributed nodes and a council node, demonstrates the feasibility of secure, decentralized market prediction using LSTM-based models and sentiment analysis. Experimental results show that our network effectively handles diverse workloads, achieving up to 298.7 TPS in burst traffic with zero failed transactions. The federated learning model achieved superior accuracy with a 0.65% deviation from actual prices, compared to 3.35% for a centralized LSTM model.

References

[1] E. Androulaki *et al.*, "Hyperledger fabric: A distributed operating system for permissioned blockchains," in *Proceedings of the Thirteenth EuroSys Conference*, 2018, pp. 1–15.

[2] K. Lee *et al.*, "A blockchain-based architecture for traffic signal control systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 6, pp. 5247–5257, 2022.

[3] L. Tan and G. Wang, "Proof-of-event recording system for autonomous vehicles: A blockchain-based solution," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 6, pp. 4256–4267, 2021.

[4] H. Guo, W. Li, and M. Nejad, "A hierarchical and location-aware consensus protocol for iot-blockchain applications," *IEEE Transactions on Network and Service Management*, vol. 19, no. 3, pp. 2972–2986, 2022.

[5] Y. Kim *et al.*, "Predicting fluctuations in cryptocurrency transactions based on user comments and replies," *PLOS ONE*, vol. 12, no. 8, p. e0180945, 2017.

[6] Z. Chen *et al.*, "Sentiment analysis for financial news and its application to stock movement prediction," *Quantitative Finance*, vol. 19, no. 9, pp. 1397–1405, 2019.

[7] Y. Luo *et al.*, "A hybrid blockchain-edge architecture for electronic health records management with attribute-based cryptographic mechanisms," *IEEE Access*, vol. 7, pp. 105 678–105 689, 2019.

[8] H. Guo *et al.*, "B2sfl: A bi-level blockchained architecture for secure federated learning-based traffic prediction," *IEEE Transactions on Services Computing*, vol. 16, no. 6, pp. 4360–4374, 2023.

[9] S. Wang *et al.*, "Adaptive traffic prediction at the its edge with online models and blockchain-based federated learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 7, pp. 4470–4482, 2021.

[10] H. B. McMahan *et al.*, "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*. PMLR, 2017, pp. 1273–1282. [Online]. Available: http://proceedings.mlr.press/v54/mcmahan17a.html

[11] T. Li *et al.*, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, 2020.

[12] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, no. 1, pp. 1–8, 2011.

[13] W. Xiong, Y. Wang, W. Li, Y. Zhang, J. Zhang, and H. Guo, "An advanced pricing mechanism for non-fungible tokens (nfts) based on rarity and market dynamics," *IEEE Transactions on Computational Social Systems*, 2024, early Access.

[14] W. Xiong, Y. Wang, W. Li, J. Zhang, and H. Guo, "Pricing mechanism of non-fungible token (nft) driven by rarity design," in *IEEE International Conference on Blockchain (Blockchain 2023)*, Hainan, China, December 17–21 2023.

[15] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[16] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *European Journal of Operational Research*, vol. 270, no. 2, pp. 654–669, 2018.

[17] M. M. Akrout and A. Moussa, "Predicting stock market movement direction with sentiment analysis and lstm neural networks," in *IEEE International Conference on Intelligent Systems*, 2020, pp. 1–6.