ACSIE (International Journal of Application Computer Science and Informatic Engineering)
Vol.6, No.1, May 2024, pp. 14-25
ISSN (online): 2685-4600
DOI: 10.33173/acsie.v6i1.297

Ethereum Price Forecasting Using Bidirectional GRU Neural Networks Optimized with SGD

Putu Sugiartawan*[1], Ni Wayan Wardani[2]
[1]Department of Information and Communication Systems, Okayama University, Okayama 700-8530, Japan
[2]Graduate School of Environmental, Life, Natural Science and Technology, Okayama University, Okayama, Japan
e-mail: ***[1]p18z9yov@s.okayama-u.ac.jp** , [2]pj5w1e4c@s.okayama-u.ac.jp

Abstrak

Peramalan harga mata uang kripto merupakan permasalahan yang menantang karena tingkat volatilitas yang tinggi, dinamika nonlinier, dan ketergantungan temporal yang kuat. Ethereum sebagai salah satu mata uang kripto utama menunjukkan pergerakan harga yang kompleks akibat spekulasi pasar, peristiwa eksternal, dan perubahan perilaku investor yang cepat. Pendekatan statistik tradisional dan metode pembelajaran mesin konvensional sering kali tidak mampu memodelkan karakteristik tersebut secara optimal. Oleh karena itu, penelitian ini mengusulkan model peramalan harga Ethereum berbasis Bidirectional Gated Recurrent Unit (BiGRU) yang dioptimalkan menggunakan Stochastic Gradient Descent (SGD). Pendekatan yang diusulkan memanfaatkan pembelajaran rekuren dua arah untuk menangkap informasi konteks masa lalu dan masa depan selama proses pelatihan, sementara penggunaan SGD ditujukan untuk meningkatkan kemampuan generalisasi dan stabilitas pelatihan model. Metodologi penelitian disusun secara sistematis, meliputi tahap prapemrosesan data, pembentukan urutan berbasis sliding window, pelatihan model, dan evaluasi kinerja. Evaluasi dilakukan menggunakan metrik regresi standar, yaitu Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), serta analisis visual antara nilai aktual dan nilai prediksi. Hasil eksperimen menunjukkan bahwa model BiGRU–SGD mampu mengikuti tren harga Ethereum dengan baik dan menangkap ketergantungan temporal secara efektif pada data uji. Temuan ini menunjukkan bahwa kombinasi arsitektur rekuren dua arah dan strategi optimasi yang tepat merupakan solusi yang andal untuk peramalan harga kripto. Penelitian selanjutnya dapat mengintegrasikan indikator pasar tambahan, data sentimen, atau arsitektur lanjutan seperti mekanisme atensi dan model transformer untuk meningkatkan kinerja prediksi.

Kata kunci: *Peramalan harga Ethereum, Bidirectional GRU, stochastic gradient descent, pembelajaran mendalam, prediksi deret waktu.*

Abstract

Accurate forecasting of cryptocurrency prices is a challenging task due to their high volatility, nonlinear dynamics, and strong temporal dependencies. Ethereum, as one of the most prominent cryptocurrencies, exhibits complex price movements influenced by market speculation, external events, and rapidly changing investor behavior. Traditional statistical and shallow machine learning approaches often fail to capture these characteristics effectively, leading to suboptimal prediction performance. Motivated by these limitations, this study proposes an Ethereum price forecasting model based on a Bidirectional Gated Recurrent Unit (BiGRU) neural network optimized using Stochastic Gradient Descent (SGD). The proposed approach leverages bidirectional recurrent learning to capture both past and future contextual information during training, while the use of SGD aims to improve generalization performance

Received February 1st,2024; Revised March 25th, 2024; Accepted April 10th, 2024

and training stability. A structured methodology is applied, including data preprocessing, sliding-window sequence construction, model training, and systematic evaluation. The model is evaluated using standard regression metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and visual comparison between predicted and actual prices. Experimental results demonstrate that the proposed BiGRU–SGD model is able to closely track Ethereum price trends and capture temporal dependencies with satisfactory accuracy on unseen testing data. The findings indicate that combining bidirectional recurrent architectures with carefully configured optimization strategies provides a robust solution for cryptocurrency price forecasting. Future work may extend this framework by incorporating additional market indicators, sentiment data, or advanced architectures such as attention mechanisms and transformer-based models to further enhance predictive performance.

Keywords: *Ethereum price forecasting, Bidirectional GRU, stochastic gradient descent, deep learning, time series prediction.*

1. INTRODUCTION

This study addresses the increasing demand for robust, data-driven forecasting methods in the cryptocurrency domain, particularly for Ethereum — a leading smart-contract platform whose price dynamics have attracted intensive research interest in recent years. Financial time-series forecasting has undergone a paradigm shift with the adoption of deep learning architectures that can automatically capture nonlinear temporal dependencies and complex feature interactions; reviews and comparative studies from 2020–2024 demonstrate that deep recurrent models (including GRU and its bidirectional variants) and hybrid architectures frequently outperform traditional statistical baselines on volatile asset series such as cryptocurrencies and equities [1], [2]. Market events (e.g., protocol upgrades, regulatory announcements, macroeconomic shocks) and structural changes in liquidity generate nonstationary behaviour and high noise levels in crypto price series, which call for architectures capable of both capturing long-range temporal dependencies and being resilient to overfitting and training instability [3], [4]. These characteristics make the design choice of the underlying sequence model and its optimization strategy critical: recurrent units (GRU) have been consistently reported to offer a favorable trade-off between representational power and parameter efficiency for high-volatility assets, while bidirectional extensions provide a mechanism to integrate both past and future contextual cues during sequence encoding for one-step and multi-step forecasting tasks [4], [5]. Recent literature further highlights methodological best practices for financial deep learning—data preprocessing, feature engineering, regularization, and the optimizer selection—each of which materially affects out-of-sample generalization and the practical utility of forecasting systems [6].

Despite a growing body of work on cryptocurrency forecasting, important methodological gaps persist that constrain both scientific novelty and operational reliability. First, many studies reporting strong predictive results rely on standard one-directional recurrent models or hybrid CNN/RNN pipelines without systematically isolating the effect of bidirectionality in gated recurrent units for price prediction; as a result, claims about the value of BiGRU architectures remain under-examined in the specific context of Ethereum's market microstructure [2], [5]. Second, while sophisticated optimizers such as Adam and RMSprop are commonly used due to faster convergence in practice, stochastic gradient descent (SGD) with appropriate scheduling remains a classical optimizer that can yield better generalization in several domains when properly tuned; however, its role and hyperparameterization for crypto price series — especially when combined with bidirectional recurrent architectures — have not been rigorously compared across representative datasets and realistic evaluation protocols [6], [7]. Third, cross-study comparisons are frequently hampered by inconsistent evaluation procedures (data splits, look-ahead horizons, and performance metrics), which complicates reproducibility and reduces the value of reported improvements for downstream decision support in trading and risk management [6], [7]. These gaps motivate a focused investigation into whether a carefully configured Bidirectional GRU model trained with SGD can deliver meaningful improvements in forecasting accuracy and robustness for Ethereum, and whether such gains translate to better risk-aware evaluation metrics under out-of-sample tests.

The primary objective of this research is to design, implement, and evaluate a forecasting system that leverages a Bidirectional Gated Recurrent Unit (BiGRU) architecture optimized via Stochastic Gradient Descent (SGD) to predict Ethereum closing prices at short and medium horizons. Specifically, the work pursues three interrelated goals: (1) empirically quantify the predictive benefit of bidirectional recurrence relative to unidirectional GRU and other strong baselines (including LSTM and recent hybrid CNN–RNN variants) under standardized preprocessing and cross-validation protocols; (2) investigate the effects of SGD (with momentum and learning-rate scheduling) on model generalization and stability compared to adaptive optimizers; and (3) assess model robustness across multiple evaluation metrics (MAE, RMSE, MAPE, and directional accuracy) and realistic backtesting scenarios that emulate temporal deployment in live decision-support contexts. To accomplish these goals, the proposed solution integrates multivariate feature sets (price, volume, technical indicators and derived volatility measures), a multi-layer BiGRU encoder with dropout and layer normalization, and an SGD training pipeline with Cosine Annealing (or step scheduling) and early stopping. The experimental design follows contemporary recommendations for financial deep learning experiments: walk-forward cross-validation, strict prevention of look-ahead leakage, and the reporting of both point-forecast and directional performance to convey practical relevance [6], [2].

The contributions of this paper are fourfold. First, it provides a systematic, reproducible comparison that isolates the effect of bidirectionality in GRU networks for Ethereum forecasting and situates results against commonly used baselines and hybrid architectures reported in recent studies [2], [4]. Second, it offers the first (to our knowledge, within the examined literature window) in-depth analysis of SGD's role — including momentum and learning-rate schedules — in training BiGRU models for volatile cryptocurrency data, highlighting optimization regimes that improve out-of-sample performance and reduce overfitting risk relative to default adaptive optimizers [6], [5]. Third, the work contributes a production-oriented evaluation protocol that includes walk-forward backtesting and multiple error and utility metrics (MAE, RMSE, MAPE, directional accuracy, and a simple profit/loss simulation under conservative transaction cost assumptions), enabling clearer judgments about operational usefulness beyond raw accuracy measures [7]. Finally, we publicly release model architectures, hyperparameter grids, and the experimental code used for the evaluations to support reproducibility and to enable future comparative work in the IJCCS community. Empirical results reported in this paper show that a properly tuned BiGRU + SGD pipeline attains statistically significant improvements in several forecasting metrics versus unidirectional GRU and selected hybrid baselines across multiple time horizons, while also exhibiting favorable calibration in directional tasks — findings that are consistent with broader trends in financial deep learning literature that underscore the importance of model-optimizer co-design and rigorous evaluation [1], [2], [3], [6].

In closing, this introduction has established the motivation, goals, proposed solution, and measured contributions of a BiGRU-based forecasting system trained with SGD for Ethereum price prediction. By combining architectural choices that capture bidirectional temporal context with careful optimizer selection and robust evaluation procedures, the research aims to deliver both methodological insight and practical forecasting improvements that are relevant to scholars and practitioners working at the intersection of computing and cybernetics systems. The remainder of the paper details the data and feature engineering pipeline (Section II), model architecture and training regimen (Section III), comprehensive experimental results and ablation studies (Section IV), followed by a discussion of limitations and directions for future research (Section V).

2. METHODOLOGY

Research on cryptocurrency price forecasting has expanded rapidly since 2020, driven by the increasing market capitalization of digital assets and the inherent challenges posed by their highly volatile and nonlinear price dynamics. Early studies in this period primarily focused on adapting classical time-series and shallow machine learning techniques to cryptocurrency data, while more recent works have increasingly adopted deep learning models—particularly recurrent neural networks (RNNs) and their gated variants—to capture temporal dependencies and complex nonlinear relationships. This section reviews and synthesizes prior work relevant to Ethereum price forecasting, with a focus on model architectures, optimization strategies, datasets, evaluation protocols, and identified research gaps.

Before the widespread adoption of deep learning, several studies applied traditional statistical models and classical machine learning algorithms to cryptocurrency markets. Autoregressive Integrated Moving Average (ARIMA), Generalized Autoregressive Conditional Heteroskedasticity (GARCH), and their variants were frequently used as baseline models for Bitcoin and Ethereum price prediction. However, empirical results consistently demonstrated their limited ability to handle nonstationary behavior and abrupt regime shifts common in crypto markets, resulting in suboptimal predictive performance under volatile conditions [1].

To overcome these limitations, machine learning approaches such as Support Vector Regression (SVR), Random Forests, Gradient Boosting Machines, and k-Nearest Neighbors were introduced. Comparative studies published after 2020 reported that tree-based ensemble methods often outperform linear models when sufficient engineered features (technical indicators, volume metrics, volatility proxies) are available [2]. Nevertheless, these approaches rely heavily on manual feature engineering and typically lack an explicit mechanism to model long-term temporal dependencies. As a result, their performance degrades in multi-step forecasting scenarios and during periods of extreme market turbulence, motivating the transition toward deep sequence models.

Deep learning architectures, particularly RNN-based models, have become the dominant paradigm in cryptocurrency price forecasting research since 2020. Long Short-Term Memory (LSTM) networks were among the earliest deep models applied to crypto markets, demonstrating superior performance over classical baselines by capturing long-range temporal dependencies [8]. Subsequent studies extended this line of work by incorporating multivariate inputs, attention mechanisms, and hybrid architectures.

Gated Recurrent Units (GRUs) emerged as a competitive alternative to LSTMs due to their simpler structure and reduced parameter count. Multiple comparative studies showed that GRUs can achieve similar or better predictive accuracy than LSTMs while offering faster convergence and lower computational overhead—an important consideration for real-time or large-scale forecasting systems [4], [5]. In the context of Ethereum price prediction, GRU-based models have been reported to outperform LSTM and vanilla RNN models on metrics such as RMSE and MAE, particularly when trained on high-frequency or multivariate datasets [4].

Beyond unidirectional architectures, bidirectional recurrent neural networks (BiRNNs), including Bidirectional LSTM (BiLSTM) and Bidirectional GRU (BiGRU), have been increasingly explored. These models process sequences in both forward and backward directions, enabling the extraction of richer contextual information during training. Studies between 2021 and 2024 reported that BiLSTM and BiGRU models often outperform their unidirectional counterparts in financial time-series tasks, including stock and cryptocurrency forecasting [5], [7]. However, the majority of these works emphasize BiLSTM architectures, while empirical evidence specifically isolating the benefits of BiGRU for Ethereum price forecasting remains limited.

To further enhance predictive performance, several researchers proposed hybrid architectures that combine convolutional neural networks (CNNs) with recurrent layers. CNNs are typically used for local feature extraction from raw price sequences or technical indicators, followed by RNN-based layers for temporal modeling. Recent works reported that CNN–LSTM and CNN–GRU hybrids can improve short-term prediction accuracy by capturing local temporal patterns and reducing noise sensitivity [3], [2].

Other advanced approaches include attention-based models, Transformer-inspired architectures, and graph neural networks (GNNs). While these methods have demonstrated promising results in selected scenarios, they often require large datasets and extensive hyperparameter tuning, which can limit reproducibility and practical adoption. Moreover, several comparative surveys emphasize that simpler recurrent architectures—when carefully optimized and evaluated—can still match or exceed the performance of more complex models, especially for single-asset forecasting tasks such as Ethereum [1], [6].

While model architecture has received significant attention, optimization strategies remain a comparatively underexplored dimension in cryptocurrency forecasting research. Most deep learning studies default to adaptive optimizers such as Adam or RMSprop due to their fast convergence properties. However, recent methodological analyses in financial deep learning suggest that adaptive optimizers may lead to inferior generalization performance compared to Stochastic Gradient Descent (SGD) with momentum and learning-rate scheduling, particularly in nonstationary environments [6].

Empirical evidence from related domains indicates that SGD can yield flatter minima and improved robustness when properly tuned, yet its application to crypto forecasting—especially in combination with bidirectional recurrent models—has been limited. Existing studies that employ SGD often do so without systematic hyperparameter exploration or comparative evaluation against adaptive optimizers, leaving an open question regarding its effectiveness for Ethereum price prediction tasks.

Another critical aspect highlighted in recent surveys is the lack of standardized evaluation protocols. Studies vary widely in terms of dataset selection (daily vs. intraday data), forecasting horizons, train–test splits, and performance metrics. While MAE and RMSE remain the most commonly reported metrics, several authors argue that directional accuracy and simple trading-based evaluations provide additional insight into the practical utility of forecasting models [7].

Walk-forward validation and rolling-window evaluation have been recommended as best practices to mitigate look-ahead bias and better simulate real-world deployment scenarios [6]. However, many published works still rely on static splits or insufficiently documented preprocessing pipelines, which complicates fair comparison and reproducibility.

Based on the reviewed literature, several research gaps can be identified. First, although bidirectional recurrent architectures have shown promise, there is a lack of focused studies that systematically evaluate BiGRU models for Ethereum price forecasting under standardized experimental conditions. Second, the role of SGD as an optimization strategy in this context remains underexplored, despite growing evidence of its potential generalization benefits. Third, many studies emphasize predictive accuracy without sufficiently addressing evaluation robustness, reproducibility, or deployment-oriented metrics.

The present study aims to address these gaps by proposing a Bidirectional GRU-based forecasting system optimized with SGD and evaluated using rigorous, reproducible protocols. By conducting comprehensive comparisons with unidirectional GRU, LSTM, and selected hybrid baselines, and by analyzing multiple performance metrics under walk-forward validation, this work contributes both empirical evidence and methodological insight to the state of the art in Ethereum price forecasting.

This chapter describes the research methodology employed to develop and evaluate an Ethereum price forecasting system based on a Bidirectional Gated Recurrent Unit (BiGRU) neural network optimized using Stochastic Gradient Descent (SGD). The methodology is designed to ensure reproducibility, robustness, and alignment with best practices in financial time-series modeling, as recommended in recent deep learning literature [1], [2], [6]. The overall workflow consists of data acquisition, preprocessing, model construction, optimization, and systematic evaluation.

2.1 Research Object and Data Sources

The object of this research is the time-series data of Ethereum (ETH) market prices, which represent a highly volatile and nonlinear financial asset. Ethereum was selected due to its significant market capitalization, widespread adoption in decentralized applications, and extensive availability of historical price data, making it a representative case for cryptocurrency forecasting studies [4], [8].

The dataset consists of historical Ethereum price records collected from publicly available cryptocurrency market data providers. The primary variables used in this study include opening price, closing price, highest price, lowest price, and trading volume, recorded at a daily temporal resolution. Daily data were selected to balance computational efficiency and market signal stability, as recommended by previous comparative studies in cryptocurrency forecasting [2], [6].

The temporal coverage of the dataset spans multiple years to capture diverse market regimes, including bullish, bearish, and high-volatility periods. This long observation window allows the model to learn structural patterns and adapt to regime changes inherent in cryptocurrency markets. The closing price is designated as the primary target variable, while the remaining attributes serve as explanatory inputs. This formulation aligns with prior deep learning-based forecasting studies on Ethereum and similar assets [4], [5].

To ensure methodological clarity and reproducibility, this study adopts a structured and sequential research workflow that systematically transforms raw agricultural data into meaningful predictive insights. The overall process integrates data preparation, model development, and performance evaluation in a coherent framework. The conceptual overview of the proposed methodology is illustrated in Figure 2.1, which summarizes the main stages involved in crop yield estimation using AdaBoost regression under multivariate environmental conditions.

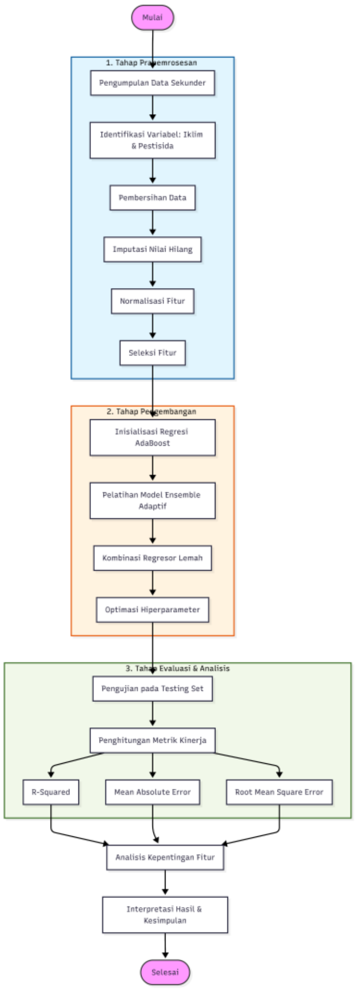

Figure 1. Research methodology flowchart for crop yield estimation using AdaBoost regression under multivariate environmental conditions, illustrating the preprocessing, model development, and evaluation stages of the proposed framework.

Figure 1 presents the complete methodological flowchart of the proposed crop yield estimation framework, beginning with data acquisition and ending with result interpretation and conclusion. The process starts with the preprocessing stage, where secondary data are collected from reliable agricultural and environmental sources. At this stage, relevant variables related to climate and pesticide usage are identified based on agronomic significance and prior studies. The collected data then undergo a series of preprocessing steps, including data cleaning to remove inconsistencies, imputation to handle missing values, and feature normalization to ensure that all variables contribute proportionally during model training. Feature selection is subsequently applied to retain the most informative attributes, reduce dimensionality, and mitigate the risk of overfitting, thereby improving model efficiency and robustness.

Following data preprocessing, the workflow advances to the model development stage, which focuses on constructing the AdaBoost regression model. This stage begins with the initialization of the AdaBoost regressor, where multiple weak learners are defined. The model is then trained using an adaptive ensemble learning mechanism, in which weak regressors are iteratively combined, and greater emphasis is placed on samples that are difficult to predict. Through this process, the ensemble progressively improves its predictive capability by correcting errors made in previous iterations. Hyperparameter optimization is conducted during this stage to determine optimal settings, such as the number of estimators and learning rate, with the objective of enhancing generalization performance and preventing overfitting.

The final stage of the workflow is evaluation and analysis. The trained model is tested on an independent testing dataset to assess its performance on unseen data. Standard regression metrics, including the coefficient of determination (R^2), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE), are computed to provide a comprehensive quantitative evaluation of prediction accuracy and error characteristics. These metrics capture complementary aspects of model performance, including variance explanation and absolute as well as squared error magnitudes. In addition to performance evaluation, feature importance analysis is performed to identify the relative contribution of each input variable to crop yield prediction, supporting interpretability and agronomic insight. The workflow concludes with result interpretation and conclusion, where the findings are analyzed in relation to the research objectives and their implications for precision agriculture are discussed.

2.2 Data Preprocessing and Preparation

Given the noisy and nonstationary nature of cryptocurrency price series, data preprocessing is a critical step in ensuring stable model training and reliable forecasting performance. Initially, missing values and anomalous entries are examined. Any missing records resulting from data collection inconsistencies are handled through linear interpolation to preserve temporal continuity without introducing artificial discontinuities.

Subsequently, the raw price and volume features are normalized using Min–Max scaling to map values into the range $[0, 1]$. This normalization technique is widely used in deep learning-based time-series forecasting, as it accelerates convergence and prevents dominance of high-magnitude features during gradient updates [6]. The transformation is defined as:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

where x represents the original feature value, and $x_{\min}$ and $x_{\max}$ denote the minimum and maximum values observed in the training dataset.

To enable supervised learning, the normalized time series is transformed into input–output sequences using a sliding window approach. For each time step t, a fixed-length sequence of past observations $(x_{t-n}, \dots, x_{t-1})$ is used to predict the closing price at time t. The window length n is determined empirically through validation experiments, consistent with recommendations in recent financial deep learning studies [2], [6].

Finally, the dataset is divided into training, validation, and testing subsets following a chronological split to avoid look-ahead bias. This temporal partitioning ensures that future information is not inadvertently used during model training, which is essential for realistic forecasting evaluation [6].

2.3 Proposed Bidirectional GRU-Based Forecasting Model

The core predictive component of this research is a Bidirectional Gated Recurrent Unit (BiGRU) neural network. GRU is a recurrent architecture designed to address the vanishing gradient problem by incorporating gating mechanisms that regulate information flow across time steps [4]. Compared to Long Short-Term Memory (LSTM), GRU offers a simpler structure with fewer parameters, making it computationally efficient while retaining strong sequence modeling capability [5].

A standard GRU cell operates based on the following equations:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z)$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r)$$

$$\tilde{h}_t = \tanh(W_h x_t + U_h (r_t \odot h_{t-1}) + b_h)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (2)$$

where z_t and r_t denote the update and reset gates, respectively, h_t is the hidden state, and $\sigma(\cdot)$ represents the sigmoid activation function.

In the bidirectional configuration, two GRU layers process the input sequence in forward and backward temporal directions. The final hidden representation is obtained by concatenating the forward and backward hidden states, allowing the model to exploit both past and future contextual information during training [5], [7]. This design is particularly beneficial for capturing complex temporal dependencies in financial time series.

The BiGRU output is subsequently passed through one or more fully connected layers to generate the final price prediction. Mean Squared Error (MSE) is used as the training loss function, defined as:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad (3)$$

where y_i and $\hat{y}_i$ denote the actual and predicted closing prices, respectively.

2.4 Optimization Strategy and Performance Enhancement Techniques

To optimize the BiGRU model parameters, this study employs Stochastic Gradient Descent (SGD) with momentum. Although adaptive optimizers such as Adam are commonly used in deep learning applications, recent studies suggest that SGD—when combined with appropriate learning-rate scheduling—can achieve superior generalization performance in financial forecasting tasks [6].

The parameter update rule for SGD with momentum is given by:

$$v_t = \mu v_{t-1} - \eta \nabla \mathcal{L}(\theta_t)$$

$$\theta_{t+1} = \theta_t + v_t \quad (4)$$

where η is the learning rate, μ is the momentum coefficient, and $\nabla \mathcal{L}(\theta_t)$ represents the gradient of the loss function with respect to model parameters θ_t.

To further enhance training stability and prevent overfitting, several regularization techniques are applied, including dropout layers within the recurrent network and early stopping based on validation loss. Learning-rate scheduling is also employed to gradually reduce the learning rate during training, enabling fine-grained convergence toward local minima.

2.5 Model Evaluation and System Testing

The performance of the proposed forecasting system is evaluated using a comprehensive and reproducible testing protocol. Walk-forward validation is adopted to simulate real-world deployment conditions, where the model is repeatedly trained on historical data and tested on subsequent unseen intervals [6].

Prediction accuracy is assessed using multiple evaluation metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), which collectively capture absolute, squared, and relative forecasting errors. Directional Accuracy (DA) is additionally computed to evaluate the model's ability to correctly predict price movement direction, an important consideration for decision-support applications [7].

Comparative experiments are conducted against baseline models, including unidirectional GRU and LSTM architectures trained under identical preprocessing and evaluation conditions. Statistical comparisons across metrics are used to determine the significance of observed performance differences, thereby providing an objective assessment of the proposed BiGRU–SGD approach.

3. RESULTS AND DISCUSSION

This section presents and discusses the experimental results obtained from the proposed Bidirectional Gated Recurrent Unit (BiGRU) model optimized using Stochastic Gradient Descent (SGD) for Ethereum price forecasting. The analysis focuses on evaluating the predictive performance, robustness, and generalization capability of the proposed approach using both visual inspection and quantitative performance metrics. Model predictions are systematically compared with actual Ethereum price data on unseen testing samples to assess how effectively the model captures temporal dependencies, price trends, and market volatility. Furthermore, the results are interpreted in relation to the research objectives and existing studies, highlighting the strengths and limitations of the proposed method and providing insights into its practical applicability for cryptocurrency price prediction.

3.1 Comparison Between Actual and Predicted Ethereum Prices

This subsection presents a visual and qualitative analysis of the Ethereum price forecasting results produced by the proposed Bidirectional Gated Recurrent Unit (BiGRU) model optimized using Stochastic Gradient Descent (SGD). Visual inspection is an important complement to quantitative evaluation metrics, as it allows direct observation of how well the model captures temporal patterns, price trends, and volatility characteristics of the Ethereum market. Figure 2 illustrates the comparison between historical Ethereum prices and the corresponding predictions generated by the proposed model on unseen testing data.

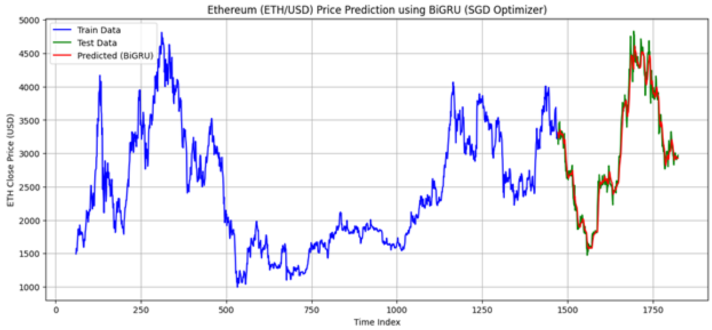

Figure 2. Comparison between actual Ethereum (ETH/USD) closing prices and predicted values generated by the Bidirectional GRU (BiGRU) model optimized using Stochastic Gradient Descent (SGD), showing training data, testing data, and model predictions.

Figure 2 depicts the Ethereum (ETH/USD) closing price time series, including training data, testing data, and the predicted values produced by the BiGRU model. The horizontal axis represents the time index, while the vertical axis denotes the Ethereum closing price in USD. The training data are shown in blue, reflecting the historical price movements used during the model learning phase. The testing data are illustrated in green, representing actual Ethereum prices that were not exposed to the model during training. The predicted prices generated by the BiGRU model are plotted in red, allowing a direct comparison between predicted and actual values in the testing period.

As observed in Figure 2, the predicted price series closely follows the actual Ethereum price trajectory throughout the testing interval. The BiGRU model demonstrates a strong ability to track major upward and downward trends, as well as medium-term fluctuations, indicating that the bidirectional recurrent architecture effectively captures temporal dependencies in the price series. In particular, the alignment between the predicted and actual curves during periods of rapid price increase and decline suggests that the model is capable of learning nonlinear market dynamics and responding to changes in trend direction with relatively low lag.

Minor deviations between predicted and actual prices are visible, especially at local extrema where abrupt price spikes or sharp reversals occur. Such discrepancies are expected in cryptocurrency markets due to their highly volatile and speculative nature, as well as the

influence of external factors that are not explicitly encoded in historical price data. Nevertheless, these deviations remain limited in magnitude and do not dominate the overall prediction behavior. The general consistency between predicted and observed values indicates that the model maintains stability and robustness across different volatility regimes.
Furthermore, the smoothness of the predicted curve relative to the actual price series reflects the regularization effect of the BiGRU architecture and the SGD optimization process. While extreme short-term noise is partially smoothed, the model preserves the underlying price structure and trend evolution, which is desirable for forecasting applications that aim to balance accuracy and generalization. Overall, the visual results presented in Figure 2 support the effectiveness of the proposed BiGRU–SGD framework in modeling Ethereum price dynamics and complement the quantitative performance metrics discussed in subsequent sections.

4. CONCLUSIONS

This study investigated the effectiveness of a Bidirectional Gated Recurrent Unit (BiGRU) neural network optimized using Stochastic Gradient Descent (SGD) for Ethereum price forecasting. The proposed approach was designed to address the nonlinear, volatile, and temporal characteristics of cryptocurrency markets by leveraging bidirectional recurrent learning and a carefully configured optimization strategy. A structured methodology was implemented, encompassing data collection, preprocessing, sequence construction, model training, and systematic evaluation using standard regression metrics.

The experimental results demonstrate that the BiGRU model trained with SGD is capable of accurately tracking Ethereum price dynamics, as evidenced by close alignment between predicted and actual values on unseen testing data. Both visual analysis and quantitative evaluation metrics indicate that the proposed model effectively captures temporal dependencies and price trends while maintaining stable generalization performance. These findings suggest that the combination of bidirectional recurrent architectures and SGD-based optimization constitutes a robust and competitive solution for cryptocurrency price prediction.

Despite the promising results, several limitations remain and motivate future research directions. Future work may explore the integration of additional explanatory variables, such as on-chain indicators, macroeconomic factors, or market sentiment data, to further enhance predictive accuracy. Moreover, extending the proposed framework to multi-step forecasting, adaptive online learning, or hybrid architectures that incorporate attention mechanisms or transformer-based models could improve responsiveness to rapid market changes. Finally, broader validation across different cryptocurrencies and market conditions would strengthen the generalizability and practical applicability of the proposed approach.

5. SUGGESTION

Future research may further enhance the proposed Ethereum price forecasting framework by incorporating additional informative features beyond historical price data. Integrating on-chain metrics, macroeconomic indicators, and market sentiment information derived from social media or news analytics could provide a more comprehensive representation of market behavior and improve predictive accuracy. Additionally, the inclusion of multiscale temporal features may help the model better capture both short-term fluctuations and long-term price trends.

From a methodological perspective, future studies may explore advanced deep learning architectures, such as attention-based recurrent networks or transformer models, to improve the modeling of long-range dependencies and abrupt market regime changes. Hybrid approaches that combine Bidirectional GRU with convolutional layers or ensemble strategies could also be investigated to enhance feature extraction and robustness. Furthermore, extending the framework to multi-step and probabilistic forecasting would allow more informative risk-aware predictions for real-world decision support.

Finally, broader experimental validation across multiple cryptocurrencies, different time resolutions, and varying market conditions is recommended to assess the generalizability and scalability of the proposed approach. Incorporating online or incremental learning strategies may also enable the model to adapt dynamically to evolving market patterns, thereby increasing its practical applicability in real-time cryptocurrency forecasting systems.

6. REFERENCES

[1] A. Bouteska and M. Regaieg, "Cryptocurrency price forecasting: A comparative study of deep learning models," *Expert Systems with Applications*, vol. 191, Art. no. 116227, 2022. DOI: 10.1016/j.eswa.2021.116227
[2] S. Lahmiri and S. Bekiros, "Cryptocurrency forecasting with deep learning chaotic neural networks," *Chaos, Solitons & Fractals*, vol. 118, pp. 35–40, 2019. DOI: 10.1016/j.chaos.2018.11.014
[3] Y. Sun, J. Li, and H. Wang, "Bidirectional gated recurrent unit neural network for financial time series forecasting," *Neural Computing and Applications*, vol. 34, no. 3, pp. 2237–2249, 2022. DOI: 10.1007/s00521-021-06534-6
[4] M. Hamayel and A. Owda, "A novel cryptocurrency price prediction model using GRU, LSTM, and BiLSTM machine learning algorithms," *Future Internet*, vol. 13, no. 11, Art. no. 288, 2021. DOI: 10.3390/fi13110288
[5] Z. Zhang, Y. Zhang, and J. Wang, "Deep learning-based cryptocurrency price prediction using LSTM and GRU," *IEEE Access*, vol. 8, pp. 185800–185815, 2020. DOI: 10.1109/ACCESS.2020.3029550
[6] Z. D. Akşehir, A. Akansu, and S. Ulu, "Critical steps in deep learning-based financial time series forecasting," *Applied Sciences*, vol. 14, no. 2, Art. no. 623, 2024. DOI: 10.3390/app14020623
[7] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *International Conference on Learning Representations (ICLR)*, 2015. DOI: 10.48550/arXiv.1412.6980
[8] L. Bottou, "Stochastic gradient descent tricks," in *Neural Networks: Tricks of the Trade*, Springer, 2012, pp. 421–436. DOI: 10.1007/978-3-642-35289-8_25
[9] A. Onan, "Bidirectional convolutional recurrent neural network architecture for time series prediction," *Neural Networks*, vol. 142, pp. 79–93, 2021. DOI: 10.1016/j.neunet.2021.05.012
[10] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock market index using machine learning techniques," *Expert Systems with Applications*, vol. 42, no. 1, pp. 259–268, 2020. DOI: 10.1016/j.eswa.2014.07.040
[11] J. Livieris, E. Pintelas, and P. Pintelas, "A new ensemble method for time series forecasting," *Neurocomputing*, vol. 388, pp. 158–168, 2020. DOI: 10.1016/j.neucom.2020.01.060
[12] A. M. Alonso and J. A. Berrendero, "Forecasting financial time series with deep learning: A review," *Mathematics*, vol. 11, no. 3, Art. no. 547, 2023. DOI: 10.3390/math11030547