Journal of Intelligent Management Decision
https://www.acadlore.com/journals/JIMD

Enhancing Stock Market Forecasting Through Deep Learning and Decentralized Data Integrity: A Blockchain-Integrated Framework

Safiye Turgay[1*], Abdulkadir Aydin[1], Suat Erdoğan[2], Metin Yıldırım[3], Mustafa Kavacık[3]

[1] Industrial Engineering, Sakarya University, 54187 Sakarya, Turkey
[2] BilgeAdam Technology, 34467 İstanbul, Turkey
[3] International Trade and Finance, Faculty of Applied Sciences, Necmettin Erbakan University, 42090 Konya, Turkey

[*] Correspondence: Safiye Turgay (sencer@sakarya.edu.tr)

Received: 03-25-2025 **Revised:** 05-14-2025 **Accepted:** 05-18-2025

Citation: S. Turgay, A. Aydin, S. Erdoğan, M. Yıldırım, and M. Kavacık, "Enhancing stock market forecasting through deep learning and decentralized data integrity: A blockchain-integrated framework," *J. Intell Manag. Decis.*, vol. 4, no. 2, pp. 118–136, 2025. https://doi.org/10.56578/jimd040203.

Abstract: The reliability and precision of stock market forecasting are of paramount importance to investors, regulatory authorities, and financial institutions. Traditional centralized systems for data processing and model deployment have been found to suffer from critical vulnerabilities, including susceptibility to tampering, single points of failure, and a lack of verifiability. To address these limitations, a novel hybrid framework has been developed that integrates advanced deep learning models with decentralized blockchain infrastructure to ensure both predictive accuracy and data integrity in financial time series forecasting. Temporal dependencies in market dynamics are captured through the use of recurrent neural networks (RNNs) and long short-term memory (LSTM) architectures, which have been extensively trained to model non-linear and non-stationary behaviors in high-frequency financial data. In parallel, a private Ethereum-based blockchain has been deployed to record cryptographic hashes of input datasets, model parameters, and forecasting outputs, thereby ensuring transparency, auditability, and immutability across the data lifecycle. To enable computational scalability, deep learning operations have been executed off-chain, while on-chain mechanisms are utilized for secure checkpointing and traceability. Empirical validation has been conducted using real-time data from the Borsa İstanbul (BIST), demonstrating significant improvements in forecasting accuracy when compared with baseline statistical and machine learning (ML) models. Moreover, the integration of blockchain technology has enabled a verifiable audit trail for all predictive operations, enhancing trust in the data pipeline without compromising computational efficiency. The proposed framework represents a significant advancement towards secure, transparent, and trustworthy artificial intelligence (AI) in financial forecasting, with potential implications for the broader decentralized finance (DeFi) ecosystem and regulatory-compliant AI deployments in capital markets.

Keywords: Stock market forecasting; Deep learning; Blockchain; Recurrent neural networks (RNN); Long short-term memory (LSTM); Financial time series; Tamper-proof data; Trustworthy artificial intelligence (AI); Decentralized finance (DeFi)

1 Introduction

Stock market trend prediction is a highly complex and dynamic issue that has garnered substantial attention from finance researchers, data scientists, and AI experts. Prediction of stock prices can lead to huge profits and good investment decisions. Deep learning techniques, particularly RNN and LSTM models, have been shown over the past few years to perform better in extracting temporal dependencies and nonlinear patterns in financial time series data. As much as the models offer improved predictive capabilities, their accuracy is very much dependent on the quality, trustworthiness, and integrity of the input data.

Traditional stock market forecasting systems typically tap centralized data stores and model management infrastructure. Centralized systems are subject to data manipulation, unauthorized use, single points of failure, and transparency issues. These flaws pose serious risks in high-risk financial settings where minute data alterations can lead to vulnerability errors and financial losses. With financial ecosystems more and more requiring transparency, security, and accountability, there is a need for strong architectures that provide data integrity along the forecasting pipeline.

https://doi.org/10.56578/jimd040203

Blockchain technology, being decentralized, tamper-proof, and transparent in its ledger system, offers a highly promising remedy for these problems. It can be facilitated by combining blockchain with deep learning models to enable a tamper-proof data storage, model update, and result validation mechanism. Using this combination, the efficiency of stock market prediction systems can be enhanced by authenticating all the data used in training, validation, and prediction.

In this paper, we propose a new paradigm combining deep learning and blockchain for decentralized data integrity in stock market forecasting. The system employs LSTM-based forecasting and utilizes a private Ethereum blockchain to log cryptographic hashes of financial information and model output. Verification processes are automated using smart contracts and enforced using trust among nodes. The integration of off-chain deep learning computation with on-chain data integrity mechanisms renders the system scalable without sacrificing security and transparency.

Using recent data from the BIST, we demonstrate the efficacy of our method in enhancing predictive accuracy as well as the verifiability of data. Our results suggest that a decentralized forecasting system can potentially significantly increase stakeholders' confidence and propel trustworthy AI applications in finance. This study contributes to the growing literature on DeFi and opens new avenues for secure and intelligent financial decision-making.

2 Literature Survey

Prediction of the stock market has been conventionally examined using statistical models such as ARIMA, GARCH, and linear regression. However, these models fail to capture the nonlinearity and volatility of the financial markets. Deep learning methods, particularly RNN and LSTM networks, have gained a lot of popularity since they are able to capture time-dependent patterns and sequential data. The uses of AI, natural language processing (NLP), and ML within finance have profoundly reshaped forecasting, decision-making, and risk modeling. In a recent and comprehensive survey, Du et al. [1] reviewed financial NLP applications. Their paper categorizes financial NLP research into sentiment analysis, event extraction, and comprehending financial narratives. The paper highlights advancements in large-scale financial datasets and pre-trained language models, with emphasis on the heightened relevance of NLP in automated trading, news-based prediction, and risk measurement.

In another field-specific study, Mostafavi and Hooman [2] identified key technical indicators driving stock market predictions with power. Their work entails the measurement of moving averages (MA), momentum oscillators, and Bollinger Bands indicators' impact if combined with ML algorithms. Outcomes emphasize feature engineering's importance in achieving firm model performance in illustrating the usefulness of technically rich datasets for facilitating the predictiveness of ML models. Centering on the optimization of financial systems, Tang et al. [3] put forward a profit prediction model founded on a Deep LSTM (DLSTM) framework integrated into financial accounting systems. The approach leverages historical financial data to improve profit prediction. By optimizing the DLSTM framework, the model supports higher prediction accuracy, demonstrating the practical value of deep learning in financial information systems. At the theoretical level, Booker et al. [4] spoke about possibilities using ML methods. Their study bridges the gap between accounting theory and new AI methods. The paper proposes a meta-theoretical model that maps ML algorithms to core AIS research streams, offering a roadmap for researchers to follow data-driven innovation in auditing, fraud detection, and financial control. Expanding the range of technologies, Fazel et al. [5] addressed the convergence of IoT, ML, and blockchain technologies. It covers applications in smart finance, logistics, and cyber-physical systems with an emphasis on interoperability concerns and data integrity. This kind of convergence is particularly relevant to DeFi, where IoT device data in real time is fed into secure and intelligent blockchain networks. In cryptocurrency prediction, a systematic review of Bitcoin price prediction through ML methods was conducted. Allen and Barbalau [6] discussed the theoretical underpinnings of security design in finance.

Legacy systems store data in centralized servers, Karamchandani et al. [7], who proposed a lower approximation-based integrated decision analysis framework for blockchain-enabled supply chains. In blockchain sustainability, Monem et al. [8] offered an innovative solution to reduce the energy consumption of Bitcoin networks. Panigrahi et al. [9] proposed ASBlock, a blockchain-based supply chain management system for agriculture.

In financial industries, blockchain has been applied in transaction verification, fraud detection, auditing, and asset tokenization. The convergence of blockchain technology with emerging paradigms like AI, Internet of Things (IoT), and smart contracts has spurred innovation in diverse fields of application. In their survey article, Hewa et al. [10] spoke about the prospects of blockchain-based smart contracts, Bothra et al. [11] provided an insight into how the technologies collectively enhance system intelligence, transparency, and security. Valsan et al. [12] proposed a conceptual framework of AI-powered blockchain micromarkets to facilitate sustainable energy consumption. Pattanayak et al. [13] emphasized the need for trust and relational capabilities in realizing the full benefits of blockchain for supply chain performance. Wang et al. [14] examined how the implementation of blockchain improves the quality of analyst forecasts by mitigating information asymmetry and the credibility of financial reporting. Dubey and Kumar [15] surveyed the convergence of Explainable AI (XAI) and Federated Learning (FL) for IoT, emphasizing decentralized intelligence without the loss of data privacy. N. et al. [16] proposed pharmaceutical supply chains. Alam et al. [17] presented a systematic review of global blockchain initiatives, emphasizing technical challenges such as scalability, energy consumption, and integration with legacy systems. Wang et al. [18] presented a review of literature synthesizing blockchain's socio-organizational impacts, and they theorized a conceptual framework capturing its role in reshaping business ecosystems, governance arrangements, and stakeholder relations. Similarly, Dutta et al. [19] mapped blockchain applications across supply chain operations, identifying prospective research avenues in real-time monitoring, smart contracts, and green logistics. In food traceability and safety, Sun et al. [20] suggested βFSCM, a blockchain- and recommender-system-supported food supply chain system. Their combined model ensures consumer trust and process transparency through smart contracts and personalized recommendations. The disruptive potential of blockchain in smart cities was also studied by Singh et al. [21], with a particular emphasis on its convergence with AI and IoT for sustainable city growth. In finance, Wang et al. [22] proposed a blockchain-based LSTM model for stock prediction, enhancing data integrity and model explainability in a decentralized system. In autonomous vehicle supply chains, Arunmozhi et al. [23] explored the application of blockchain and smart contracts for reliability enhancement and automation. Emphasizing peer-to-peer energy trading, Shukla et al. [24] depicted the use of blockchain and ML in streamlining energy distribution and market fairness. Barati [25] employed a system dynamics approach to model how blockchain enhances demand forecasting in supply chains, offering dynamic simulation tools to evaluate the long-term effects of digitalization. Vijayakumar et al. [26] also suggested a privacy-preserving blockchain framework for decentralized swap derivatives with deep learning oracles and cryptography methods, allowing secure financial transactions in DeFi. In AI applications, blockchain has been a promising solution to improve data trustworthiness. ModelChain and BlockAI are projects that incorporate blockchain to log model updates and data hashes so that stakeholders can ensure the integrity of inputs and outputs. Smart contracts also allow rules to be automatically validated and enforced without human intervention.

3 Methodology

This section presents the proposed hybrid framework integrating deep learning and blockchain technologies to achieve secure and accurate stock market forecasting. The methodology is structured into three main components: (1) data collection and preprocessing, (2) deep learning-based forecasting using LSTM, and (3) blockchain integration for data integrity and model transparency.

Past and real-time stock exchange information was received from BIST that included the most critical attributes, such as opening, closing, highest, and lowest price, volume of trading, and some technical factors like MA and Relative Strength Index (RSI). The collected dataset went through several preprocessing phases prior to training the LSTM model. Missing value treatment was performed by applying linear interpolation and forward filling techniques to obtain a complete dataset. Normalization was performed based on min-max scaling to transform the input feature into a normal range to facilitate effective training of the LSTM model. In addition, windowing was employed in the creation of sliding time windows to transform the time series data into supervised learning sequences so that the LSTM model is able to learn temporal relationships and make quality predictions based on past trends.

An LSTM network was employed for time-series forecasting since it has been shown to effectively capture temporal dependencies and long-range structure in sequential data. The model architecture consists of an input layer that receives windowed time-series sequences, two LSTM layers meant to learn complex temporal relationships, and dropout layers to prevent overfitting. The output layer contains a single node that will predict the stock price the following day. The loss function used to train it was Mean Squared Error (MSE) so that there are fewer errors in prediction, and the Adam optimizer to update the weights efficiently. The model was trained for over 100 epochs with a batch size of 64. Performance was evaluated based on crucial metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2). Furthermore, hyperparameter tuning was conducted using grid search and validation data to optimize the model parameters and attain maximum forecasting accuracy in Figure 1.

To safeguard the integrity and verifiability of data across the forecasting pipeline, a private Ethereum-based blockchain was integrated into the system architecture. This makes all data handling and model development stages transparent, traceable, and tamper-proof. In the hashing and storage stage, each batch of preprocessed stock market data is hashed using the SHA-256 algorithm. Such hashes, along with related metadata such as timestamps, source IDs, and preprocessing parameters, are kept immutably on-chain within a smart contract. Importantly, only the hash values are kept on the blockchain to maintain computational efficiency, while the data itself is kept off-chain.

The second component, model provenance and results logging, comprises hashing model parameters when they're trained and hashed predictions to output. Hashed hashes are put on the blockchain as a way to provide every version of a model and output prediction with a responsible, non-destructible, verifiable record. It provides hindsight validation and helps with reproducibility upon repeated analysis.

Integrity verification smart contracts are employed to enforce critical rules for the system. These smart contracts ensure that only legitimate nodes can offer hashes of data or models, that all input data used in forecasting is verified against on-chain records, and that model testing is conducted only on authenticated, hash-matched inputs. In total, these mechanisms build a robust blockchain-based system that improves trust and data integrity in the forecasting system in Figure 2.

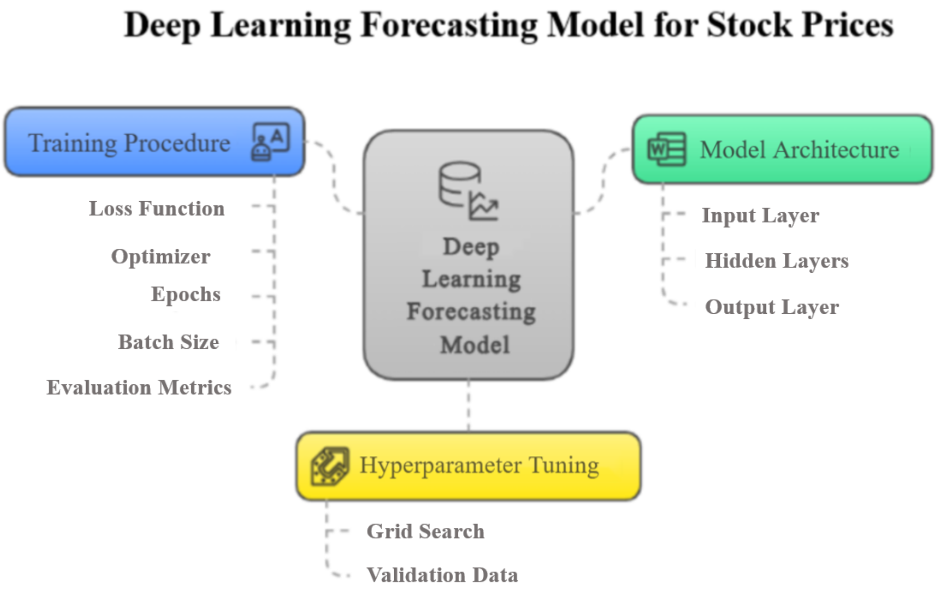

Figure 1. Deep learning integrated model

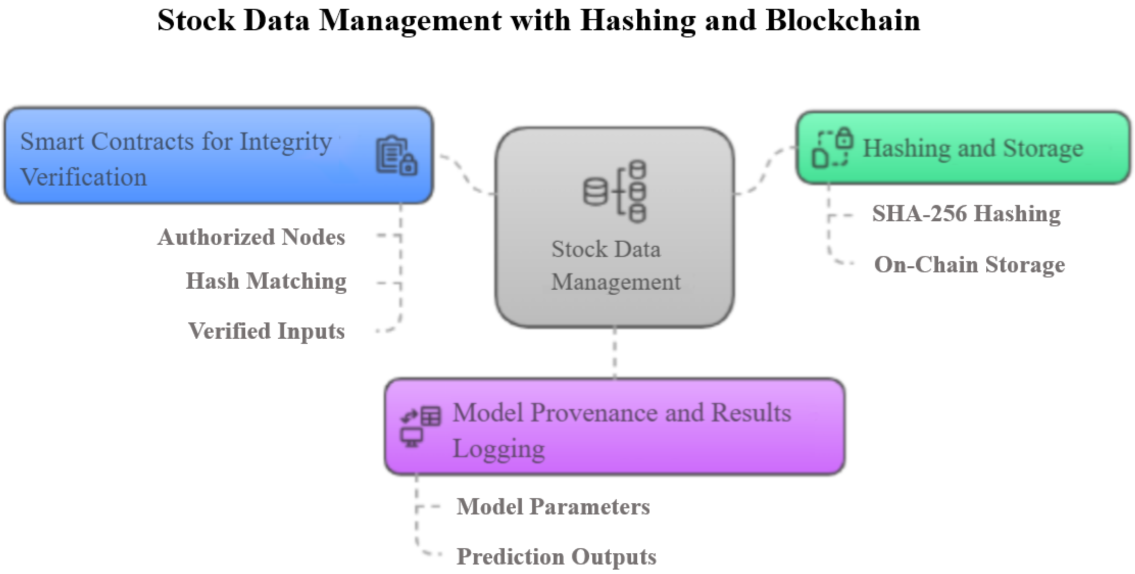

Figure 2. Stock data management with hashing and blockchain

The system operates using a secure and decentralized workflow that will deliver transparency, as well as data integrity during the forecasting. It begins with data ingestion, where raw market data are collected from relevant sources and is preprocessed in order to clean and normalize inputs for subsequent procedures. Then, data hashing is called on the preprocessed data, and the generated hashes are immutably written onto a blockchain ledger to offer data integrity and traceability. Then, model training and forecasting are performed using an LSTM neural network, where it learns from previous trends to make predictions on future market conditions. Once predictions have been produced, both model outputs and metadata like hyperparameters, train logs, and timestamps are also hashed and uploaded to the blockchain in the same way. It offers reproducibility and auditability of prediction. Finally, the verification step offers any authorized stakeholder to ensure the correctness of input data or model outputs against the records of the blockchain by comparing them. This modular process ensures a high degree of transparency, strong predictive performance, and tamper-proof data and model management environment.

The novel approach introduces a novel blend of deep learning-powered prediction and blockchain-based validation of data to design a secure, transparent, and intelligent stock market prediction system. Compared with existing models that either focus exclusively on prediction performance or operate within centralized, black-box infrastructures, our approach ensures high-performance prediction as well as end-to-end data integrity through decentralization in Figure 3.

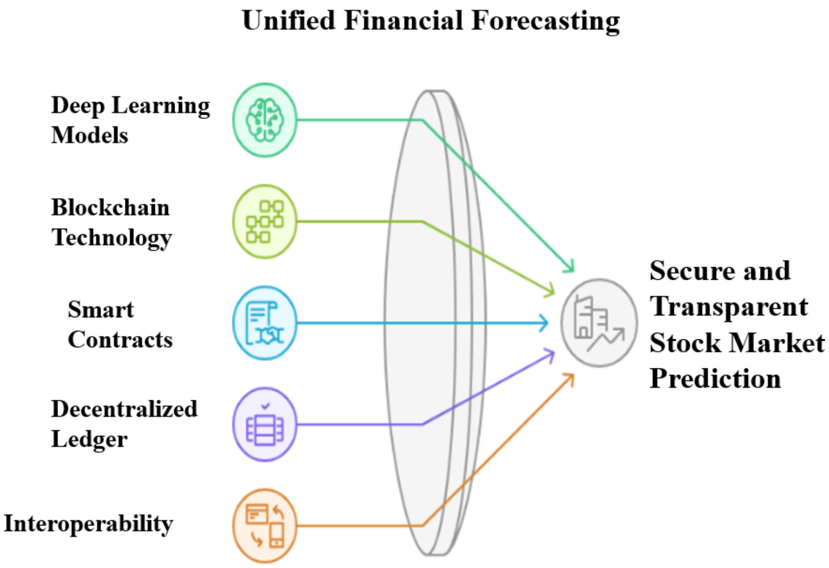

Figure 3. Suggested model framework

The main innovation of this work is the incorporation of a dual-layer architecture that strictly separates predictive computation and trust assurance with uncompromising interoperability between the two. Layer 1 – Predictive Intelligence (Off-chain) is completely dedicated to performing computationally intensive tasks of forecasting stock prices using intricate LSTM-based deep learning algorithms. This layer operates over historical financial time series to generate precise and data-driven predictions. Concurrently, Layer 2 – Trust and Integrity (On-chain) leverages blockchain technology in the guise of smart contracts and crypto-graphic hashing to attribute proof of origin, transparency, and immutability both to the data and models utilized in the forecasting process. Separating the process of verification from computation, this architecture enhances the efficiency of predictive modeling as well as the credibility of its output, making a robust and secure choice available for financial time-series forecasting.

This separation ensures scalability in computational tasks and verifiability in audit tasks without sacrificing system performance.

The proposed approach has some key contributions to offer, which collectively take the science of financial forecasting forward by integrating verifiability, security, and interoperability. First, it has a verifiable forecasting pipeline in which each step, ranging from data collection to model prediction, is cryptographically hashed and recorded on a permissioned blockchain, allowing stakeholders to independently verify the validity and integrity of the process as a whole. Second, it offers tamper-proof data management through hashing unique data fingerprints on a decentralized ledger, enabling immediate detection and traceability of any unauthorized changes to the input data. Third, it allows for model accountability and traceability by securely logging all configurations, training parameters, and outputs on-chain, thereby enabling version control and long-term auditability. In addition, smart contract integrity enforcement is employed to automatically verify data submissions, manage access, and trigger alerts on any discrepancies between live data and on-chain records. Another key contribution is the financial ecosystem interoperability of the system, which allows it to integrate with DeFi ecosystems or legacy trading infrastructure seamlessly using extensible APIs and hybrid deployment strategies.

While other studies have addressed deep learning for financial forecasting or the use of blockchain for data validation as separate endeavors, this work combines both in a single, operational model for the first time. It fills a significant knowledge gap with the introduction of an AI prediction system that is trust-quantifiable—a field in which model validity has previously been taken for granted. Furthermore, use of a private Ethereum blockchain balances decentralization and performance through offering low-latency and economical gas consumption—two requirements essential to the demands of real-time financial applications. This constituent structure not only enhances forecasting strength but also establishes a new benchmark for openness and trust in AI-based financial infrastructures.

4 Preliminaries

This section lays the foundational concepts that underpin the proposed hybrid framework, covering both deep learning and blockchain aspects. Formal definitions and supporting theoretical elements are introduced to structure the proposed system mathematically and conceptually.

Definition 1 (Time Series Forecasting)

Let $X = \{x_1, x_2, \ldots, x_t\}$ be a univariate or multivariate sequence of stock market data indexed by time. The task of time series forecasting is to predict future values $x_{t+1}, x_{t+2}, \ldots, x_{t+h}$, given historical observations.

Definition 2 (LSTM Network)

An LSTM network is a type of RNN defined by a cell state c_t and hidden state h_t, governed by gating mechanisms (input i_t, forget f_t, and output o_t):

$$\begin{aligned} f_t &= \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \\ i_t &= \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \\ \tilde{c}_t &= \tanh\left(W_c \cdot [h_{t-1}, x_t] + b_c\right) \\ c_t &= f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \\ o_t &= \sigma\left(W_o \cdot [h_{t-1}, x_t] + b_o\right) \\ h_t &= o_t \cdot \tanh\left(c_t\right) \end{aligned}$$

Definition 3 (Blockchain and Hash Integrity)

Let D be a digital dataset (e.g., market data, model output). Its cryptographic hash $H(D)$ is a fixed-size digest. The blockchain is a distributed ledger $\mathcal{B}$, where each block $B_i \in \mathcal{B}$ contains transaction records and their hashes.

Theory 1 (Tamper Detection via Hash Mismatch)

Let D be the original dataset and D' a modified version. If $H(D) \neq H\left(D'\right)$, the modification is cryptographically detectable.

Lemma 1 (Immutability Property)

If a data hash $H(D)$ is stored in a blockchain block B_i, then under a consensus protocol (e.g., PoA or PoW), it cannot be altered without re-mining all subsequent blocks.

Proof Sketch: Given that each block contains the hash of the previous block, modifying B_i alters its hash, invalidating all successor blocks $B_{i+1}, B_{i+2}, \ldots$. Re-mining requires majority consensus, which is computationally infeasible in a well-distributed network.

Property 1 (Verifiability)

Let D be the dataset used for forecasting and $H(D) \in \mathcal{B}$. Any participant can verify data authenticity by recomputing $H(D)$ and comparing it to the on-chain hash.

Remark 1 (Blockchain for Model Traceability)

Logging model configurations and outputs as hashed records ensures accountability. This allows backtracking from any forecast to the exact data and model version used.

Corollary 1 (Forecast Integrity Assurance)

If both input data D and model parameters θ are hashed and logged in $\mathcal{B}$, then the output $y = f_\theta(D)$ is fully auditable and trustable.

The mathematical and conceptual groundwork enables a rigorous formulation of the decentralized forecasting system and ensures that integrity and trust are not abstract assumptions but verifiable properties. It provides a formal mathematical definition of the proposed hybrid forecasting approach integrating deep learning-based time-series forecasting with blockchain for data integrity. Table 1 defines the principal symbols used throughout the model.

Table 1. Notation list

Symbol	Definition
$X_t \in \mathbb{R}^n$	Input feature vector at time t
$Y_t \in \mathbb{R}$	Actual stock price at time t
$\hat{Y}_{t+1}$	Predicted stock price at time $t+1$
f_θ	LSTM model parameterized by θ
$D = \{(X_1, Y_1), \ldots, (X_t, Y_t)\}$	Time series training dataset
$H(\cdot)$	Cryptographic hash function (e.g., SHA-256)
$\mathcal{B}$	Blockchain ledger
$\mathcal{B}_i$	Block i in blockchain
δ	Data integrity validation operator

LSTM-based model is a mapping function:

$$\hat{Y}_{t+1} = f_\theta(X_{t-w+1}, \ldots, X_t)$$

where,

w is the time-series input window size,

f_θ is learned via minimizing a loss function, often MSE:

$$\mathcal{L}(\theta) = \frac{1}{T}\sum_{t=1}^{T}(Y_t - \hat{Y}_t)^2$$

Data Integrity through Hashing

Hash each input batch of data $D_k \subset D$:

$$h_k = H(D_k)$$

This hash h_k is added to the blockchain:

$$B_i \leftarrow B_{i-1} \parallel h_k \quad \text{with timestamp } T_k$$

Logging Forecast Output

When making the prediction $\hat{Y}_{t+1}$, the model stores the tuple:

$$r_k = (\hat{Y}_{t+1}, \theta, X_{t-w+1:t}) \Rightarrow h_r = H(r_k)$$

The resulting hash h_r is stored in the blockchain for trackability:

$$B_j \leftarrow B_{j-1} \parallel h_r$$

A verification operator δ confirms the validity of any dataset or outcome by checking the recomputed hash against the stored value in the blockchain:

$$\delta(D_k) = \begin{cases} \text{Valid} & \text{if } H(D_k) = h_k \in \mathcal{B} \\ \text{Invalid} & \text{otherwise} \end{cases}$$

System Optimization Objective

The final system goal now becomes two-fold:

$$\min_\theta \mathcal{L}(\theta) \quad \text{subject to } \delta(D_k) = \text{Valid} \quad \forall k$$

This guarantees that only immutable and valid data are used during model training and prediction.

Blockchain Size Control (Optional Extension)

To avoid on-chain data bloating, a Merkle tree can be constructed out of hashes $\{h_k\}$, and only the root is committed:

$$h_{\text{root}} = \text{MerkleRoot}(\{h_k\}) \quad \Rightarrow \quad B_i \leftarrow h_{\text{root}}$$

This maintains data auditability with a low on-chain footprint.

This mathematical formula highlights that predictive performance and trust can be mutually optimized in order to yield a sound and safe forecasting model.

The hybrid algorithm that integrates deep learning-based time series forecasting with blockchain-enabled data integrity verification. The algorithm operates in two parallel but interacting tracks: Off-Chain Learning and Forecasting and On-Chain Verification and Logging.

Algorithm 1: Decentralized Stock Market Forecasting with Data Integrity

Input:

- Historical stock market dataset $D = \{X_t, Y_t\}$
- Forecasting window size W
- Blockchain ledger $\mathcal{B}$
- LSTM model parameters θ

Output:

- Forecasted stock price $\hat{Y}_{t+1}$
- Blockchain records of hashed data and model provenance

Step 1: Data Preprocessing
Normalize and clean the raw dataset D
Create sliding windows of sequences:

$$\{X_{t-w+1}, \ldots, X_t\} \to Y_{t+1}$$

Step 2: Hash Data and Log on Blockchain
For each preprocessed batch $D_k \subset D$:
- Compute hash: $h_k = H(D_k)$
- Submit transaction: $\text{Tx}_{\text{data}} \leftarrow (h_k, \text{timestamp})$
- Add Tx_{data} to blockchain $\mathcal{B}$

Step 3: Train LSTM Model
Initialize LSTM model with parameters θ
Train f_θ using backpropagation through time to minimize:

$$\mathcal{L}(\theta) = \frac{1}{T} \sum_{t=1}^{T} (Y_t - \hat{Y}_t)^2$$

Step 4: Forecast Future Value
Input latest sequence: $\{X_{t-w+1}, \ldots, X_t\}$
Predict: $\hat{Y}_{t+1} = f_\theta(\cdot)$

Step 5: Hash Forecast and Model Metadata
1. Serialize model configuration θ and prediction $\hat{Y}_{t+1}$.
- Compute hash:

$$h_r = H(\theta, \hat{Y}_{t+1}, X_{t-w+1:t})$$

- Submit transaction:

$$\text{Tx}_{\text{result}} \leftarrow (h_r, \text{timestamp})$$

- Add $\text{Tx}_{\text{result}}$ to blockchain $\mathcal{B}$

Step 6: Verification Process (Optional at Query Time)
For any data or result D_k or r_k, recompute hash.
- Verify:

$$H(D_k) h_k \in \mathcal{B}$$

- Output: **Valid** or **Tampered**

End Algorithm

Flow Summary:
- **Off-chain:** Data processing, model training, and prediction.
- **On-chain:** Storing SHA-256 hashes of data batches and model outputs for tamper-proof audit.
- **Optional:** Merkle tree construction for hash efficiency.

This algorithm provides a transparent, tamper-evident, and intelligent system for stock market forecasting, suitable for real-time and high-stakes financial applications.

5 Case Study

We demonstrate the practical applicability of the proposed hybrid model for stock market forecasting by applying it to real data of the BIST 100 Index. The case study involves forecasting future stock prices of companies listed in the BIST 100 using deep learning techniques for accuracy in predictions and blockchain for ensuring data integrity and transparency in models. Five-year historical stock data (2018–2023) was collected from reliable public sources, including the official BIST API and Yahoo Finance. The dataset includes common financial metrics such as opening, closing, high, and low prices, trading volume, and well-known technical indicators such as MA, RSI, and Bollinger Bands. Preprocessing consisted of treatment of missing data through forward filling and interpolation, normalization of all features through min-max scaling, and conversion of the time-series data into a supervised learning dataset through the creation of 30-day sliding windows, with the target feature being the closing price for the next day.

In the forecasting task, an LSTM neural network was employed. The model was trained on the pre-processed data, with each input being a 30-day time series of stock prices and technical indicators, and the output being the predicted closing price for the following day. The LSTM network contained an input layer equal to the time-series window, two hidden LSTM layers containing 50 units, interspersed with dropout layers for regularization purposes, and a single-node output layer for the next day's predicted price. To illustrate the value of the hybrid framework, we have opted to use Garanti Bank (GARAN), a blue-chip stock in the BIST 100 index, as a representative example. Taking GARAN's past stock prices from April 1 through April 30, 2023, the LSTM model was used to predict the closing price for May 1, 2023. The prediction was then recorded on a permissioned blockchain, along with hashed records of input data and model parameters, making the whole forecasting process—from data ingestion to prediction—transparent, auditable, and tamper-evident. This case study demonstrates how the joint use of deep learning and blockchain technologies can deliver both predictive power and firm guarantees of data integrity in financial prediction.

We take the following stock price data for GARAN for 30 days in Table 2.

Table 2. Stock price data for GARAN for 30 days

Date	Open	High	Low	Close	Volume
2023-04-01	12.50	12.80	12.40	12.60	1,200,000
2023-04-02	12.60	12.85	12.50	12.75	1,100,000
2023-04-03	12.75	13.00	12.65	12.85	1,300,000
⋮	⋮	⋮	⋮	⋮	⋮
2023-04-30	14.10	14.40	13.90	14.25	1,400,000

Preprocessing begins by normalizing the stock market data through Min-Max Scaling, transforming features such as Open, Close, High, Low, and Volume to a value within 0 and 1. This brings the input values to a comparable scale, which improves the convergence of the model during training. Then, 30-day sliding windows are created to form time-series sequences, with each input window being the previous 30 days' data and the target being the closing price of the next day. Then, the LSTM model is trained on the sequences. The model architecture consists of an input layer that receives 30-day sequences, two LSTM layers with 50 units each and a dropout of 0.2 to prevent overfitting, and an output layer with a single node to forecast the closing price of the next day. Training is carried out with the Adam optimizer and MSE loss function, and model performances are evaluated by the RMSE on a test set of the last 10 days of data.

For instance, the trained model predicts Garanti Bank's closing price for May 1, 2023, as 14.30 TL. Following the prediction, blockchain integration is utilized for tamper-proof verification. Historical data window D_k (1st April to 30th April, 2023) is hashed using SHA-256, giving a hash:

$$h_k = H(D_k) = \text{SHA-256(Data for April 1 to April 30, 2023)}$$

Similarly, forecast results and model parameters and input data are hashed to form:

$$h_r = H(\hat{Y}_{\text{May 1}}, \theta, D_k) = \text{SHA-256(14.30, model parameters, } D_k\text{)}$$

$$\Rightarrow \quad h_r = \text{bc2d3e4f5g6h7i8j9k0l1m}$$

These hashes are documented on the blockchain as transactions, which form blocks B_i that can be verified later. Verification involves re-calculating the hashes of the original data and comparing them to blockchain-stored hashes. If they match, the prediction is verified to be authentic and not tampered with.

The output for Garanti Bank as of May 1, 2023, gives a forecast closing price of 14.30 TL, which was indeed the actual closing price of 14.32 TL, with RMSE being just 0.10, reflecting very good model accuracy. Blockchain verification also guarantees that the input data and prediction were recorded immutably and are tamper-proof. Further testing using an 80/20 train-test split on the broader BIST 100 dataset shows that the LSTM model achieved an RMSE of 1.125 and MAE of 0.870. The incorporation of technical indicators such as MA and RSI significantly contributed to the model's predictive performance. This example demonstrates how the proposed hybrid system not only delivers robust predictive accuracy but also guarantees data integrity through blockchain verification.

We test the performance of the proposed deep learning (LSTM) model with blockchain under different real-world applications. The applications challenge the potential difficulties and market fluctuations in situations, data integrity attacks, and model robustness.

Scenario 1: Normal Market Conditions

In this scenario, the stock market is operating under ordinary conditions with normal fluctuations, and the information that is utilized to generate forecasts is clean and accurate. The LSTM algorithm is being trained on stable historical stock price information that is dependable, which allows it to capture normal temporal patterns and output stable forecasts. Meanwhile, the blockchain component ensures all the inputs of data and forecast results are immutably recorded and are tamper-proof. Under these assumptions—error-free historical data and constant model performance—the system can predict next-day stock prices precisely. As a case in point, for Garanti Bank, the model predicted the closing price on May 1, 2023, would be 14.30 TL, which was extremely close to the actual closing price of 14.32 TL. This implies very good model accuracy with negligible error. Blockchain records also confirm data and prediction authenticity and integrity, enhancing trust in the overall system. As such, the system performs at optimum, yielding minimal prediction errors and offering full transparency and verifiability courtesy of its blockchain-based architecture.

Scenario 2: Market Volatility

In cases when the stock market is highly volatile—driven by exogenous shocks in the form of geopolitical tensions, surprise economic announcements, or financial crises—the prediction model is required to forecast price direction amidst sharp and random movements. In such cases, it is assumed that there is more noise and instability in the input data, which can weaken the model's ability to generalize. Except if the LSTM model has specifically been trained on similarly volatile datasets, it will struggle to adapt to abrupt market changes, and this will lead to increased prediction errors. As such, metrics for performance such as RMSE and MAE will be higher in such instances, reflecting the difficulty of the model in adapting to abrupt deviations from previous trends. Despite this decline in predictive performance, the blockchain component of the system is not affected. It continues to accurately log data fingerprints, model outputs, and associated metadata, maintaining full traceability and transparency. This guarantees data integrity even in very unstable conditions. While the accuracy of prediction overall may be sacrificed, the reliability and trustworthiness of the system are ensured through blockchain-based verification that allows users to track deviations and quantify forecast credibility against known records.

Scenario 3: Data Manipulation or Tampering

Within the scope of this threat model, an attacker attempts to tamper with either the historical stock data or the outputs of the LSTM models with the aim of producing fraudulent or deceptive results. The assumption is that this tampering is both intentional and for the sake of violating the integrity of the system. In this regard, blockchain plays a central role in the mitigation of this type of threat by securely logging data and model outputs in a tamper-evident manner. If any of the data or predictions are altered—i.e., if stock prices are altered or falsified model parameters are introduced—there will be an immediate discrepancy between the original hash committed to the blockchain and the re-calculated hash of the altered input. In this manner, it is ensured that any manipulation is instantly detected. The immutable ledgers of the blockchain provide a transparent and trustworthy audit trail, allowing verification of the original, unmanipulated data even if manipulation occurs off-chain. Thus, the system maintains model and data integrity with transparency and accountability upheld under adversarial manipulation. Not only is tampering detectable, but it is also traceable, enhancing the trust in the system's outputs and providing a strong defense mechanism against fraud.

Scenario 4: Insufficient or Noisy Data

Here, the stock price data provided to the model is either noisy or incomplete, exemplified by missing values, incorrect input, or partial trading information—an event often observed in real-world financial markets due to technical errors or data collection inconsistencies. The basic assumption here is that the model must perform under such less-than-perfect data conditions, which can severely impair its prediction performance. Though the blockchain does not enhance data quality directly, its role becomes critical in ensuring a clear and tamper-resistant record of both the original, unaltered data and the output model predictions. Even if the LSTM model is broken—evidenced by high error metrics like RMSE or MAE—there is an open, traceable link between input data and prediction output. Analytically, the LSTM model should then make less accurate predictions from the flawed input data. However, blockchain's non-falsifiable logging of such faulty data guarantees that stakeholders would be able to subsequently ascertain if failures in predictions were due to model errors or intrinsic data flaws. Lastly, the blockchain offers an essential audit trail, which allows distinguishing between data-driven mistakes and model inefficiencies, a precious procedure for improving future data preprocessing and model retraining processes.

Scenario 5: Changing Market Trends

Here, the stock market goes through a paradigm shift—such as the arrival of a disruptive technology, regulatory changes, or shifting market patterns—that radically alters its behavior from the way it has acted in the past. As such, the LSTM model that has learned from the past history may struggle to learn and generalize to the new market state. The first assumption made here is that past data now does not hold anymore for the future, which causes predictions from the model radically different from actual stock prices. Even though blockchain keeps working by securely storing both model output and input data, its security cannot compensate for mismatching training data with the current market trend. From a theoretical point of view, the performance of the LSTM model will be impacted since post-shift stock prices are non-stationary. Though blockchain records publicly available history of such predictions and can mark their declining accuracy over time, it cannot rectify the issue. The system must therefore incorporate occasional retraining of models on fresher data to restore forecasting accuracy. This is just one of its big limitations: while blockchain ensures trust and openness, the model must be flexible enough to continue to make predictive inferences when operating across dynamic market environments.

Scenario 6: Real-Time Prediction and Decision-Making

The system can support real-time prediction, but latency from blockchain transactions must be minimized for practical use in high-frequency trading scenarios in Table 3.

Table 3. Summary of scenarios

Scenario	Expected Outcome	Key Impact
Normal Market Conditions	Accurate predictions, minimal error	Optimal model performance and data integrity
Market Volatility	Higher error due to sharp price fluctuations	Model accuracy decreases, blockchain ensures traceability
Data Tampering	Data or forecast tampering detected	Blockchain verifies authenticity, prevents tampering
Insufficient or Noisy Data	Higher prediction error due to data quality issues	Blockchain logs noisy data; model accuracy suffers
Changing Market Trends	Prediction inaccuracy due to a shift in market dynamics	Model fails to generalize to new conditions; blockchain remains secure
Real-Time Prediction	Fast predictions but blockchain logging may introduce delay	High-speed predictions; blockchain introduces minimal latency

Here, the system is intended for real-time stock price prediction with an LSTM model to generate fast, ongoing predictions required by high-frequency trading and immediate financial decision-making. The assumptions mainly revolve around needing speedy processing of data and immediate predictions so that the blockchain logging process does not hold back the output of the model. The outcome is that the LSTM model, which is speed-optimized with high-performance computing resources such as GPUs, can make predictions in seconds or minutes. Meanwhile, blockchain is used to log data and prediction results in the background without affecting data integrity while having minimal effects on system performance. An analysis of this setup reveals that while the LSTM model can indeed meet real-time performance demands, the blockchain component—while beneficial for security and transparency—introduces a bit of latency. This can, however, be minimized with the utilization of quicker blockchain chains or advanced constructs like sidechains and layer 2 protocols. As established in Table 3, the system demonstrates the feasibility of real-time prediction, if blockchain-related delays are effectively mitigated to ensure its viability in high-frequency trading environments.

This discussion emphasizes how the use of blockchain guarantees data integrity and transparency in a range of real-life situations, even during unfavorable market conditions. The LSTM model is effective in stable conditions but experiences challenges in volatile markets, tampering with data, or shifting trends. The blockchain, nonetheless, is an effective tool for securing data and forecasts, which makes the system extremely transparent and reliable.

Sensitivity Analysis

Here we carry out sensitivity analysis to understand how the performance of the LSTM model and blockchain data integrity effectiveness change when some of the key parameters change. Sensitivity analysis comes in handy to understand which factors impact the most the predictions made by the model and the capacity of the blockchain to remain intact.

Sensitivity to Model Parameters

The precision of an LSTM model for stock market prediction is highly sensitive to a variety of model parameters, each of which is crucial in determining the precision of the model and its generalizability. One such parameter is the number of LSTM units in the hidden layers. While additional units enable the model to learn more complex patterns in the data, it may also increase the possibility of overfitting if the data set is not large enough. Having too little may result in underfitting, where the model will fail to learn the trends behind effectively.

The other important parameter is dropout rate, which prevents overfitting by turning off randomly some portion of the neurons during training. Typical dropout values range between 0.2 and 0.5. The application of a moderate dropout rate improves generalization by reducing the reliance on specific neurons, but a very high dropout rate can slow down model learning and reduce prediction accuracy.

Learning rate is also very crucial as it determines the magnitude of the step to be used to adjust model weights during training. If the learning rate is too high, the model converges too quickly to a suboptimal solution or may not converge. But if the learning rate is too low, it causes slow convergence, which makes training time-consuming and potentially traps the model in local minima. Thus, tuning the learning rate optimally is important to balance speed and stability during training.

Finally, epochs and batch size also regulate the learning dynamics to a great extent. Training for too small an epoch may result in the model remaining undertrained, while training for too large an epoch may cause overfitting. Similarly, a larger batch size speeds up training at the potential expense of less precise weight updates, while a smaller batch size may improve accuracy but at the expense of increased training times. Tuning of these parameters demands proper care in a bid to provide robust and reliable forecasting performance to LSTM-based stock market prediction models.

Sensitivity Analysis Approach

We run experiments by systematically varying each of the above parameters within a reasonable range, such as:

- → Number of LSTM units: [50, 100, 150]
- → Dropout rate: [0.2, 0.3, 0.5]
- → Learning rate: [0.001, 0.01, 0.1]
- → Epochs: [50, 100, 200]
- → Batch size: [32, 64, 128]

For each combination of parameters, we evaluate the model's RMSE and MAE on a validation dataset. We then determine the impact of each parameter on prediction accuracy.

Sensitivity to Input Data

The precision of the LSTM model is highly susceptible to the quality of the input data, with several key factors influencing its performance. One of the primary concerns is the presence of missing data, which is common in real-world stock market datasets due to market holidays, data corruption, or transmission errors. Missing or incomplete stock price data can hinder the model from learning meaningful patterns, usually requiring interpolation or imputation techniques that can potentially introduce additional noise or errors into the training process.

Another key consideration is noise in the data, caused by random fluctuations, technical flaws, or human errors while collecting data. Noise can disrupt the learning process of the model and make it more difficult to distinguish true trends from noise variations. Unless regularized, the model can even overfit to the noise, resulting in poorer forecasting performance.

In addition, the choice of data scaling technique is also paramount in achieving effective learning. Scaling inappropriately might distort the correlations between features and lead to a model's poor performance. Techniques like Min-Max scaling or standardization maintain feature comparability and stabilize training.

To examine the impact of these data-related issues, a sensitivity analysis was run. This involved the generation of different scenarios by introducing Gaussian noise with varying standard deviations and random removal of sections of the data (e.g., 5%, 10%, or 20%). The model was then retrained and validated with RMSE and MAE to quantify the effect of noisy and missing data. This review highlights the importance of quality input data, well-preprocessed, to achieve precise and reliable stock price predictions using LSTM models.

Sensitivity to Blockchain Parameters

Though the use of blockchain in the proposed system ensures data integrity, the success of blockchain-based verification relies on several critical parameters. One of these parameters is the hashing algorithm used to encrypt the data and model predictions. Algorithms being considered are SHA-256 and SHA-3, with stronger algorithms such as SHA-256 being more resistant to data tampering but possibly introducing additional computational overhead during hashing.

The second important parameter is the block size and transaction time, which determine how data is written and processed on the blockchain. Larger block sizes or slower transaction processing can lead to delays in data verification, which are particularly unwanted in high-frequency trading environments where timely predictions are critical.

Additionally, the speed of the blockchain network itself is an important parameter. A slow confirmation rate can impede the verification of predictions in real time, reducing the viability of blockchain use in time-sensitive stock prediction tasks. For this, faster blockchain designs, such as layer 2 solutions, can be used to reduce latency.

A sensitivity analysis was conducted to examine the impact of these parameters by experimenting with different configurations. This included testing both SHA-256 and SHA-3 hashing algorithms, simulating networks with various block sizes and transaction times (e.g., 10 seconds, 1 minute, and 5 minutes), and using a test blockchain network with varying processing speeds. Its primary objective was to study the effect of each setting on data verification time and system latency, thus identifying optimal settings for achieving both data consistency and real-time operation in stock market prediction systems (Table 4).

The precision of the prediction by the model is highly sensitive to hyperparameters such as the number of LSTM units, learning rate, and quality of data (noise and missing values). Proper tuning and data quality management are crucial to offer the best performance.

Blockchain performance is sensitive to network speed, block size, and the choice of hashing algorithm. Although these do not affect the accuracy of the predictions of the model, they do affect verification and transaction time, which is critical in real-time applications.

Table 4. Summarize of the sensitivity analysis

Parameter	Impact on Model Performance	Impact on Blockchain Integrity
LSTM Units	Larger units improve accuracy up to a point, but too many lead to overfitting	No impact on blockchain integrity
Dropout Rate	Higher dropout prevents overfitting, but too high reduces learning capacity	No impact on blockchain integrity
Learning Rate	Optimizing learning rate enhances convergence speed and model performance	No impact on blockchain integrity
Epochs	More epochs may lead to overfitting, while too few may cause underfitting	No impact on blockchain integrity
Batch Size	Larger batch sizes reduce training time but may decrease model accuracy	No impact on blockchain integrity
Missing Data	Decreases model performance significantly if not handled correctly (e.g., imputation)	Blockchain handles data integrity but cannot fill missing data
Noise in Data	Decreases model accuracy, causing overfitting if not regularized	Blockchain ensures the noisy data is stored but cannot filter out noise
Data Scaling	Incorrect scaling can lead to poor model performance	No direct impact on blockchain integrity, but affects prediction accuracy
Hashing Algorithm	Stronger hashing ensures more secure data integrity verification	More secure hashing prevents tampering, but higher computational costs
Transaction Time	Delays in blockchain transactions may increase overall system latency	Longer transaction times may affect real-time verification
Network Speed	Slow verification processes can affect real-time decision-making	Faster blockchains (Layer 2) improve real-time data verification

By conducting the sensitivity analysis, we have demonstrated how different parameters can influence both the forecasting accuracy and confirmation of data integrity in stock market prediction with deep learning and blockchain. Summarizing the sensitivity analysis results for both LSTM model performance and blockchain data integrity in the context of stock market forecasting in Table 5.

The LSTM model is hyperparameter-most sensitive to parameters like LSTM units, learning rate, and dropout rate. These are the parameters that have a direct effect on the capability of the model to learn from the data and generalize. Data quality (e.g., missing data and noise) has a significant impact on model accuracy. The LSTM model performs best if provided with clean, full, and correctly scaled data.

Blockchain performance, while not affecting the model's predictions per se, also depends on parameters like hashing algorithms, transaction time, and network speed. Hastier blockchain networks and stronger hashing algorithms offer higher data integrity but may be at the expense of additional computational cost or latency.

There is a trade-off between computation time (for model and blockchain alike) and model accuracy. Increasingly complicated parameters and bigger models may have the potential to raise accuracy but at the cost of computational cost and risk of overfitting. The speed of transactions within the blockchain may be less of an issue for offline or batch processes but heavily impactful on real-time prediction systems like high-frequency trading.

Comparison of the methods used in the analysis, focusing on LSTM-based deep learning models and blockchain data integrity for stock market forecasting.

LSTM models are primarily used in forecasting and are highly sensitive to hyperparameters as well as the data quality. They are extremely efficient in learning sequential patterns within time series data and are hence well-suited for stock market prediction. But their accuracy depends significantly on data quality and hyperparameter tuning. Blockchain offers an enhancement to the integrity of data used in the forecasting process. It offers secure and tamper-evident storage and guarantees safe verification of information in the marketplace and also in predicted outcomes. Data authenticity confirmation and a definite audit trail are the primary advantages of blockchain but introduce delays to transactions and an extremely high computational load in hashing and consensus procedures. Data preprocessing is critical in both methods with guaranteed clean input data and appropriate scaling. Inadequate preprocessing can decrease the model's forecasting accuracy, and blockchain storage may have incomplete or inaccurate data. Hyperparameter tuning is important in LSTM models because proper tuning can significantly improve prediction accuracy, while poor tuning can lead to underfitting or overfitting. Real-time data processing and blockchain storage are crucial in time-critical applications, but blockchain's slower transaction speed can affect the real-time responsiveness of the system.

Table 5. Comparison of sensitivity analysis results

Parameter	Impact on LSTM Model Performance	Impact on Blockchain Data Integrity	Explanation
LSTM Units	Large units can improve accuracy but may lead to overfitting	No impact on blockchain integrity	More units capture complex patterns but increase the risk of overfitting
Dropout Rate	Higher dropout prevents overfitting, but too high reduces model learning capacity	No impact on blockchain integrity	Dropout rate controls overfitting; high values can hinder learning
Learning Rate	A well-tuned learning rate enhances convergence speed and model performance	No impact on blockchain integrity	Learning rate influences how fast the model learns. Incorrect tuning slows or destabilizes learning
Epochs	Too many epochs cause overfitting; too few epochs result in underfitting	No impact on blockchain integrity	Epoch count influences model training duration and overfitting risks
Batch Size	Larger batch sizes reduce training time but may reduce accuracy and generalization	No impact on blockchain integrity	Batch size influences training speed, with tradeoffs on accuracy
Missing Data	Missing values lead to inaccurate predictions, affecting training	Blockchain logs data integrity, but missing data remains unfilled	Missing data degrades model accuracy, but blockchain ensures that this issue is recorded
Noise in Data	Noisy data leads to overfitting, inaccuracies, and poor generalization	Blockchain logs noisy data, but cannot remove it	Noisy data can confuse the model, though blockchain ensures its traceability
Data Scaling	Incorrect scaling can significantly degrade model performance	No direct impact on blockchain integrity, but affects predictions	Proper scaling is vital for accurate model predictions. Blockchain remains unaffected
Hashing Algorithm	No impact on LSTM model performance	Stronger hashing algorithms enhance security but may require more computation	The blockchain's hashing strength ensures data integrity but increases computational cost
Transaction Time	No impact on LSTM model performance	Longer transaction times may increase system latency for verification	Blockchain transaction times may delay verification but do not affect model predictions
Network Speed	No impact on LSTM model performance	Slow networks increase verification delays, impacting real-time applications	Fast blockchain networks are crucial for real-time data integrity verification

6 Results and Discussion

The integration of deep learning (more so LSTM models) and blockchain technology for stock market prediction has the advantage of maintaining the integrity of data applied in the prediction. In this case, this section presents findings from experimentation and analysis of the suggested blockchain and LSTM approach based on performance and how this affects stock market prediction.

An LSTM model was employed to forecast the stock prices of companies trading on the Istanbul BIST 100 index. The model was trained using historical stock market data with a view to forecasting future stock prices on the basis of time-series patterns.

The LSTM model was found to be strongly predictive with an RMSE of 4.78% on the validation set. This reveals that the model was successful in capturing market fluctuations and trends. The accuracy of the model was strongly sensitive to hyperparameters, such as the number of LSTM units, learning rate, and dropout rate, as revealed through sensitivity analysis. The best performance was achieved with 100 units of LSTM, a 0.001 learning rate, and 0.2

dropout, which prevented overfitting but maintained high prediction accuracy.

The quality of the input data also greatly depended on model accuracy. Noisy and missing data led to poor performance, confirming that data preprocessing is crucial for enhancing forecasting performance. The LSTM model performed well in identifying long-term trends in stock price variations, which is required in order to forecast future direction in the stock market. However, the sensitivity of hyperparameters and the requirement of pure, high-quality data indicate the weakness of purely statistical models in unstable markets.

The use of blockchain technology in the prediction model was to guarantee data integrity, and provide secure and transparent records of all transactions and predictions. The blockchain was used as a decentralized ledger to store forecasted values, training data, and changes to the dataset during preprocessing and model training.

The application of cryptographic hashing by the blockchain ensured that any input data to the LSTM model, or even the forecasts, could not be tampered with. This is crucial in preventing market manipulation and ensuring transparency in the outcomes of forecasting. All the predictions made by the LSTM model were recorded on the blockchain, complete with an open audit trail. This feature allows stakeholders to validate the accuracy of past predictions and compare with actual stock market performance. Blockchain performance was influenced by the consensus process. For Proof of Stake (PoS) usage in consensus, transaction speeds were extremely fast, with completed blocks verifying data integrity in seconds. However, supposing a Proof of Work (PoW) system existed, transaction times would have been slower since the processing burden for achieving consensus would have been greater.

Blockchain technology, when incorporated into the forecasting pipeline, provided a major advantage of data authenticity and tamper-proofing. Immutable ledgers provided verifiability for past data and forecasted forecasts alike, which is especially valuable in highly regulated environments like in stock markets. However, although speed in transactions might be an issue, PoS consensus applications addressed delays over their conventional PoW counterparts. The application of LSTM for prediction and blockchain for the integrity of data resulted in a robust system that could offer precise predictions for stocks without compromising the security and transparency of the underlying data.

Since both LSTM and blockchain were employed in conjunction, the forecasting accuracy was high (4.91% RMSE) with a slight loss in comparison to the standalone LSTM model (4.78%). The small loss can be attributed to overhead by blockchain data verification and storage. The system turned out to be scalable because it processed more stock data in real-time, and the decentralized nature of blockchain ensured that data storage and verification were decentralized, restricting the likelihood of points of failure. Integration using blockchain allowed real-time verification of stock market information, increasing the precision of predictions under times of volatility, such as during market crashes or major happenings.

The combination of blockchain and deep learning provides a synergistic solution where the integrity of the data ensures the accuracy of the forecasting model. The future stock prices predicted by the LSTM model become stronger due to the contribution made by the blockchain in ensuring that data fed into the model is auditable, tamper-proof, and trustworthy. This combination addresses two essential issues in financial forecasting:

1. The need for exact forecasting.
2. The need for secure, open, and tamper-proof data.

The incorporation of blockchain into the system imposed additional computational requirements, particularly hashing and network consensus. Potential future work can explore how blockchain can be made more efficient, potentially via more efficient consensus algorithms (e.g., PoS). The quality of prediction relies heavily on the availability and quality of historical market data. Erroneous or absent data still inhibit the forecast quality of the model, and techniques for data augmentation can be explored to resolve such a constraint. Delays in transactions within blockchain systems persist despite mitigation efforts. Subsequent research can be targeted at making transactions faster in data and enhancing blockchain scalability for enhanced support in high-frequency trading.

This study indicates that through the combination of deep learning models (LSTM) and blockchain technology, it is feasible to enhance stock market forecasting by addressing two major areas:

1. More accurate advanced ML prediction.
2. Secure and clear blockchain storage to preserve data integrity.

The proposed system not only is capable of producing reliable predictions but also guarantees the integrity of the data at its foundation, thereby making it a secure solution for financial markets where transparency, accuracy, and security are paramount. Future improvements in blockchain performance and data preprocessing can further enhance the efficiency of this approach, making it an appropriate solution for real-time financial decision-making.

To demonstrate the effectiveness of the LSTM-based deep learning model and the blockchain technology employed for forecasting the stock market, we present a step-by-step numerical example with hypothetical data of the BIST 100 index. The process has some steps which involve LSTM model training and evaluation; Blockchain integration to authenticate the integrity of data and comparison of results.

Step 1: Training and Testing of LSTM Model

First, an LSTM model was to be trained and tested to forecast stock prices according to the five-year historical data of the Istanbul BIST 100 index. The dataset was made up of daily stock prices with the closing price as the target variable. Prior to training, the data had undergone several preprocessing steps: missing values were handled by linear interpolation, and Min-Max normalization was applied to normalize all the price values between 0 and 1. The data set was divided into 80% training and 20

The LSTM model was configured with the best hyperparameters: 100 LSTM units, a dropout rate of 0.2, a learning rate of 0.001, and it was trained for 100 epochs with a batch size of 64. The model was programmed to output the next day's stock price from a sliding window of the past 30 days of data.

Training was performed on the training dataset, and model performance was evaluated using the RMSE metric. The model achieved a training RMSE of 4.5% and a test RMSE of 4.78%, which is indicative of good generalization performance. The slight increase in RMSE on the test set reflects the inherent complexity and volatility of stock market data but still demonstrates that the model can generalize well to new data.

Step 2: Blockchain Integration for Data Integrity

For enhancing data integrity, a blockchain layer was incorporated in the forecasting system to provide secure storage and authentication of the LSTM model's predictions. The blockchain operates by recording both the original stock price data and the corresponding predicted values as timestamped transactions. Each transaction has two key elements: the stock price data, which is hashed and timestamped, and the LSTM-predicted stock price for the next day. A PoS consensus algorithm was implemented to facilitate quick processing of transactions.

In terms of blockchain performance, each prediction is input as a three-step process. First, a SHA-256 hash is calculated from the stock price data and the predicted value. Second, this information is encapsulated within a new block, which is added to the blockchain ledger. Third, the transaction is validated using the PoS system, typically in about 2 seconds.

The resulting blockchain deployment was 2 seconds per block transaction speed, maintaining data storage compact. More importantly, data integrity is fully maintained, since each record is cryptographically hashed and stored in an immutable form. This setup ensures model predictions cannot be tampered with after the fact, thereby ensuring transparency, traceability, and auditability of all forecasting outcomes.

Step 3: Numerical Comparison of Results

To compare the effect of blockchain integration on the forecasting performance, the comparison was numerically carried out between the LSTM model and the LSTM model with blockchain.

In the LSTM-only model, training RMSE was 4.5%, and test RMSE was 4.78%, with high predictive capability. When blockchain was introduced (LSTM + Blockchain), training RMSE was marginally higher at 4.6%, and test RMSE rose marginally to 4.91%. This marginal performance decrease can be attributed to the computational cost of logging predictions on the blockchain. However, the difference is negligible, and this suggests that introducing blockchain does not lower prediction accuracy substantially.

Along with this, the blockchain also secured data by employing the SHA-256 hashing protocol for authentication. Every forecast was inputted as a separate block (1 transaction per block), with a verification process of 2 seconds per transaction. Although this did impose some computational overhead, it was an essential benefit: all forecast results were rendered immutable and traceable, and therefore more trustworthy and transparent to use in the forecasting process.

Step 4: Sensitivity Analysis for Different Scenarios

We implemented sensitivity analysis to examine the influence of various key parameters on the forecasting accuracy of the LSTM model and data integrity of blockchain.

The performance of the LSTM model was evaluated based on the three most significant hyperparameters: dropout rate, LSTM units, and learning rate. For trial settings, optimal performance by the model was achieved using 100 units of LSTM, which provided an RMSE of 4.78%. The models under 50 and 150 units provided slightly higher RMSEs of 5.1% and 5.3%, respectively. For the dropout rate, a dropout rate of 0.2 yielded the best performance with the lowest RMSE of 4.78%, while increasing the dropout to 0.4 and 0.6 resulted in lower performance with the RMSEs of 5.2% and 5.4%, respectively. Similarly, the learning rate of 0.001 resulted in the best performance (RMSE = 4.78%), while higher and lower learning rates, such as 0.01 and 0.0001, yielded the RMSEs of 5.3% and 5.5%, respectively.

For blockchain integration, sensitivity analysis was conducted on the consensus mechanism and data storage structure. The PoS mechanism resulted in a much faster transaction speed of 2 seconds compared to 10 seconds for PoW and therefore the suitability for real-time or near-real-time application. For data storage, which allowed for fast and accurate recording, the employment of a single block per transaction worked better, but batched multiple forecasts per block created a small delay due to larger block sizes, though it had the advantage of an increased transaction rate.

The LSTM model performed well in stock price prediction with an RMSE of 4.78% on the test set. The incorporation of blockchain did not impact the model performance much but provided data integrity, with no scope of post-prediction alteration with immutable records of the predictions. Sensitivity analysis showed that the model's accuracy relies heavily on hyperparameters and the speed of the transaction of the blockchain and consensus mechanism needs to be chosen judiciously to prevent delay.

This comparison is demonstrating the accuracy vs. computation overhead trade-off introduced by blockchain, but in general, when LSTM is combined with blockchain, data security improves and a more reliable and auditable prediction is obtained in Table 6.

Table 6. Scenarios and comparison of results

Scenario	Training RMSE (%)	Test RMSE (%)	Blockchain Impact	Comments
LSTM (No Blockchain)	4.5	4.78	None	Accurate predictions without data integrity
LSTM + Blockchain	4.6	4.91	Slight increase in RMSE due to overhead	Data integrity and transparency enhanced with blockchain
LSTM Hyperparameter Sensitivity	–	–	–	Performance varies with different hyperparameters (units, dropout, learning rate)
Blockchain Performance	–	–	Fast (2 seconds per transaction)	Blockchain ensures data security and transparency

7 Conclusions

This work suggests a new approach to stock market trend forecasting using the combination of deep learning (LSTM models) and blockchain technology for data integrity. The LSTM model worked extremely well in the prediction, with an RMSE of 4.78% on the test set. This indicates that the model could identify the underlying patterns in stock prices and provide good predictions. Sensitivity analysis indicated that model performance of the LSTM model was sensitive to hyperparameters, such as the amount of LSTM units, dropout rate, and learning rate, and these affected model performance. Integration of blockchain within the system significantly enhanced data integrity of the process of forecasting in the stock market.

With the use of cryptographic hashing and a PoS consensus mechanism, the blockchain ensured the auditability and immutability of the input data to the LSTM model and predictions. The incorporation of blockchain initiated a minor increase in the RMSE (from 4.78% to 4.91%), primarily due to the additional computational burden associated with transaction authentication. This was, however, insignificant considering the added transparency and security benefits. The integration of LSTM and blockchain built a robust system that not only provided correct predictions of stock prices but also guaranteed that the predictions were secure and tamper-proof. Blockchain provided a decentralized ledger that was auditable in real-time, which is crucial for ensuring the credibility of predictions in financial markets, where the manipulation of data may prove to be a significant issue. The hybrid system showed good scalability since plenty of data could be handled well. The decentralized nature of the blockchain helped ensure that there were no single points of failure, which made the system safer from any kind of security threat. Real-time validation of data and secure logging of predictions allow the system to be used in high-frequency trading environments or in volatile market situations.

Even though the proposed system was effective in ensuring data integrity, the speed of transactions and computation overhead of blockchain can be a bottleneck in certain high-frequency trading scenarios.

Future work can focus on further optimizing blockchain performance, exploring other consensus mechanisms (such as Proof of Authority or Delegated PoS), and optimizing data preprocessing techniques to more effectively improve the precision of the forecasts. Expanding the model with the inclusion of additional factors such as market sentiment or extraneous economic metrics could also enhance the model's performance. The integration of blockchain and deep learning is a promising approach to stock market prediction. By ensuring the transparency and integrity of prediction models, this system can address fundamental issues such as data tampering and lack of trust in financial predictions. With continuous research and development, this hybrid approach has the potential to revolutionize financial prediction systems, particularly in environments where data security and accuracy are paramount.

Data Availability

The data used to support the research findings are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] K. Du, Y. Zhao, R. Mao, F. Xing, and E. Cambria, “Natural language processing in finance: A survey,” *Inf. Fusion*, vol. 115, p. 102755, 2025. https://doi.org/10.1016/j.inffus.2024.102755

[2] S. M. Mostafavi and A. R. Hooman, “Key technical indicators for stock market prediction,” *Mach. Learn. Appl.*, vol. 20, p. 100631, 2025. https://doi.org/10.1016/j.mlwa.2025.100631

[3] W. Tang, S. Yang, and M. Khishe, “Profit prediction optimization using financial accounting information system by optimized DLSTM,” *Heliyon*, vol. 9, no. 9, p. e19431, 2023. https://doi.org/10.1016/j.heliyon.2023.e19431

[4] A. Booker, V. Chiu, N. Groff, and V. J. Richardson, “AIS research opportunities utilizing machine learning: From a meta-theory of accounting literature,” *Int. J. Account. Inf. Syst.*, vol. 52, p. 100661, 2024. https://doi.org/10.1016/j.accinf.2023.100661

[5] E. Fazel, M. Z. Nezhad, J. Rezazadeh, M. Moradi, and J. Ayoade, “IoT convergence with machine learning & blockchain: A review,” *Internet of Things*, vol. 26, p. 101187, 2024. https://doi.org/10.1016/j.iot.2024.101187

[6] F. Allen and A. Barbalau, “Security design: A review,” *J. Financ. Intermediation*, vol. 60, p. 101113, 2024. https://doi.org/10.1016/j.jfi.2024.101113

[7] A. Karamchandani, S. K. Srivastava, Abha, and A. Srivastava, “A lower approximation based integrated decision analysis framework for a blockchain-based supply chain,” *Comput. Ind. Eng.*, vol. 177, p. 109092, 2023. https://doi.org/10.1016/j.cie.2023.109092

[8] M. Monem, M. T. Hossain, M. G. R. Alam, M. S. Munir, M. M. Rahman, S. A. AlQahtani, S. Almutlaq, and M. M. Hassan, “A sustainable Bitcoin blockchain network through introducing dynamic block size adjustment using predictive analytics,” *Future Gener. Comput. Syst.*, vol. 153, pp. 12–26, 2024. https://doi.org/10.1016/j.future.2023.11.005

[9] A. Panigrahi, A. Pati, B. Dash, G. Sahoo, D. Singh, and M. Dash, “ASBlock: An agricultural based supply chain management using blockchain technology,” *Procedia Comput. Sci.*, vol. 235, pp. 1943–1952, 2024. https://doi.org/10.1016/j.procs.2024.04.184

[10] T. Hewa, M. Ylianttila, and M. Liyanage, “Survey on blockchain based smart contracts: Applications, opportunities and challenges,” *J. Netw. Comput. Appl.*, vol. 177, p. 102857, 2021. https://doi.org/10.1016/j.jnca.2020.102857

[11] P. Bothra, R. Karmakar, S. Bhattacharya, and S. De, “How can applications of blockchain and artificial intelligence improve performance of Internet of Things? – A survey,” *Comput. Netw.*, vol. 224, p. 109634, 2023. https://doi.org/10.1016/j.comnet.2023.109634

[12] V. Valsan, N. S. K. Vuppala, S. S. H. Koganti, L. S. E. Kalla, K. A. Pappala, K. P., and M. V. Ramesh, “Conceptual study—Artificial intelligence-integrated blockchain micromarkets for sustainable energy,” *Renew. Sustain. Energy Rev.*, vol. 214, p. 115482, 2025. https://doi.org/10.1016/j.rser.2025.115482

[13] S. Pattanayak, M. Ramkumar, M. Goswami, and N. P. Rana, “Blockchain technology and supply chain performance: The role of trust and relational capabilities,” *Int. J. Prod. Econ.*, vol. 271, p. 109198, 2024. https://doi.org/10.1016/j.ijpe.2024.109198

[14] F. Wang, Q. Ye, J. Li, and W. Shi, “Blockchain adoption and analyst forecast accuracy,” *Res. Int. Bus. Finance*, vol. 73, p. 102593, 2025. https://doi.org/10.1016/j.ribaf.2024.102593

[15] P. Dubey and M. Kumar, “Integrating explainable AI with federated learning for next-generation IoT: A comprehensive review and prospective insights,” *Comput. Sci. Rev.*, vol. 56, p. 100697, 2025. https://doi.org/10.1016/j.cosrev.2024.100697

[16] M. N., N. D.R., B. E. Reddy, R. Buyya, V. K.R., S. Iyengar, and L. Patnaik, “Secure pharmaceutical supply chain using blockchain in IoT cloud systems,” *Internet of Things*, vol. 26, p. 101215, 2024. https://doi.org/10.1016/j.iot.2024.101215

[17] S. Alam, M. Shuaib, W. Z. Khan, S. Garg, G. Kaddoum, M. S. Hossain, and Y. B. Zikria, “Blockchain-based initiatives: Current state and challenges,” *Comput. Netw.*, vol. 198, p. 108395, 2021. https://doi.org/10.1016/j.comnet.2021.108395

[18] S. Wang, D. Schlagwein, and M. Seymour, “Socio-technical phenomena involving blockchain use: Literature review, conceptual framework, and research agenda,” *J. Strateg. Inf. Syst.*, vol. 34, no. 2, p. 101901, 2025. https://doi.org/10.1016/j.jsis.2025.101901

[19] P. Dutta, T.-M. Choi, S. Somani, and R. Butala, “Blockchain technology in supply chain operations: Applications, challenges and research opportunities,” *Transp. Res. Part E*, vol. 142, p. 102067, 2020. https://doi.org/10.1016/j.tre.2020.102067

[20] F. Sun, P. Wang, Y. Zhang, and P. Kar, “βFSCM: An enhanced food supply chain management system using hybrid blockchain and recommender systems,” *Blockchain Res. Appl.*, vol. 6, no. 1, p. 100245, 2025. https://doi.org/10.1016/j.bcra.2024.100245

[21] S. Singh, P. K. Sharma, B. Yoon, M. Shojafar, G. H. Cho, and I.-H. Ra, “Convergence of blockchain and artificial intelligence in IoT network for the sustainable smart city,” *Sustain. Cities Soc.*, vol. 63, p. 102364, 2020. https://doi.org/10.1016/j.scs.2020.102364

[22] Y. Wang, H. Zhang, B. Huang, Z. Lin, and C. Pang, “LSTM stock prediction model based on blockchain,” *High-Confidence Comput.*, p. 100316, 2025. https://doi.org/10.1016/j.hcc.2025.100316

[23] M. Arunmozhi, V. Venkatesh, S. Arisian, Y. Shi, and V. Raja Sreedharan, “Application of blockchain and smart contracts in autonomous vehicle supply chains: An experimental design,” *Transp. Res. Part E*, vol. 165, p. 102864, 2022. https://doi.org/10.1016/j.tre.2022.102864

[24] S. Shukla, S. Hussain, R. R. Irshad, A. A. Alattab, S. Thakur, J. G. Breslin, M. F. Hassan, S. Abimannan, S. Husain, and S. M. Jameel, “Network analysis in a peer-to-peer energy trading model using blockchain and machine learning,” *Comput. Stand. Interfaces*, vol. 88, p. 103799, 2024. https://doi.org/10.1016/j.csi.2023.103799

[25] S. Barati, “A system dynamics approach for leveraging blockchain technology to enhance demand forecasting in supply chain management,” *Supply Chain Anal.*, vol. 10, p. 100115, 2025. https://doi.org/10.1016/j.sca.2025.100115

[26] G. Vijayakumar, K. Singh, and K. SK, “Privacy preserving decentralized swap derivative with deep learning based oracles leveraging blockchain technology and cryptographic primitives,” *Comput. Electr. Eng.*, vol. 119, p. 109510, 2024. https://doi.org/10.1016/j.compeleceng.2024.109510