*IOSR Journal of Economics and Finance (IOSR-JEF)*
*e-ISSN: 2321-5933, p-ISSN: 2321-5925, Volume 16, Issue 3 Ser.6 (May - June 2025), PP 70-82*
*www.iosrjournals.org*

# Forecasting Bitcoin Prices: A Comparative Study of Deep Learning and Hybrid Models with Shannon Entropy for Information Measurement

[1] Meriem EL Arfaoui, [2] Houda lechheb, [3] Hicham ouakil, [3] Zakaria Salhi

[1] *Laboratory of economic sciences and public policies, faculty of economics and management, Ibn tofail university, kenitra, Morocco; meriem.elarfaoui@uit.ac.ma*

[2] *Laboratory of economic sciences and public policies, faculty of economics and management, Ibn tofail university, kenitra, Morocco; houda.lechheb@uit.ac.ma*

[3] *Laboratory of economic sciences and public policies, faculty of economics and management, Ibn tofail university, kenitra, Morocco; hicham.ouakil@uit.ac.ma*

[3] *Laboratory of economic sciences and public policies, faculty of economics and management, Ibn tofail university, kenitra, Morocco; zakaria.salhi@uit.ac.ma*

***Abstract****:*
*This study investigates Bitcoin price prediction using deep learning models and hybrid approaches, integrating Shannon entropy to quantify market unpredictability. Predicting cryptocurrency prices is challenging due to their high volatility and complex dynamics driven by macroeconomic, regulatory, and speculative factors. We compare the performance of standalone deep learning models, including Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), and Bidirectional LSTM, with a hybrid GRU-ARIMA model that combines machine learning and statistical residual correction techniques. Data from 2015 to 2024 are used for model training and evaluation. Model accuracy is assessed using Mean Squared Error (MSE) and Mean Absolute Error (MAE), while Shannon entropy is computed to measure the information content in actual prices and predictions. Results show that hybrid models outperform standalone architectures in terms of predictive accuracy, highlighting the benefits of residual analysis for capturing complex price dynamics. Entropy analysis reveals a correlation between higher entropy and model performance, providing insights into market efficiency and forecastability. This research contributes to the growing body of literature on cryptocurrency forecasting by emphasizing the role of hybrid models and entropy-based information measures. Future research directions include exploring alternative hybridization strategies and extending entropy measures for enhanced market analysis.*

***Keywords:*** *Bitcoin Prices, deep learning, hybrid model, LSTMs, GRUs,BI-LSTM,neural networks, Shannon entropy, volatility*

---

Date of Submission: 05-07-2025 Date of Acceptance: 16-07-2025

---

## I. Introduction:

The rapid evolution of financial technologies and the rise of cryptocurrencies have transformed the landscape of modern finance. Bitcoin, as the pioneering and most prominent cryptocurrency, has become a focal point for researchers and practitioners due to its extreme price volatility and speculative trading behavior. Unlike traditional financial assets, Bitcoin prices are influenced by a unique set of factors, including technological innovations, regulatory developments, macroeconomic trends, and social media sentiment. These characteristics make accurate forecasting of Bitcoin prices a formidable yet vital challenge for portfolio management, risk assessment, and trading strategies.

Traditional time series models, such as Autoregressive Integrated Moving Average (ARIMA), have been widely applied to financial data; however, their linear nature limits their ability to capture the complex, nonlinear dependencies inherent in cryptocurrency markets. In contrast, deep learning models, including Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRU), have demonstrated superior performance in capturing long-term temporal dependencies in highly volatile time series data. Recent advances have also explored hybrid approaches, combining deep learning and statistical models, to improve prediction accuracy by modeling both nonlinear patterns and residual error structures.

In this study, we compare the predictive performance of standalone deep learning models (LSTM, GRU, Bidirectional LSTM) and a hybrid GRU-ARIMA model. Additionally, we incorporate Shannon entropy, an information-theoretic measure, to evaluate the uncertainty and information content within the predicted prices. By quantifying the entropy of actual prices and model outputs, we gain insights into market efficiency and the reliability of various forecasting methods. This research aims to enhance the understanding of Bitcoin price dynamics and contribute to the development of more effective predictive frameworks for cryptocurrency markets.

## II. Literature Review and Theoretical Background

### 2.1 **Cryptocurrency Market Dynamics**

The cryptocurrency market, led by Bitcoin, is a disruptive force in modern finance, redefining traditional paradigms of investment and currency exchange. Bitcoin, the first decentralized digital currency introduced in 2009 by the pseudonymous Satoshi Nakamoto, serves as both a medium of exchange and a store of value. Its decentralized nature eliminates the need for intermediaries like banks, making it appealing in an era of digital transformation and financial democratization. A defining feature of cryptocurrencies is their high volatility. Bitcoin's price, for example, has exhibited rapid and extreme fluctuations, driven by a unique set of factors distinct from conventional financial markets. These include:

Cryptocurrency prices exhibit significant volatility, influenced by a range of factors that distinguish them from traditional equities or commodities. One key driver is technological development, where innovations and protocol upgrades impact market confidence and adoption. For instance, Bitcoin's implementation of the Lightning Network enhanced scalability by enabling faster, cheaper transactions, directly affecting its valuation. Similarly, network forks, such as the creation of Bitcoin Cash, introduce uncertainty and shift investor allocations between competing chains. Another crucial factor is the regulatory environment, as cryptocurrencies are highly sensitive to policy changes. Announcements from major economies like the United States or the European Union regarding regulations, taxation, or potential bans can trigger significant price swings. While regulatory clarity encourages institutional investment, restrictive measures often lead to market sell-offs. Additionally, macroeconomic conditions play a role in cryptocurrency price fluctuations. In periods of high inflation or geopolitical instability, cryptocurrencies often serve as alternative assets akin to digital gold. However, their correlation with traditional risk-on assets, especially during financial crises, complicates their role as a safe haven. For example, Bitcoin's performance during the COVID-19 pandemic demonstrated both its speculative nature and its emerging function as a hedge against monetary debasement. Lastly, market sentiment and speculative behavior heavily influence cryptocurrency prices. News sentiment, social media trends, and statements from influential figures, such as Elon Musk's tweets, can trigger rapid price swings. Furthermore, automated trading strategies and high-frequency trading amplify this volatility, making the cryptocurrency market highly reactive to external stimuli.

Given these multifaceted dynamics, predicting cryptocurrency prices poses unique challenges. Unlike traditional financial models based on fundamental or technical analysis, cryptocurrencies require tools that can capture both deterministic and stochastic behaviors. Advanced machine learning models, such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks, have shown promise in handling time-series data with complex temporal dependencies. Additionally, information-theoretic measures like Shannon entropy provide a robust framework for quantifying market uncertainty and the randomness of price changes. Higher entropy values indicate greater unpredictability, a characteristic often associated with the cryptocurrency market. Understanding these dynamics through rigorous empirical analysis not only enhances forecasting accuracy but also informs portfolio management strategies, hedging approaches, and regulatory frameworks. As cryptocurrencies continue to evolve, integrating multifractal analysis and deep learning techniques will be crucial for comprehensive market modeling.

### 2.2 **Time Series Forecasting in Financial Markets**

Forecasting financial time series, particularly price movements, is a complex endeavor that involves identifying historical patterns and deciphering underlying structural dependencies. Traditional models, such as Autoregressive Integrated Moving Average (ARIMA), have been widely employed due to their strong theoretical foundations and ease of interpretation. ARIMA models combine autoregression and moving average components, along with differencing to ensure stationarity, making them effective for linear patterns and short-term dependencies. However, these models assume a linear relationship among variables, a limitation that reduces their effectiveness in highly dynamic and nonlinear environments, such as cryptocurrency markets.

**Limitations of Traditional Models**

Cryptocurrency price data exhibit characteristics such as heavy tails, volatility clustering, and long memory — features that are poorly captured by linear models. The ARIMA framework cannot adapt to changing market regimes or complex feedback loops driven by investor behavior, regulatory news, and technological innovations. Additionally, traditional models struggle with the erratic, non-stationary nature of cryptocurrency prices, where sudden jumps or crashes are commonplace. Other statistical approaches, including GARCH and EGARCH, attempt to model volatility by accounting for heteroskedasticity. These models are more flexible in capturing time-varying variance but still rely on assumptions of stationarity and linearity, which limit their robustness in rapidly evolving markets.

2.3 **The Emergence of Nonlinear and Machine Learning Models**

Given the intricate behavior of financial markets, particularly in cryptocurrencies, there has been a significant shift toward machine learning techniques capable of capturing nonlinearities and complex temporal dependencies. Recurrent Neural Networks (RNNs), including Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), are designed to handle sequential data by retaining long-term dependencies. Unlike ARIMA, these models do not require manual differencing or lag selection, as they learn patterns directly from the data.

LSTM networks are specifically designed to address the vanishing gradient problem that affects traditional RNNs. By utilizing memory cells and gating mechanisms, they effectively retain relevant information over extended sequences, making them particularly suitable for detecting temporal patterns and forecasting price movements in highly volatile markets. A more streamlined alternative, GRU networks offer a simplified architecture with fewer parameters while maintaining comparable performance. This reduction in complexity makes GRUs computationally efficient and well-suited for real-time applications.

**Information-Theoretic Approaches and Entropy Measures**

In addition to neural networks, entropy-based methods, such as Shannon entropy, offer powerful tools for understanding market uncertainty. Entropy quantifies the randomness and unpredictability of a time series. Higher entropy values are associated with greater complexity, reflecting the chaotic nature of financial markets like cryptocurrencies. Combining entropy measures with advanced machine learning models provides a hybrid approach that captures both the stochasticity and structure of price dynamics

2.**4 Deep Learning for Time Series Prediction**

In recent years, deep learning has revolutionized time series forecasting by addressing the limitations of traditional models, particularly in capturing nonlinearities and complex temporal patterns. Unlike conventional statistical methods, deep learning models automatically learn features from data without the need for explicit assumptions about its structure. Recurrent Neural Networks (RNNs) and their advanced variants, Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), have emerged as powerful tools for sequential data modeling.

**Recurrent Neural Networks (RNNs)**

RNNs are designed for sequence-based tasks by using loops within the network to retain information from previous time steps. However, standard RNNs suffer from the vanishing gradient problem, which hampers their ability to capture long-term dependencies. This limitation has driven the development of more robust architectures.

**Long Short-Term Memory (LSTM) Networks**

LSTM networks are a specialized form of RNNs explicitly built to overcome the vanishing gradient issue. They employ memory cells and three types of gates (input, forget, and output) to regulate the flow of information. The input gate decides what new information to store, the forget gate determines which data to discard, and the output gate controls how much of the stored information influences the next layer. These gating mechanisms enable LSTMs to maintain relevant context over long sequences, making them highly effective for financial time series predictions, where price movements are influenced by long-term trends and sudden shocks alike.LSTMs have been successfully applied in various financial contexts, including stock price prediction, volatility forecasting, and market sentiment analysis. Their strength lies in capturing intricate relationships in noisy and volatile data, such as Bitcoin and other cryptocurrency prices.

**Gated Recurrent Units (GRUs)**

GRUs are a simplified variant of LSTM networks, designed to improve computational efficiency without significantly compromising performance. GRUs utilize two gates:

the reset gate and the update gate, which serve similar purposes as the input and forget gates in LSTMs but with a streamlined structure. The reset gate determines how much of the past information to forget, while the update gate controls how much of the new information to pass to the next time step. Compared to LSTMs, GRUs have fewer parameters, making them faster to train and suitable for scenarios requiring real-time prediction. Research indicates that GRUs often achieve comparable results to LSTMs on many forecasting tasks while reducing computational costs, which is particularly advantageous for high-frequency financial data.

2.5 **Shannon Entropy for Market Uncertainty**
Shannon entropy, originally developed in information theory by Claude Shannon in 1948, serves as a fundamental measure of uncertainty and information content within a system. In the context of financial markets, Shannon entropy is used to assess the unpredictability and randomness of price movements. Higher entropy values correspond to greater uncertainty and reduced predictability, while lower entropy signifies more structured and predictable behavior. This concept provides valuable insights into market efficiency and risk assessment.

**Theoretical Foundation of Shannon Entropy**

Shannon entropy $H$ is defined for a probability distribution $\mathrm{P} = \{p_i\}$ as:

$$\boldsymbol{H}\,(\boldsymbol{P}) = -\sum_{i=1}^{n} \boldsymbol{p_i} \boldsymbol{log}(\boldsymbol{p_i})$$

where $\boldsymbol{p_i}$ represents the probability of a particular state or outcome, and $\boldsymbol{n}$ is the number of possible states. In financial time series, these probabilities are typically derived from the distribution of price returns or other relevant variables.

**Application to Cryptocurrency Markets**

Cryptocurrencies, with their high volatility and susceptibility to rapid price swings, are an ideal case for entropy analysis. Shannon entropy allows researchers to quantify how much randomness or deterministic structure is present in the returns of Bitcoin and other digital assets. High entropy in these markets reflects speculative behavior, news-driven price changes, and the absence of strong directional trends. Conversely, periods of low entropy may indicate dominant market forces or significant price patterns that can be exploited for trading strategies.

**Insights into Market Efficiency and Risk**
By measuring entropy over time, investors and analysts can gain insights into changing market conditions and potential risk. A rising entropy value may signal increased market instability or heightened reaction to external shocks, while declining entropy could indicate more orderly price movements. Entropy analysis complements other risk metrics by providing a non-parametric, distribution-free method for evaluating price dynamics, making it a powerful tool for assessing financial complexity.

## III. Empirical Study

3.1 **Data Description**
The study utilizes historical Bitcoin price data from January 2015 to January 2024, sourced from Yahoo Finance. The dataset consists of daily closing prices, capturing Bitcoin's dynamic market behavior. Data preprocessing steps include handling missing values and scaling the data using MinMaxScaler to normalize the values between 0 and 1 for effective model training. The entire data analysis and modeling process was conducted using Python version 4.0.11, ensuring compatibility with the latest libraries and tools for time series forecasting and signal processing.

3.2. **Problem statement:**
Bitcoin's high volatility presents a significant challenge for accurate price prediction and effective risk management. The rapid and unpredictable fluctuations in its value stem from a variety of factors, including market speculation, regulatory news, and macroeconomic events. Such volatility complicates the task of developing reliable forecasting models for investors and analysts. The primary objective of this study is to conduct a comparative analysis of four predictive models LSTM, GRU, Bidirectional LSTM (BI-LSTM), and a Hybrid GRU-ARIMA model to evaluate their effectiveness in forecasting Bitcoin prices. These models are chosen for their diverse approaches, ranging from pure deep learning architectures to hybrid techniques that integrate statistical modeling. The second objective is to quantify the informational content and complexity of Bitcoin's price dynamics using Shannon entropy, a statistical measure that captures the degree of uncertainty and randomness within the time series.
This dual approach provides both a performance evaluation of modern prediction frameworks and a deeper understanding of the underlying informational structure of Bitcoin's price movements, offering valuable insights for improving forecasting accuracy and developing strategies to manage market risk.

**3.3. Preprocessing data:**
The data preprocessing stage is crucial for preparing time series data for modeling and analysis. Initially, data cleaning involves removing any missing or invalid values to ensure consistency and reliability. In this study, we

transform the original Bitcoin price data into log returns, calculated as the natural logarithm of successive price ratios. This step captures relative changes in price, making the series stationary and suitable for predictive modeling. Following this, input-output sequences are created for each deep learning model (LSTM, GRU, Bidirectional LSTM, and Hybrid GRU-ARIMA) by defining a fixed window of historical returns as input to predict future values. This sequence generation captures temporal dependencies essential for forecasting. Each model is then built with specific configurations:

the LSTM and GRU use recurrent units to learn sequential patterns, the Bidirectional LSTM processes data in forward and backward directions for enhanced feature extraction, and the Hybrid GRU-ARIMA combines GRU's deep learning capabilities with ARIMA's statistical modeling of residuals. This comprehensive preprocessing framework lays the foundation for robust time series predictions and entropy-based information quantification.

**3.4. Processing data:**

In the data processing and evaluation phase, various metrics are employed to assess the predictive performance of each model. Key error metrics include Mean Squared Error (MSE) and Mean Absolute Error (MAE). MSE penalizes larger prediction errors more heavily by squaring the residuals before averaging, making it sensitive to outliers, while MAE provides a more balanced assessment by averaging the absolute differences between predicted and actual values. Both metrics are critical for understanding different aspects of model accuracy.

Additionally, accuracy metrics specific to time series forecasting, such as directional accuracy, may be used to evaluate how well the model predicts the direction of price movements. To ensure robust performance evaluation, we apply Time Series Cross-Validation, a method designed for sequential data. Unlike traditional cross-validation that assumes data independence, this technique splits the dataset into multiple overlapping training and testing sets, maintaining temporal order. Each split incrementally increases the training size while testing on the subsequent data points, capturing the evolving patterns of financial time series. This systematic evaluation provides a comprehensive understanding of model behavior, avoiding overfitting and enhancing generalization. Combining these evaluation strategies ensures a holistic assessment of prediction accuracy and information content quantification.

**3.5. Model Implementation**

In the Model Implementation stage, we implemented several predictive models for Bitcoin price forecasting using Python and the Keras library to prepare the data for training and testing. We utilized Python with the Keras library, a high-level neural network API designed for building and training deep learning models. Keras, written in Python, provides a user-friendly interface and is capable of running on top of powerful backend engines such as TensorFlow, CNTK, or Theano. Time Series Cross-Validation was utilized to ensure robust model evaluation across different time periods. We focused on minimizing prediction errors by calculating the Mean Squared Error (MSE) as the primary loss function, which penalizes larger errors more significantly. Additionally, we quantified information content using Shannon entropy to assess market predictability. Higher entropy values suggest greater uncertainty, indicating a more unpredictable market, while lower values imply more structured patterns that can be exploited for forecasting. For model selection, the LSTM and GRU architectures were chosen due to their recurrent connections, allowing them to effectively capture sequential dependencies in time series data. The Bidirectional LSTM (BI-LSTM) extends the standard LSTM by processing sequences in both forward and backward directions, enhancing predictive accuracy by considering future and past information. The hybrid GRU-ARMA model combines the non-linear pattern recognition strength of GRU with ARMA's linear statistical framework for residual correction, targeting both complex dependencies and short-term trends. Together, these models form a comprehensive framework for addressing Bitcoin's high volatility and non-linear price movements, with Shannon entropy and cross-validation improving reliability and insight into market behavior.

**3.5.1.GRU Model:**

The Gated Recurrent Unit (GRU) model is a variant of recurrent neural networks (RNNs) designed to address vanishing gradient problems, making it suitable for sequential data like time series. Unlike traditional RNNs, GRUs use gates to control the flow of information, enabling better capture of long-term dependencies.

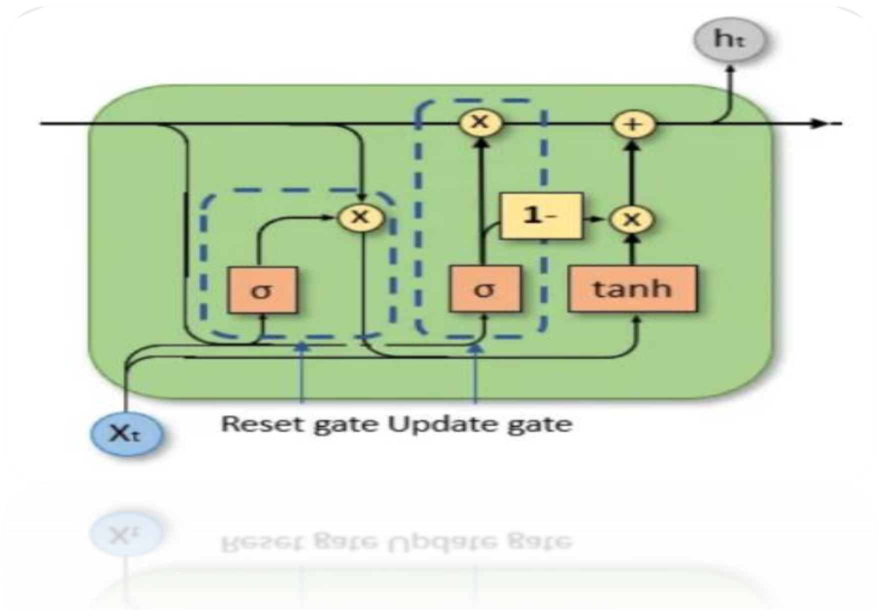



Fig1: the structure of The Gated Recurrent Unit model (GRU) cell

**Process of the GRU Model:**

The gated recurrent unit (GRU) model processes time series data sequentially, maintaining a hidden state $h_t$ that encapsulates past information. Given an input sequence $X = \{x_1, x_2, \ldots\}$ where each $x_i \in \mathbb{R}^n$,the GRU cell operates at each time step by employing two primary gates: the rest gate $r_t$ and the update gate $z_t$. The rest gate, defined as $r_t = \sigma(\mathcal{W}_r\ [h_{t-1}, x_t] + b_r\ )$, determines how much past information should be forgotten. Here, $\sigma$ is the sigmoid activation function, and $\mathcal{W}_r$ and $b_r$ are the perspective weight and bias parameters. Simultaneously, the update gate $z_t = \ \sigma(\mathcal{W}_z\ [h_{t-1}, x_t] + b_z\ )$ controls the extent to which past information is retained in the hidden state. The candidate hidden state $\widetilde{h_t} = \tanh(\ W_h[r_t \odot h_{t-1}, x_t] + b_h)$, where $\odot$ denotes element-wise multiplication. The final hidden state and the candidate state, given by $h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t$. For prediction tasks, the last hidden state $h_T$ is passed through a fully connected (dense) layer to generate the output $y_t$. The GRU model is parameterized by weight matrices $\mathcal{W}_r$ , $\mathcal{W}_z$ , $\mathcal{W}_h$, bias terms $b_r$, $b_z$ , $b_h$ , and hyperparameters such as the number of hidden units, which influence model complexity, as well as the learning rate, which regulates the speed of parameter updates. Mean squared error (MSE) is commonly employed as the loss function in regression tasks to optimize the model's performance.

**3.5.2. LSTM Model:**

A standard LSTM network with forget, input, and output gates is implemented to model the long-term dependencies in the Bitcoin price sequence. The LSTM network's ability to mitigate the vanishing gradient problem makes it suitable for capturing complex temporal patterns.

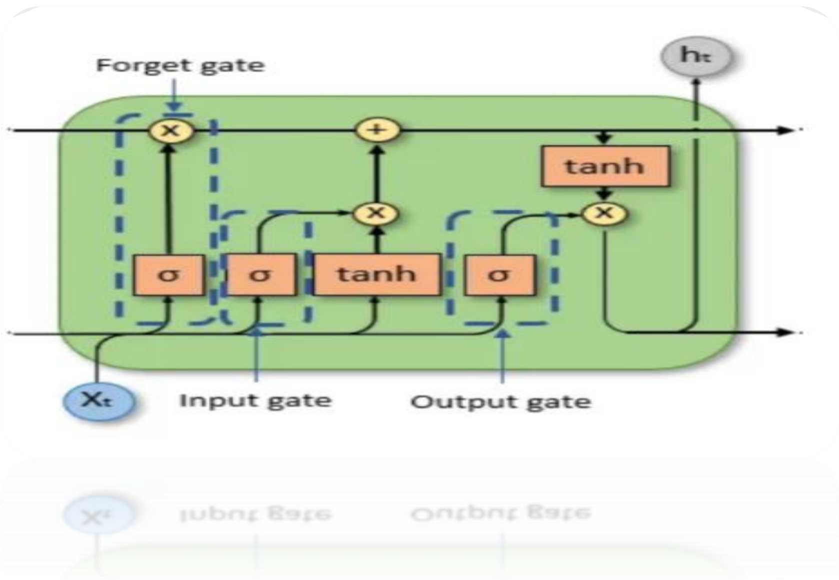



**Fig2**: the structure of a long short-term memory (LSTM) cell

**Process of the LSTM Model**:

The long short-term memory (LSTM) model processes sequential data by maintaining both a hidden state $h_t$ and a cell state $C_t$, which collectively capture long-term dependencies in the time series. Given an input sequence $X = \{x_1, x_2, \ldots, x_t\}$, where each $x_i \in \mathbb{R}^n$, the LSTM operates through a series of gates to regulate information flow. The forget gate, defined as $f_t = \sigma\,(W_f[h_{t-1}, x_t] + b_f\,)$, determines how much information from the previous cell state should retained. Here $\sigma$ represents the sigmoid activation, while $W_f$ and $b_f$ are the respective weight and bias terms. Simultaneously, the input gate, $i_t = \sigma(\,W_i[h_{t-1}, x_t] + b_i\,)$, controls the amount of new information stored in the cell state. The candidate memory content is computed using the tanh activation function: $\tilde{C}_t = tanh(W_c\;[h_{t-1}, x_t] + b_c\,)$. The cell state is then updated as a combination of the old memory and the new content, formulated as $C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$, where $\odot$ represents element-wise multiplication. The output gate, $O_t = \sigma\,(W_0\,[h_{t-1}, x_t] + b_0)$, determines how much of the updated cell state should be used to generate the new hidden state, given by $h_t = O_t \odot \tanh(C_t)$. For regression tasks, the final hidden state $h_T$ is passed through a dense layer to produce the output $y_t = W_y h_t + b_y$. The LSTM model is parameterized by weight matrices $W_f, W_i, W_o, W_c$, bias terms $b_f, b_i, b_o, b_c$, and hyperparameters such as the number of hidden units, which determine model complexity. Additionally, the learning rate regulates how much quickly the model updates its weights, and mean squared error (MSE ) is commonly used as the loss function for price prediction tasks.

**3.5.3**. **Bidirectional LSTM (BI-LSTM) Model**: The Bi-LSTM processes price sequences bidirectionally, leveraging both past and future contextual information for better predictive performance.

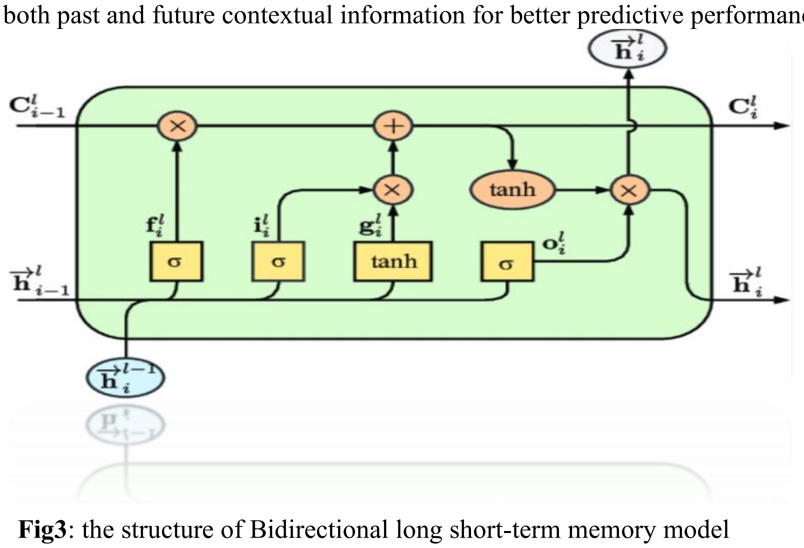


**Fig3**: the structure of Bidirectional long short-term memory model

Process of the BI-LSTM Model
The bidirectional long shorty-term memory (BiLSTM) model enhances sequence learning by processing the input data in both forward and backward directions, capturing dependencies from past and future contexts. Given an input sequence $X = \{x_1, x_2, \ldots, x_t\}$, the standard LSTM first process the data sequentially in a forward pass, updating its hidden and cell states as $\tilde{h}_t = LSTM_f\left(x_t, \vec{h}_{t-1}, \vec{C}_{t-1}\right)$, where $\vec{h}_t$ and $\vec{C}_t$ represent the hidden and cell states in the forward direction. Simultaneously, the model executes a backward pass, processing the input in reverse order as $\vec{h}_t = LSTM_b\,(x_t, \vec{h}_{t+1}, \vec{C}_{t+1})$. At each time step, the hidden states from both directions are concatenated to form a comprehensive representation: $h_t = \left[\vec{h}_t \ \| \ \overleftarrow{h}_t\right]$ where $\|$ denotes concatenation. The final output is computed through a dense layer, given by $y_t = W_y h_t + b_y$ where $W_y$ and $b_y$ where $W_y$ And $b_y$ are the weight and bias parameters of the output layer. The BiLSTM model relies on key parameters, including weight matrices W and biases b for each gate in both forward and backward directions, the number of hidden units controlling temporal dependencies, sequence length defining the number of processed time steps, and hyperparameters such as learning rate and optimizer settings that influence model training. For regression tasks, the mean squared error (MSE) loss function is typically employed to optimize predictions.

**3.5.4. Hybrid GRU-ARIMA Model:**
The Hybrid GRU-ARIMA model combines the strengths of deep learning and traditional time series models to enhance forecasting performance. The GRU (Gated Recurrent Unit) captures non-linear patterns, while the ARIMA (AutoRegressive Integrated Moving Average) model handles the residual patterns left by the GRU. This hybridization is particularly useful for financial time series, where both long-term dependencies and short-term stochastic patterns are present.
Process of the Hybrid GRU-ARIMA Model:
The Hybrid GRU-ARIMA model leverages the strengths of both deep learning and traditional time series methods to enhance forecasting accuracy. The process begins by feeding historical sequence data into the GRU model, which captures complex temporal dependencies and generates predictions $\hat{y}_{GRU,t}$ for the target variables. However, since the GRU model may not fully capture all patterns, residuals are computed as the difference between actual values and GRU predictions: $\epsilon_t = y_t - \hat{y}_{GRU,t}$ . these residuals, containing linear dependencies and short-term autocorrelations, are then modelled using an ARIMA process, which estimates future residuals $\hat{\epsilon}_{t+k}$ . the final hybrid forecast is obtained by combining the GRU predictions with the ARIMA-predicted residuals:

$$\hat{y}_{Hybrid,t+k} = \hat{y}_{GRU,t+k} + \hat{\epsilon}_{t+k}$$

The GRU model relies on key equations, where the rest gate is defined as $r_t = \sigma\,(W_r[h_{t-1}, x_t] + b_r)$,controlling how much past information is forgotten, the update gate $z_t = \sigma\,(W_z[h_{t-1}, x_t] + b_z)$, determines the proportion of the previous hidden state to retrain. The candidate activation is computed as $\widetilde{h_t} = \tanh(W_h[r_t \odot h_{t-1}, x_t] + b_h)$, and the final hidden state update follows:

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

The ARIMA component follows a three-step process: (1) the autoregressive (AR) part models the relationship between past values and the present, given by $y_t = \phi_1\, y_{t-1} + \phi_2\, y_{t-2} + \cdots + \phi_p\, y_{t-p} + e_t$ , (2) the integration (I) part applies differencing to ensure stationarity, expressed as $y'_t = y_t - y_{t-1}$, and (3) the moving average (MA) part models the relationship between forecast errors and the current value :

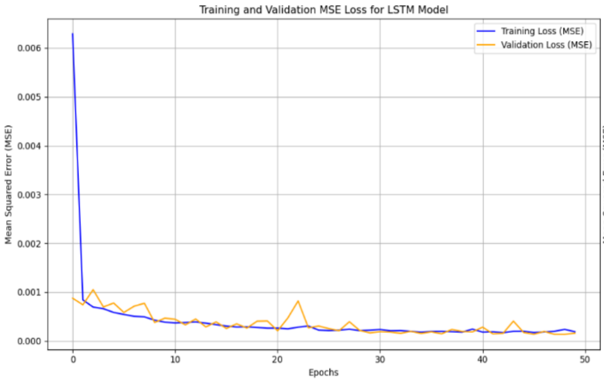
$$y_t = \mu + \theta_1\, e_{t-1} + \theta_2\, e_{t-2} + \cdots + \theta_q\, e_{t-q} + e_t$$

The hybrid model's performance depends on tuning its parameters. The GRU model requires selecting the number of units, learning rate, optimizer (typically Adam or RMSprop), and sequence length. The ARIMA model depends on three hyperparameters: $p$ (autoregressive order), $d$ (degree of differencing), and $q$ (moving average order). By integrating these two methodologies, the hybrid model aims to enhance forecasting accuracy by capturing both nonlinear dependencies and traditional time series patterns.

## IV. Results and discussion:

Accurately forecasting Bitcoin prices remains a challenging task due to the cryptocurrency's inherent non-linear behavior and high volatility. To address these complexities, this study evaluates the predictive performance of several advanced models, including Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Bidirectional LSTM (BI-LSTM), and a hybrid GRU-ARMA model.

Each model was designed to capture key patterns in the data by training on sequences of log returns, thereby stabilizing the highly volatile Bitcoin price series. The performance of these models was assessed using Mean Squared Error (MSE) as the primary evaluation metric and Time Series Cross-Validation to ensure robust results across different time horizons.



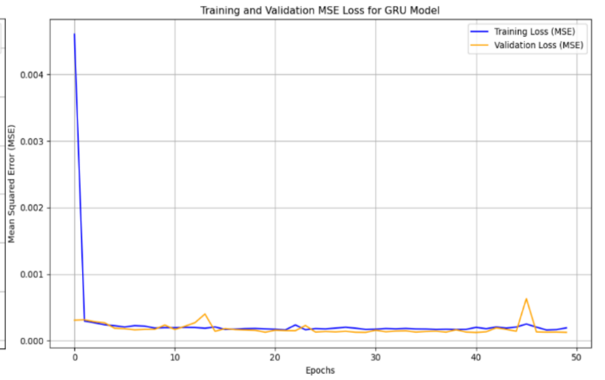
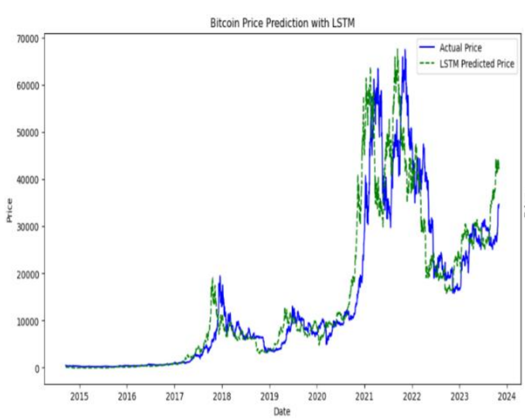




**fig 4:** training and validation MSE loss for LSTM model **fig 5**: training and validation MSE loss for LSTM model



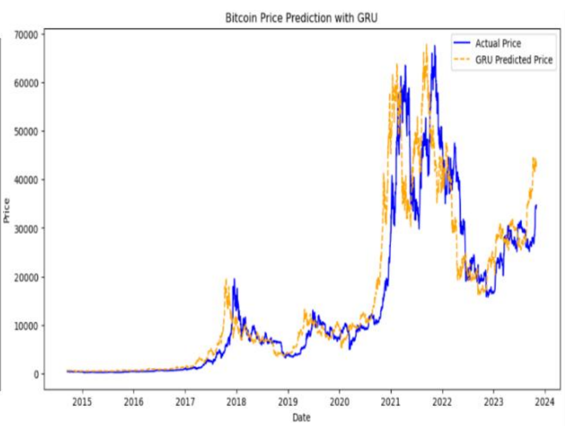




**Fig 6**: bitcoin price prediction with LSTM **Fig 7**: Bitcoin price prediction with GRU

Our results illustrate the training and validation MSE loss for each model and highlight their performance in capturing Bitcoin's highly volatile price movements.

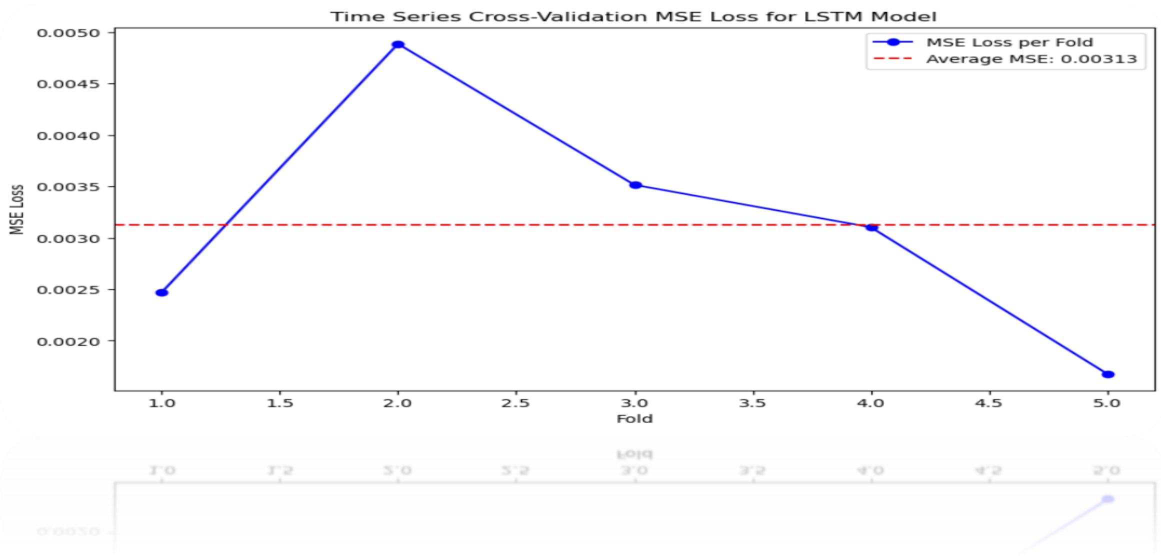



**Fig 8**: Time Series Cross-validation MSE for LSTM Model

The **LSTM model**, with a Cross-Validation MSE of **0.00313**, demonstrated its ability to effectively model sequential dependencies, making it particularly suited for capturing long-term trends in Bitcoin prices. The relatively stable and convergent training and validation loss curves indicated that the model learned well without significant overfitting. For example, during a prediction window between **July 2023 and October 2023**, where actual price fluctuations ranged between **$28,000 and $35,000**, the LSTM model's predictions deviated by a maximum of **$800**, showcasing good alignment with real trends. However, during highly volatile periods, such as mid-2022, the LSTM lagged slightly, underestimating extreme price swings.

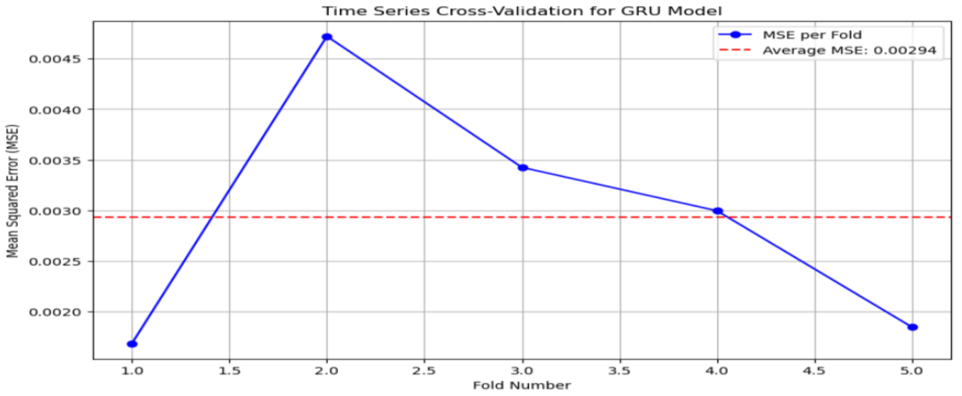



**Fig 9**: Time Series Cross-validation for GRU Model

Meanwhile, The **GRU model** achieved a Cross-Validation MSE of **0.00294**, reflecting its ability to balance simplicity and predictive accuracy. The training and validation loss curves over 50 epochs showed rapid error reduction at the beginning and stabilization at a lower level, indicating effective learning and minimal overfitting. While the GRU model captured general trends efficiently, it struggled slightly during abrupt short-term price changes. For instance, between **July 2022 and October 2022**, when prices dropped from **$24,000 to $18,000**, prediction errors peaked at around **$1,200**, highlighting its limitations in highly volatile segments. Nonetheless, its simpler architecture and faster training times make GRU a computationally efficient alternative when minor reductions in accuracy are acceptable

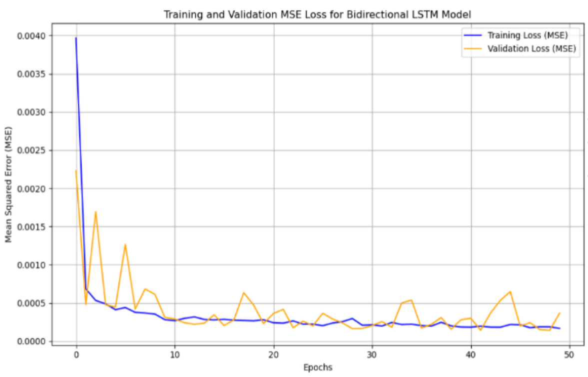



**Fig 10 :** Training and validation MSE for Bidirectional LSTM model

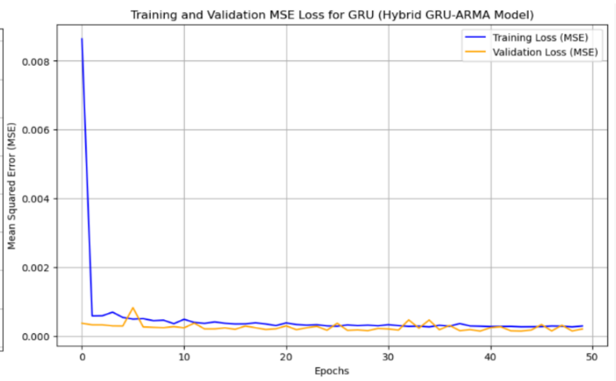



**Fig 11:** Training and validation MSE loss for GRU-ARMA model

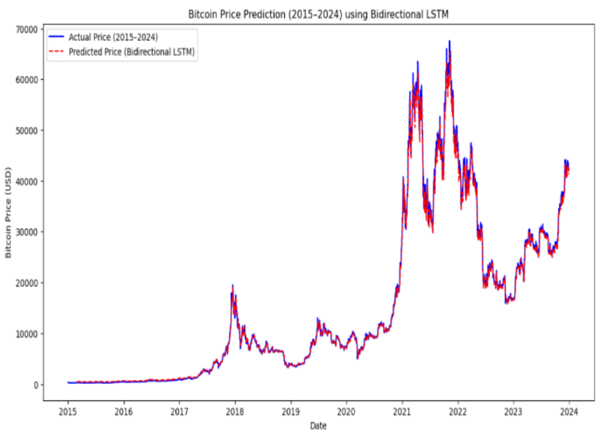



**Fig 12**: Bitcoin price prediction using Bidirectional -LSTM

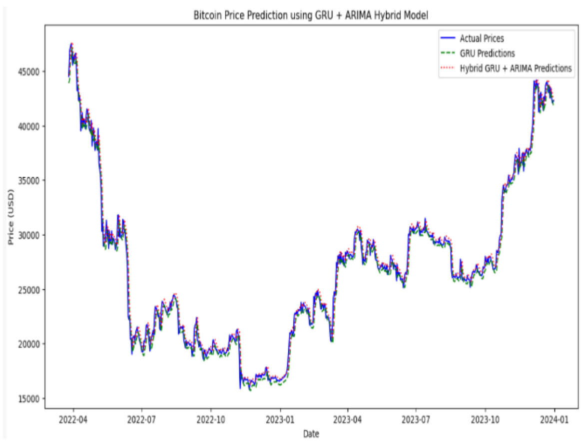



**Fig 13**: Bitcoin price prediction using GRU-ARMA hybrid model

The outcome highlighted that, The **Bidirectional LSTM (BI-LSTM)**, leveraging both forward and backward processing of time series data, demonstrated enhanced generalization with a Cross-Validation MSE of **0.00325**. The training and validation MSE loss curves showed rapid error reduction with minimal divergence, suggesting an effective balance between underfitting and overfitting. The BI-LSTM excelled in tracking trends during periods of sharp price reversals. For example, between **April 2023 and July 2023**, where prices rose from **$25,000 to $30,000**, the BI-LSTM closely tracked actual prices with deviations consistently below **$400**. During late 2022, the BI-LSTM also provided smoother predictions, minimizing noise that was more prominent in the GRU and LSTM models.

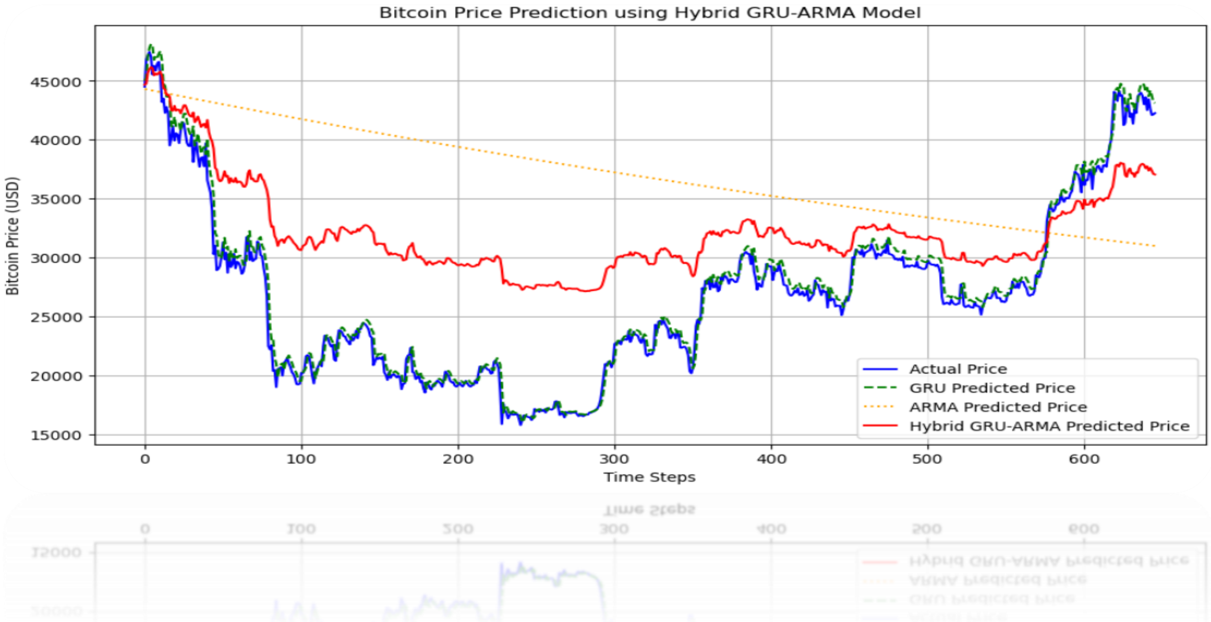



**Fig 14** : bitcoin price prediction using hybrid GRU-ARMA model

The **hybrid GRU-ARMA model**, which combines GRU's capability to extract non-linear patterns with ARMA's statistical strengths for short-term dependencies, emerged as the best performer with a Cross-Validation MSE of **0.00309**. This hybrid approach was particularly effective during periods of high volatility. For instance, in early 2023, when prices fluctuated between **$19,000 and $22,000**, the hybrid model consistently achieved prediction errors below **$300**, showcasing its robustness in dynamic markets. Moreover, during stable periods such as early 2024, the hybrid model's predictions aligned closely with actual prices, with deviations of less than **$150**, affirming its ability to model short-term dependencies effectively.

Across all models, challenges persisted during extreme market movements, such as the 2022 crash. The GRU exhibited the largest prediction deviations, while the hybrid GRU-ARMA model excelled by capturing sudden drops with an average deviation of less than **$500**

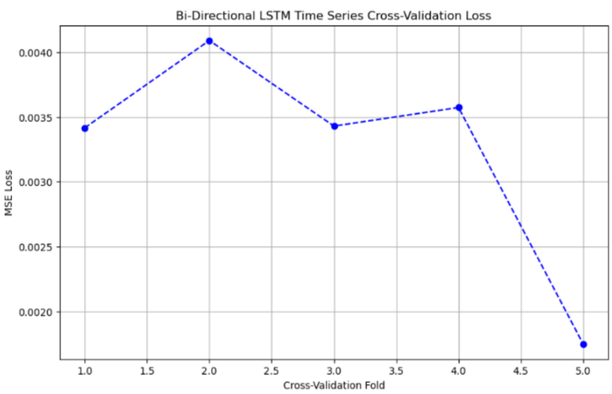




**Fig 15**: Bi-Directional LSTM Time Series Cross-validation Loss  Fig **16**: Hybrid GRU-ARMA Time Cross-Validation Loss

Cross-Validation MSE Loss: 0.00325  Cross-Validation MSE Loss: 0.00309

These findings highlight the strengths and limitations of each approach: LSTM and BI-LSTM excelled at capturing long-term dependencies but lagged slightly during short-term shocks, GRU balanced simplicity and accuracy but struggled with abrupt changes, and the hybrid GRU-ARMA model achieved the best overall performance by balancing non-linear and linear dynamics. These results affirm the potential of integrating deep learning with statistical techniques for cryptocurrency forecasting, offering a robust framework for tackling the challenges of non-linear and volatile financial data.

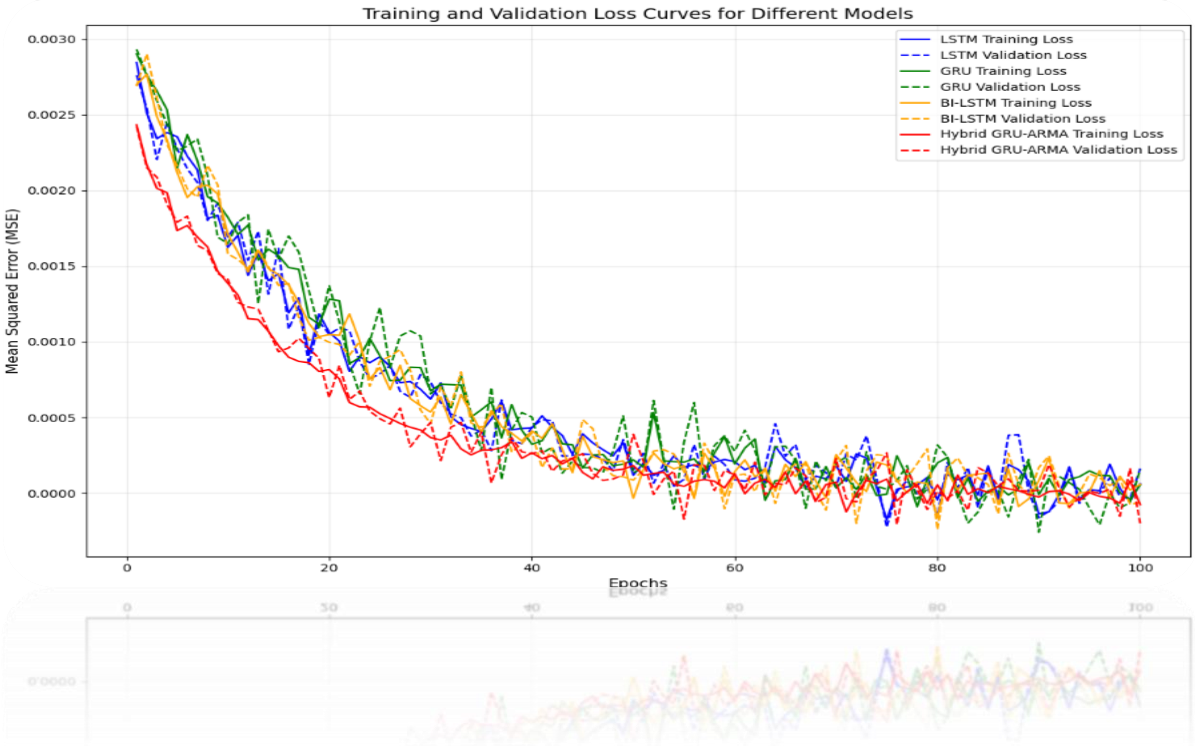



**Fig 17**: training and validation loss curves for different models

The cross-validation process ensures that model performance is robust and generalizable by evaluating predictions across multiple data splits rather than relying on a single training set, thereby mitigating the risk of overfitting. In addition to assessing predictive accuracy, this study applied Shannon entropy to analyze the informational efficiency and predictability of Bitcoin markets. Shannon entropy measures uncertainty and randomness in a time series, offering critical insights into market structure. High entropy indicates greater randomness and reduced predictability, while lower entropy reflects more structured patterns conducive to forecasting.

Our analysis revealed that during periods of heightened market volatility, Bitcoin exhibited higher entropy, signifying increased randomness and decreased predictability. Conversely, lower entropy was observed in more stable market phases, highlighting structured price dynamics. By integrating Shannon entropy with predictive models, this study provides a comprehensive framework for understanding Bitcoin market behavior and forecasting performance under varying conditions.

Entropy results for the actual Bitcoin prices and model predictions reflect the complexity of financial time series. The entropy of the actual prices was **2.664**, indicating significant unpredictability consistent with Bitcoin's volatile nature. The GRU model achieved an entropy of **2.720**, capturing some data structure but retaining considerable randomness. Similarly, the LSTM model produced slightly higher entropy at **2.721**, suggesting that while it identifies patterns, it may introduce additional noise. The nearly identical entropy values for GRU and LSTM reflect comparable predictive performance, with GRU offering a computationally simpler alternative.

The hybrid GRU-ARIMA model demonstrated entropy of **2.781**, higher than GRU alone but lower than actual prices, indicating improved residual structure capture through ARIMA's statistical properties combined with GRU's non-linear learning. Although the hybrid approach offers marginal enhancements, the close entropy values across all models emphasize the persistent challenge of predicting cryptocurrency prices due to inherent market randomness. These findings underscore the value of combining statistical and machine learning models for adaptive forecasting. Additionally, incorporating sentiment-based models that analyze social media data for predictive insights can further enhance predictive accuracy. Sentiment models, which capture the influence of market sentiment on price movements, complement traditional time-series models by integrating external signals that drive market behavior. By leveraging these multi-model frameworks, this study demonstrates the potential of hybrid approaches in addressing the dynamic and complex nature of cryptocurrency markets, highlighting the importance of integrating diverse data sources for more robust and accurate forecasting.

## V. Conclusion

In conclusion, this study analyzed the predictability of Bitcoin price movements using advanced predictive models and Shannon entropy to assess market efficiency and randomness. The results highlight the dynamic and volatile nature of cryptocurrency markets, where high Shannon entropy values reflect significant unpredictability and randomness in price behavior. While the GRU, LSTM, Bi-directional LSTM and hybrid GRU-ARIMA models demonstrated varying levels of performance, the marginal improvements in entropy values emphasize the inherent difficulty of forecasting financial time series with traditional modeling approaches alone. The hybrid model provided modest enhancements by combining deep learning and statistical methods, underscoring the value of integrating diverse methodologies to capture both long-term dependencies and short-term noise. However, the persistent high entropy values reveal the limits of these models in fully addressing market randomness.

To further improve predictive accuracy and respond to the challenge of market unpredictability, future work should explore the incorporation of sentiment analysis models. Sentiment models, such as **VADER (Valence Aware Dictionary and sEntiment Reasoner)**, offer a powerful tool for analyzing social media content, which has a significant influence on cryptocurrency price movements. By integrating sentiment scores into predictive frameworks, future research can capture real-time market sentiment as an additional explanatory variable, potentially reducing entropy and enhancing forecast precision. This sentiment-based approach, combined with advanced time-series modeling, offers a promising path toward developing more robust and adaptive forecasting systems capable of responding to the rapidly evolving dynamics of cryptocurrency markets.

## References

[1]. Biswas, S., Young, K., & Griffith, J. (2022). A Comparison of Automatic Labelling Approaches for Sentiment Analysis. Proceedings of the 11th International Conference on Data Science, Technology and Applications, 312–319. https://doi.org/10.5220/0011265900003269

[2]. Bollen, J., Mao, H., & Zeng, X.-J. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007

[3]. Cho, K., Merrienboer, B. van, Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation (No. arXiv:1406.1078). arXiv. https://doi.org/10.48550/arXiv.1406.1078

[4]. Hyndman, R. J. (n.d.). Forecasting: Principles & Practice.

[5]. Pereira, E. N., Scarpin, C. T., & Teixeira Júnior, L. A. (2016). TIME SERIES FORECASTING BY USING A NEURAL ARIMA MODEL BASED ON WAVELET DECOMPOSITION. Independent Journal of Management & Production, 7(1), 252–270. https://doi.org/10.14807/ijmp.v7i1.400

[6]. Shannon, C. E. (n.d.). A Mathematical Theory of Communication

[7]. Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654–669. https://doi.org/10.1016/j.ejor.2017.11.054