



RESEARCH ARTICLE

OPEN ACCESS

FORECASTING BITCOIN PRICE MOVEMENTS USING HISTORICAL DATA

 Ammar Ahmed Othman ^{*1}, Seddiq Hassan Al-Banna Ali ² and Mohammed Bakr Youssef ³
¹Ministry of Higher Education and Scientific Research - University of Mosul - College of Fine Arts. Iraq

^{2,3}Ministry of Higher Education and Scientific Research - University of Mosul - College of Computer Science and Mathematics. Iraq

¹<https://orcid.org/0009-0003-5736-9183>, ²<https://orcid.org/0009-0006-1178-3639>, ³<https://orcid.org/0009-0001-4646-0796>

 E-mail: ammar@uomosul.edu.iq, sedeeq.alkhazraji@uomosul.edu.iq, Mohammed.bkr@uomosul.edu.iq

ARTICLE INFO

Article History

Received on: March 05, 2026

Reviewed on: April 04, 2026

Accepted on: May 01, 2026

Published on: July 31, 2026

Keywords:
 Bitcoin Price Prediction,
 Cryptocurrency Forecasting,
 Long Short-Term Memory
 (LSTM),
 Technical Indicators,
 Time-Series Analysis,
 Cryptocurrency Volatility,
 Machine Learning.

ABSTRACT

In the digital currency, Bitcoin (BTC) is called the gold of the digital currency. It is possible to make some profits in trading of bitcoins, though this market is a very illiquid market and it is very challenging to determine the price of a bitcoin. The current work uses historical data and technical indicators to predict Bitcoin prices in a broad approach. BTC-USD price data were obtained using Yahoo Finance API and covered from 01/01/2015 till 07/01/2024. The concept of feature engineering was applied to improve the dataset by including vital financial characteristics, including Moving Averages, RSI, and Bollinger Bands for higher forecasting precision. The forward-looking model for the Bitcoin price was developed using machine learning and deep learning algorithms. The efficiency of the model was assessed with the help of Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). The overall values of Mean Absolute Error, Mean Squared Error, and Root Mean Squared Error were 0.0062, 8.39e-05 and 0.0092 respectively which suggest that the proposed model is accurate in forecasting the future prices.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

The market of the crypto currency is one of the biggest and influential markets in the world of the financial market, where trader from all over the world invest in the digital currency. Trade in this market serves multiple objectives from acting as the market that provides the required liquidity for currencies as well as enhancing the provision of the export currency for the export of goods and services to other countries to offer individuals and institutions the ability to make attractive returns from speculative trades in the global foreign exchange market. The crypto market is a large and highly liquid market in terms of trading volume, which means that investors prefer such a market when they are acting to earn quick profits [1]. This current market also brings the chance to diversify and get returns due to global market fluctuations. Taking into account the processes associated with development of technology and growth of investment conscious, trading in the currency market is not only the profession of the experts but also the area freely accessible to everyone willing to invest and manage risks more effectively [2]. The financial technology has been experiencing drastic changes especially from the beginning of the new century, and has greatly influenced the world's financial structure.

From these contributors are fairly introduced the development of the digital currency which is an additional form of the money and financial transactions. New technology in money adoption, known as digital currencies based on encryption and decentralization have received other unusual attention in both field of finance and academic mediation [3]. Digital currencies are one of the financial assets that utilize encryption and common transaction denominations to generate new ones. They are not like conventional money as they are not controlled by a central banking network, but they are documented through a network of nodes. These are currencies based on block chain technology, a transparent and secure database of currencies, hence efficient and free from fraud [4]. Bitcoin, launched in 2008, is the first digital currency that introduced the concept of blockchain as a decentralized means of recording transactions. Despite its humble beginnings, it has proven its ability to greatly influence global financial markets, leading to the emergence of many other digital currencies such as Ethereum, Ripple, and Litecoin [5].

II. DIGITAL CURRENCIES

Another aspect of the digital currencies is that the control of the type or kind of money is not centralized there is no central authority that dictates the type of money to be issued in circulation or how it will be managed. Rather, digital currencies operate through a group of people to assure the transaction's safety and increase transparency. Further, these currencies are characterized by rapidity of the financial operations and the highest level of protection due to the use of complex encryption algorithms that eliminate the need for the use of conventional financial mediators across the border [6]. Digital currencies can be primarily categorized into three main types. Cryptocurrencies: Examples are Bitcoin and Ethereum, through which value is transferred and stored. For example, Bitcoin is the first virtual currency designed to act as an alternative currency. , whereas Ethereum is not only a digital currency but also a platform for developing decentralized applications and smart contracts [7].

1. Tokens: These are based on technologies such as Ethereum and are used for various uses starting from funding projects, offering specific services or products, up to emulating traditional shares. Such tokens are Tether (USDT) and Chainlink (LINK) [8].
2. Stablecoins: meant to be less volatile, they are most often anchored on a conventional currency, for instance, the US dollar. Some of the centralized stablecoins include Tether, which is also known as USDT, and USD Coin also known as USDC. These currencies is expected to reduce the volatile nature, which is common with most of the other digital currencies for use in trading and other economic activities [9].

III. DIFFICULTY IN PREDICTING CRYPTOCURRENCY PRICES

Depending on the probability of cryptocurrency price prediction is still one of the biggest prospects that affects the investors and analysts in this highly volatile market. Currency prices are volatile and vary significantly and unpredictably, thus, predicting cryptocurrency is not an easy task which depends on the understanding of this market. Among the influential factors are the following [10]:

1. Price volatility: In general, the prices of cryptocurrencies fluctuate, which is much higher than the price fluctuations in traditional financial markets. Such a situation is explained by several reasons: Fluctuations in supply and demand, large transactions, and the emotional state of a trader. These fluctuations make it challenging to adopt standardized models to project the cryptocurrency prices faithfully [11].
2. Lack of historical data: Generally, historical information related to cryptocurrencies is limited and of a shorter time span as compared to traditional markets. This lack of long-term data hampers the analysts in the construction of a forecasting system based on quantitative analysis over a long span of time; the long-term trends cannot be predicted [10].
3. Psychological and market factors: The prices of various cryptocurrencies are greatly influenced by psychological factors. It can be also defined that news, rumors and market expectations can produce fluctuations in the price. There is empirical evidence that fast trading based on news affects the prices of stocks in a big way [12].
4. Technological developments and updates: It is clear that the further technological evolution and updates affect the projects related to cryptocurrencies. Alterations in the protocols, the security measures, or the nature of the technology in any way can result in shifts in the perceived market value of cryptocurrencies [1].
5. Regulatory considerations: Cryptocurrencies are still a new form of currency with new local and international regulatory frameworks and laws being passed often. Cryptocurrencies are unpredictable since new regulatory actions or changes in government benchmarks can impact the token's price [13].
6. Fraudulent activities and security risks: The cryptocurrency market is under high risks of fraud and security threats. These activities lead to the emergence of a situation that can dramatically change the prices of cryptocurrencies and, therefore, makes it more complicated to foresee them with a high level of certainty [14].
7. Inter-currency interaction: Often particular cryptocurrency prices like for Bitcoin may affect the prices of other cryptocurrencies. Transactions between various digital currencies create fluctuations which can be unpredictable [15].

IV. LITERATURE REVIEW

1. Deep Learning Models for Bitcoin Prediction Using Hybrid: This paper utilized more than 50,000 hourly Bitcoin price data from the period of May 2018 to January 2024 The study used complex models such as gradient descent, attention mechanisms, LSTM, and ANNs. As it was developed it complement LSTM with attention mechanisms and integrated gradient-specific optimization to further improve the predictive power of the model. The results of the work show that these hybrid models, and more specifically with the gradient-specific optimization, enhance the Bitcoin price prediction by utilizing the advantages of each approach discussed above [16].
2. Bitcoin price prediction using Deep Learning Algorithm: The goal of this paper is to forecast the Bitcoin USD price using a Bayesian Recurrent Neural Network and Long Short Term Memory Network. Using LSTM, the model had an accuracy of 52% and 8% RMSE which is better than conventional ARIMA models. Furthermore, a GRU model was applied, and utilizing deep learning models, the best results obtained with an enhancement of 94,7% in the GPU training time [17].
3. Bitcoin Price Prediction Based on Deep Learning Methods: This paper aimed at enhancing the performance of the deep learning techniques such as min-max normalization and the Adam optimizer in enhancing the accuracy of stock market prediction. When employing 56,832 observations of Bitcoin's price, the authors determined that MLP was ineffective because it lacked memory. On the other hand, LSTM and GRU models with past memory, gave better predictions [18].

V. ARTIFICIAL INTELLIGENCE AND PREDICTION

AI has achieved a lot in the area of prediction including in the areas of markets, and health, weather among other areas not mentioned. Some of the artificial intelligence methods and techniques found in prediction include the following [19].

V.1 MACHINE LEARNING (ML)

Machine learning ML is a branch in artificial intelligence that involves the use of algorithms to make predictions or decisions from data without being programmed. ML can be categorized into the following two main types [20]:

Supervised Learning: Supervised learning involves the use of training sets that is, the inputs are associated with their respective outputs or labels. During the training phase, the model acquires the knowledge of how to convert inputs into outputs, while during the testing phase, it is assessed. This method is widely applied to problems where historical labeled data is available. Description of Supervised learning algorithms is shown in the Table (1) [21].

Table 1: Description of Supervised Learning Algorithms.

Algorithm	Description	Applications
Linear Regression	Establishes a linear relationship between input features and a continuous target variable.	Predicting continuous variables, e.g., house prices, stock prices.
Support Vector Machines (SVM)	Finds a hyperplane that best separates data points into distinct classes while maximizing the margin.	Image classification, spam detection, and medical diagnosis.
Decision Trees	Splits data into subsets based on feature values, creating a tree-like structure for prediction.	Fraud detection, loan approval, and risk assessment.
Random Forests	Combines multiple decision trees to improve predictive performance and reduce overfitting.	Credit scoring, stock market prediction, and customer segmentation.

Source: [22].

The most common types of machine learning methods are the supervised learning algorithms due to their use of labelled datasets which allow the models to be trained to directly map the input variables to the known outputs. The algorithms used in predictive analytics and classification include Linear Regression, Support Vector Machines, Decision Trees, and Random Forests. The quality and quantity of labeled training data are also highly relied upon to determine their effectiveness. Such techniques are especially useful in the areas, where one has the data about the past and is capable of learning patterns in order to make predictions in the future. But this can be disadvantageous with supervised learning models as they can overfit or undergeneralize when the training data is smaller or extremely skewed.

Unsupervised Learning: In unsupervised learning the model operates using data that do not contain any labeling and is aimed at finding patterns, correlations, or structures of data without prior information about the results. It is popular for use in the initial stages of data analysis and especially for identifying trends. Table (2) presents details of Unsupervised Learning algorithms [23].

Table 2: Description of Unsupervised Learning Algorithms.

Technique	Description	Applications
Clustering Algorithms	Group similar data points into clusters, e.g., K-Means and Hierarchical Clustering.	Customer segmentation, market basket analysis, and image compression.
Dimensionality Reduction	Reduces the number of features while retaining the data's essential information, e.g., PCA.	Visualization of high-dimensional data, noise reduction, and feature selection.

Source: [24].

Unsupervised learning is an important part of the exploratory data analysis of the data when there are no labeled datasets. Clustering and dimensionality reduction are techniques by which large datasets are identified as hiding structures, correlations, and patterns in the models. Clustering, such as, is used to cluster similar data instances and this is applicable in segmentation and in anomaly detection. Principal Component Analysis (PCA) is a method of dimensionality reduction that reduces large and complicated datasets while retaining valuable data. Such methods are especially useful with high-dimensional data. Irrespective of this, it is difficult to interpret the results of unsupervised learning models because predefined labels do not exist to assess the performance of models.

V.2 DEEP LEARNING (DL)

Neural Networks with more than one layer are used in DL which is a part of machine learning. It is most useful in case of large database and problems which are nonlinear in nature. Table (3) define Deep Learning (DL) algorithms [25].

Table 3: Description of Deep Learning Algorithms.

Model Type	Description	Applications
Convolutional Neural Networks (CNNs)	Extract hierarchical features from structured data such as images and videos.	Image recognition, object detection, and video analysis.
Recurrent Neural Networks (RNNs)	Process sequential data, maintaining temporal relationships using feedback loops.	Speech recognition, stock price forecasting, and language modeling.
Long Short-Term Memory (LSTM)	Overcomes RNN limitations by capturing long-term dependencies with memory cells.	Time-series forecasting, text generation, and financial predictions.

Source: [26]

The deep learning algorithms have contributed a lot to the functionality of machine learning so that systems with deep learning can automatically extract hierarchical features representations of large datasets. Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) networks are architectures that are popular in complex tasks that involve images, videos and sequential data. CNNs are especially useful at analyzing spatial data, whereas RNN and LSTM models are oriented at the recognition of temporal relationships in the sequence of information. Such models have been incredibly successful in computer vision, speech recognition and natural language processing. Nonetheless, deep learning models are usually run on big datasets, substantial amounts of computing power, and delicate hyperparameter adjustment.

V.3 REINFORCEMENT LEARNING (RL)

This kind of machine learning is a type of learning from experience, which involves an agent and its environment in the process of accomplishing a certain task. The model just has reward for right action and punishment for wrong action and it makes corrections for its strategy and actions. Table (4) delineate the Reinforcement Learning (RL) algorithms [25].

Table 4: Description of Reinforcement Learning Algorithms.

Use Case	Description	Examples
Trading Strategies	Optimizes decision-making in buying and selling stocks based on market conditions.	Algorithmic trading, portfolio optimization.
Robotics	Improves robot autonomy and efficiency in completing complex tasks.	Autonomous navigation, object manipulation.
Game Playing	Trains models to perform well in competitive environments by learning from outcomes.	Chess, Go, video games, and reinforcement-based simulations.

Source: [27]

The difference of reinforcement learning in comparison with other machine learning paradigms lies in the fact that it aims at understanding how to act in the best way, interacting with the environment. Under this scheme, the agent will be rewarded or punished depending on its actions and will keep on improving its decision-making scheme by trial and error. The reinforcement learning has proved to be useful in robotics, automated trading and intelligent gaming systems. It has the strength of being able to acquire adaptive strategies in changing and uncertain conditions. However, reinforcement learning models are computationally costly and in many cases, they take a long period of time to train, especially when the environment is complex in terms of state and action space.

V.4 NATURAL LANGUAGE PROCESSING (NLP)

Natural Language Processing (NLP) is about the interface between computers and languages used by humans. It utilizes models of machine learning and deep learning to work with textual data. Table (5) provides description of (NLP) techniques [28].

Table 5: Natural Language Processing (NLP) techniques.

Model	Description	Applications
Transformers	Advanced models like GPT and BERT designed for natural language understanding.	Text summarization, translation, and question answering.
BERT (Bidirectional Encoder Representations from Transformers)	Captures word context bidirectionally, making it ideal for sentiment analysis.	Sentiment analysis, trend prediction, and contextual text understanding.

Source: [28].

Natural Language Processing (NLP) aims at making computers comprehend, interpret and produce human language. Deep learning architectures that utilize transformer-based models (e.g., BERT and GPT) are highly used in modern NLP systems. These models acquire contextual dependencies between words and are highly effective in the performance of language understanding tasks. Applications in the field of machine translation, sentiment analysis, question answering and text summarization are some of the common applications of NLP techniques. The use of transformer models has brought a revolution in the field due to the enhanced awareness of context and scalability. Nevertheless, NLP models can be resource-intensive (in terms of the required training datasets and computation), and, still, they can be confused by ambiguity and specifics of domain language.

V.5 ENSEMBLE METHODS

Ensemble methods enhance model performance by combining multiple learners to reduce errors and increase accuracy. Table (6) shows Ensemble Methods [29].

Table 6: Ensemble Methods.

Method	Description	Applications
Boosting	Sequentially improves weak learners, e.g., AdaBoost and Gradient Boosting Machines (GBM).	Fraud detection, loan default prediction.
Bagging	Combines predictions from multiple models using resampling, e.g., Random Forests.	Customer segmentation, product recommendation.
Stacking	Trains a meta-model on predictions from base models for better performance.	Financial forecasting, ensemble predictive analytics.

Source: [29].

Ensemble learning approaches are designed to enhance predictive accuracy through the integration of a number of machine learning models into a strong framework. Boosting, bagging, and stacking are methods that can be used to minimize the modeling errors and enhance the stability of the models. The iterative boosting techniques improve weak learners and the bagging techniques minimize variance using resampling techniques. Stacking also enhances performance based on a meta-model of combining predictions made by base learners. They are common in complicated predictive problems where one model is not likely to store all the patterns in the data. Although not the least effective, ensemble approaches can become more expensive to compute and demand selection and parameter optimization of models.

V.6 TIME SERIES ANALYSIS

Time series analysis is used for analyzing and predicting data that evolves over time. It is particularly relevant in fields like finance, weather, and inventory management. Table (7) shows Time Series Analysis [30].

Table 7: Time Series Analysis.

Technique	Description	Applications
ARIMA (Auto Regressive Integrated Moving Average)	Combines autoregressive and moving average components to model and forecast time-series data.	Forecasting stock prices, sales, and demand trends.
Prophet	A user-friendly forecasting tool designed for capturing seasonality and trend components.	Business forecasting, revenue prediction, and anomaly detection.

Source: [30].

The time series analysis is dedicated to modelling and predicting time varying data, which is critical in the financial field, energy use, and weather forecasting. Logistic regression methods, including ARIMA and Prophet, have been popular in capturing trend, seasonal and temporal dependencies in a sequence of data. ARIMA models are especially useful in statistical forecasting whereas Prophet offers a more generic model when dealing with seasonality and irregular trends. The practices help organizations to make sound decisions when future predictions are made. Nevertheless, proper prediction needs very good historical data and the forecasting models might fail in instances where there occurs an abrupt shift in the structure of the system or some other unforeseen anomalies.

V.7 BAYESIAN NETWORKS

Bayesian networks are probabilistic graphical models that represent a set of variables and their conditional dependencies. They are particularly useful in scenarios where uncertainty must be accounted for in decision-making. Table (8) shows Bayesian Networks [31].

Table 8: Bayesian Networks.

Applications	Description
Risk Assessment	Helps model and predict potential outcomes and risks.
Medical Diagnosis	Used to identify diseases based on symptoms and test results.

Source: [31].

Bayesian networks are a probabilistic model that can be used to describe complex systems that have uncertain dependencies among variables. These graphical models display the conditional dependencies in the form of directed acyclic graphs which enable researchers to learn the cause-and-effect relationships as well as do probabilistic inferences. Applications of Bayesian networks in decision support applications, medical diagnosis applications, and risk assessment applications are common. They are especially effective in uncertain conditions because they can apply the previous knowledge and update the probabilities with regard to the new information. Nevertheless, it may be computationally difficult to build the correct Bayesian network structures in the presence of many variables and complicated dependency functions.

V.8 GENETIC ALGORITHMS

Genetic algorithms are optimization techniques inspired by natural selection. They iteratively improve solutions by mimicking biological evolution. Table (9) shows Genetic Algorithms [32].

Table 9: Genetic Algorithms.

Stage	Description	Applications
Selection	Chooses the best candidates based on fitness.	Portfolio optimization, route planning, engineering design.
Crossover	Combines parent solutions to generate offspring.	Product development, genetic programming.
Mutation	Introduces random changes to maintain diversity.	Complex problem-solving, evolving trading strategies.

Source: [32].

Genetic algorithms are evolutionary algorithms that are based on the concept of natural selection and biological evolution. The processes used in searching optimal or near-optimal solutions in them include selection, crossover and mutation. These algorithms have been used especially in solving complex optimization problems, which would not be easily solved by traditional methods because of large search space or nonlinear constraints.

Genetic algorithms are used extensively in the design of engineering, financial optimization, scheduling, and artificial intelligence. Although genetic algorithms are flexible, they necessarily need to be tuned by hand, and can have a slow convergence rate, particularly if the search space is very large or the fitness function is complicated.

VI. METHODOLOGY AND IMPLEMENTATION

Figure (1) shows a full workflow of model creation and assessment of a Long Short-Term Memory (LSTM) model to make predictions or analysis using data. It is carried out through a series of steps of a structured machine learning pipeline that starts with the acquisition of raw data and moves through a series of steps such as preprocessing, feature engineering, model training, and performance evaluation. Each of the steps allows properly preparing the dataset and making sure that the resulting model will succeed in learning patterns and making reliable predictions. The first step is Data Collection which is the process of getting raw data out of the sources that are relevant. This information can be in databases, sensors, open repositories, or real time streams based on the field of application. The data can also have anomalies, missing values or other irrelevant characteristics at this point, and can therefore be further processed before modelling is done. After collecting the data, the next step of the work is Data Exploration and Visualization. This step is aimed at getting to know the form and properties of the data set gathered. Patterns, anomalies and interrelations among variables are detected using techniques like statistical summaries, correlation analysis and visual techniques like charts or plots. Data exploration is quite necessary since it assists the researchers to learn about the data set and identify the most relevant approach of preprocessing and modeling. The second step is Data Preprocessing which involves cleaning and readying of raw data to be used in machine learning. Preprocessing tasks normally involve processing any missing data, eliminating duplicates, reconciliation of inconsistencies and in case of need, the conversion of categorical data into numerical form. Good preprocessing enhances the quality of data and gives the model consistent and meaningful data.:

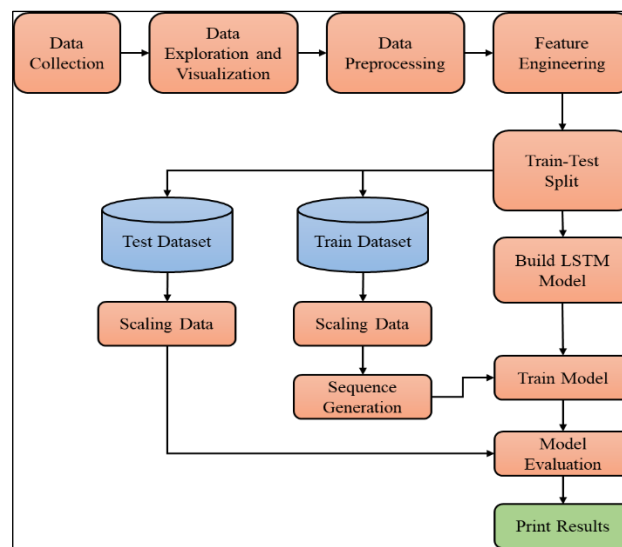


Figure 1: Methodology Block Diagram.

Source: Authors, (2026).

Following the preprocessing, Feature Engineering is used. During this phase, the appropriate features are chosen or developed to increase the learning capacities of the model on the data. The feature engineering process can be in the form of deriving new features, modifying the existing features, or doing away with redundant features. Good feature engineering could without any doubt enhance predictive power of machine learning models. After the features are ready the Train-Test Split process is used to divide the dataset. The data is divided into two subsets, the training data, which is used to train the model, and the testing data, which is used to assess the performance of the model on unseen data. Such partition can be used to measure the generalization capabilities of the model to new data and avoid overfitting. The training and testing data sets are subjected to Scaling Data which normalizes or standardizes the values of feature values to a similar range. Scaling is especially relevant to neural network models such as LSTM as it stabilized the process of training and speeded up the convergence. Once the training dataset has been scaled up, it is followed by Sequence Generation. Given that LSTM models expect that data is sequential or time-dependent, it must be converted into a form of sequences that are time-dependent. This stage will arrange the input data into temporal chains through which the LSTM network can acquire temporal dependencies and patterns with time. Then, Build LSTM Model phase specifies the neural network structure. The model normally has LSTM layers which form temporal relationship then more layers like dense layers to make predictions. This stage is used to configure model parameters, activation functions and optimization algorithms. After the construction of the model, the workflow goes to Train Model where the LSTM network is trained to learn patterns on the training data by the use of optimization. The model adapts its internal weights during training to reduce the error of prediction. After training is completed, Model Evaluation is performed using the testing dataset. Evaluation metrics such as accuracy, mean squared error, or other performance indicators are calculated to measure the model's effectiveness.

VI.1 DATA COLLECTION

The data collection phase focuses on gathering historical Bitcoin price data as the foundation for analysis and prediction. The Yahoo Finance API was used to retrieve daily data from January 1, 2015, to June 1, 2024. The dataset includes the following key columns.

Table 10: The dataset includes the following key columns.

Column	Description
Date	The timestamp for each entry, used to organize the data chronologically.
Open	The opening price of Bitcoin for the day.
Low	The lowest price recorded during the day.
High	The highest price recorded during the day.
Close	The closing price, reflecting the market's sentiment at the end of the day.
Adj Close	Adjusted closing price, accounting for stock splits or other adjustments.
Volume	The total number of Bitcoin units traded during the day, representing activity.

Source: Authors, (2026).

Table 10 offers all the basic variables needed to examine the price movements of Bitcoin and construct predictive models. Each column records a specific phenomenon of the market behavior such as price dynamics and trading activity. The open, high, low and close prices explain the daily price changes, which are fundamental to comprehend the volatility and trending. Adjusted close value is used to continue consistency in financial analysis in case of some adjustments. In the meantime, the trading volume indicates the market participation and liquidity. The combined variables will be a full depiction of the market conditions and act as essential inputs to further analysis, feature engineering, and model training phases.

VI.2 DATA EXPLORATION AND VISUALIZATION

To understand the data's underlying characteristics and trends, Exploratory Data Analysis (EDA) was conducted. Visualization tools were used to create meaningful insights:

Table 11: Data Exploration and Visualization.

Visualization	Purpose
Close Price Plot	Tracks the historical closing prices of Bitcoin to identify trends and price cycles.
Histogram and Density	Analyzes the distribution of Bitcoin closing prices to understand common price ranges.
Candlestick Plot	Combines open, high, low, and close prices, providing detailed insights into price movements.

Source: Authors, (2026).

The illustrations to be discussed in Table 11 are essential in perceiving the general dynamics of Bitcoin prices throughout the years. The close price plot enables the researcher to see long-term trends, cyclicity and abrupt changes which can point at volatility in the market or other external economic factors. The histogram and the density plots could be used to examine the statistical distribution of closing prices and reveal the common price ranges and determine potential skewness or outliers. Moreover, the candlestick chart provides a detailed display of price fluctuations of each trading day, through the combination of open, high, low and close values. The methods of exploration allow interpreting the dataset more effectively and allow making informed decisions in preprocessing and modeling phases.

VI.3 DATA PREPROCESSING

Preprocessing ensures the dataset is clean and ready for model training. Several steps were undertaken:

Table 12: Data Preprocessing.

Step	Description
Handling Missing Values	Missing data was filled using forward filling, and remaining null values were removed.
Scaling Data	Normalized using MinMaxScaler to scale all features between 0 and 1. Separate scalers were used to prevent data leakage.
Sequence Generation	Created time-series sequences with a 30-day window size, ensuring sufficient historical context for predictions.

Source: Authors, (2026).

As shown by Table 12, the following preprocessing tasks are necessary to prepare the dataset to be easy to train in machine learning. The process of working with missing values will make sure that incomplete records will not harm the learning process of the model. Forward filling is also useful in preserving the continuity of a time-series data with minimal loss of information. Scaling of the data set with MinMaxScaler standardizes the features in a similar range, thus enhancing the stability of the training of the neural network and faster convergence. Also, time-series sequences generate time-series data, which are structured as input to sequential models like LSTM. These preprocessing steps are very important in improving the data quality and it also makes sure that the model is presented with sound and well-organized inputs.

VI.4 FEATURE ENGINEERING

To enhance the dataset's predictive power, key financial indicators were added. These indicators capture different aspects of market dynamics:

Table 13: Feature Engineering.

Indicator	Description
Simple Moving Average (SMA)	Calculates the average of closing prices over specified periods (e.g., 20 and 50 days) to smooth trends.
Exponential Moving Average (EMA)	Similar to SMA but gives more weight to recent prices, making it more responsive to changes.
MACD (Moving Average Convergence Divergence)	Tracks momentum using the difference between short-term and long-term EMAs.
RSI (Relative Strength Index)	Identifies overbought or oversold conditions, ranging from 0–100.
Volatility Measures	Includes ATR (Average True Range) and standard deviation to assess price volatility.
Bollinger Bands	Highlights overbought/oversold conditions using upper and lower price bands based on standard deviations.

Source: Authors, (2026).

Introducing domain knowledge into the dataset through feature engineering is essential in the quality of prediction of machine learning models. The financial indicators in Table 13 represent various market behavior components, such as market trends, market momentum, and market volatility. SMA and EMA are moving averages that are used to smooth out fluctuations and identify long-run trends. Such indicators as MACD and RSI give us the information about the momentum in the markets and possible overbought or oversold. Volatility indicators like ATR and Bollinger Bands also assist in determining times of instability of a market. The combination of these indicators provides the model with a more in-depth contextual information increasing its capability to model complex price dynamics.

VI.5 TRAIN-TEST SPLIT

The dataset, spanning 3440 days, was divided into training, validation, and test sets to ensure robust model evaluation:

Table 14: Train-Test Split.

Subset	Duration	Purpose
Training Set	January 1, 2015 – April 1, 2023	Used to train the model on historical data (3012 days).
Validation Set	April 2, 2023 – October 2, 2023	Fine-tuning hyperparameters and evaluating generalization (214 days).
Test Set	October 3, 2023 – June 1, 2024	Assessing model performance on unseen data (214 days).

Source: Authors, (2026).

The division plan of the dataset used in Table 14 provides a dependable assessment of the predictive model. The division of the dataset into training, validation and testing subset is important as it gives the model the opportunity to learn on historic data without bias in evaluation on unprecedented data. Most of the data is found in the training set, which is employed to acquire patterns and relationships among variables. The validation set aids in the hyperparameter tuning and also helps in avoidance of overfitting by testing the model in the course of development. Lastly the test set is used to give a final evaluation of the predictive power of the model on new data. This organized separation increases the accuracy and validity of the experimental findings.

VI.6 SCALING DATA

To normalize data for model training and improve performance, MinMaxScaler was applied:

Table 15:Scaling Data.

Scaling Step	Description
Feature Scaling	All features were scaled between 0 and 1 to prevent larger values from dominating the training process.
Data Leakage Prevention	Separate scalers were applied to training and test data to avoid unintentional leakage.

Source: Authors, (2026).

Preprocessing Data Scaling Data scaling is an important preprocessing step when training neural network models, especially gradient-based optimization. The use of MinMaxScaler will make sure that all the input features are put within a similar range between 0 and 1. This normalization avoids larger numeric ranges of features controlling the training process, and enhances numerical stability in model optimization. Moreover, using distinct scalers in training and testing data sets can be used to avoid data leakage, which otherwise will result in overly optimistic estimates of performance. Scaling ensures that faster convergence is achieved during the training process and enhances the overall performance and stability of the LSTM model by standardizing the values of features.

VI.7 SEQUENCE GENERATION

For time-series modeling, sequences of fixed length (30 days) were created to capture historical context. The sequence generation step involved:

Table 16: Time-Series.

Window Size	Purpose
30 Days	Provides sufficient historical context without introducing unnecessary computational overhead.
Longer Window Sizes	Tested but showed diminishing returns in prediction accuracy.
Shorter Window Sizes	Reduced model performance due to inadequate historical information.

Source: Authors, (2026).

Sequence generation is a fundamental step in preparing data for time-series models such as LSTM networks. Instead of treating each observation independently, the dataset is transformed into sequential windows that capture historical dependencies. As shown in Table 16, a 30-day window size was selected to provide sufficient historical context while maintaining computational efficiency. This approach enables the model to learn temporal patterns and relationships that influence future Bitcoin price movements. Experiments with longer windows showed limited improvements in prediction accuracy, while shorter windows lacked adequate historical information. Therefore, the chosen sequence length provides an effective balance between model complexity and predictive performance.

VI.8 MODEL BUILDING: LSTM ARCHITECTURE

The LSTM model was designed to capture temporal dependencies in Bitcoin price data. The architecture is simple yet effective:

Table 17: LSTM Architecture.

Layer	Configuration
Input Layer	150 LSTM units to process sequential data and retain long-term dependencies.
Dense Layer	1 unit with linear activation to predict the next Bitcoin price (continuous regression output).
Optimizer	Adam optimizer, chosen for its adaptive learning capabilities and stability.
Loss Function	Mean Squared Error (MSE), used to minimize prediction errors.
Training	100 epochs with a batch size of 10 to balance learning efficiency and computational cost.

Source: Authors, (2026).

The LSTM architecture described in Table 17 is designed to effectively capture temporal dependencies in Bitcoin price data. LSTM networks are particularly suitable for time-series prediction because they incorporate memory cells that retain important information over long sequences. The model begins with an LSTM layer containing 150 units, allowing it to learn complex temporal relationships within the input sequences. A dense output layer with linear activation is used to generate continuous price predictions, which is appropriate for regression tasks. The Adam optimizer ensures efficient gradient-based learning, while the mean squared error loss function minimizes prediction errors during training. This architecture provides a balance between model complexity and computational efficiency.

VI.9 MODEL EVALUATION AND PREDICTION

Thus, after the training of the model, we check its efficiency through specific criteria such as Mean Squared Error (MSE). China's predicted prices and actual prices are also cross plotted to evaluate the performance as shown: The model is then used to predict prices of Bitcoin which assist in analyzing future trends. This actually gives elaboration on each step in the methodology making it clearer and making sure that any person who reads the methodology understands the steps that are being taken in the process.

VII. RESULTS AND DISCUSSION

VII.1 VISUALIZE TRAINING AND VALIDATION LOSS

Figure (2) demonstrates the Model Loss graph of the LSTM model during training and validation epoch by epoch up to one hundred. On the y-axis we have the loss values while the x-axis has the number of epochs. The first one in blue is the training loss, and the second one in orange is the validation loss. At first, when the number of epochs is low, which is the case in the first 10 epochs in this paper, the validation loss is much higher than the training loss which suggests that the model is yet to fully learn the patterns in the data and generalize the patterns to unseen data. Yet, the training and the validation losses decrease quickly during the first 20 epochs which shows that the model is learning significant features concerning the training set. After about 20 epochs, the validation loss becomes almost flat and fluctuates around a sufficiently small constant, while the training loss remains gradually decreasing and becomes almost zero. This shows that the model learning is able to capture the patterns of the data without overfitting the dataset as the difference between training and validation loss is still reasonable. The validation loss is relatively smooth and close to zero indicating good generalization capability when learning new data, although small oscillations in loss during the later epochs might have been caused by variations in the validation batches.

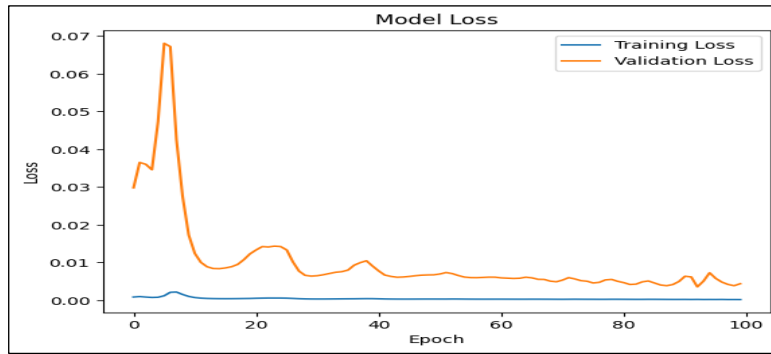


Figure 2. Model Loss.
Source: Authors, (2026).

In summary, the graph represents the LSTM model in a good performance learning in training phase and good generalization in validation phase. Such an advantage of the selected architecture, preprocessing steps, and hyper parameters has helped enhance the proposed model for accurate predictions of the given time-series data.

VII.2 MODEL EVALUATION

The performance of the LSTM model was evaluated using three key metrics: These include; Mean Absolute Error (MAE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). These metrics help understand the degree of correctness of the issued prediction by the model. The MAE quantifies the mean of the absolute differences between the estimated and the actual prices, on the other hand, the MSE quantifies the arithmetic mean of the squared differences with an inclination towards relatively large errors. RMSE again derived from MSE is more understandable as it is in the same unit as the target variable. Moreover, the observed prices were graphed with the predicted prices in order to get deeper insight and to see whether the model is capable of identifying patterns.

Table 18: Model Evaluation.

Metric	Value	Description
MAE	0.00622	The average absolute error between predicted and actual values, indicating high accuracy.
MSE	8.39651	The mean of squared errors, highlighting how well the model minimizes larger deviations.
RMSE	0.00916	The square root of MSE, showing an interpretable measure of prediction accuracy.

Source: Authors, (2026).

These metrics establish the model as a useful one because of low error values depicted below the bar for all the metrics. The comparison of the trends of the actual and predicted prices also supports the proposed model's ability to provide a good estimate for prices.

VII.3 VISUALIZATION ACTUAL VS. PREDICTED PRICES

The differential LSTM model is illustrated in the graph shown below titled "Actual vs. Predicted Cryptocurrency Prices" with days on the x-axis and normalized prices on the y-axis. The black line at the bottom is the actual values, while the grey line is the forecasted values. It accurately also reflects actual depreciation and appreciation of the model closely thus mimicking the actual price movements. In the less volatile periods (days 25–50 and 125–150), the model provides almost identical predictions to actual prices, proving the effectiveness of the model during stable periods. During more active periods (days 50–60, and 160–175), the model accurately identifies the direction and size of major price fluctuations with only slight distortion in both over- and underestimations of certain movement. Taken together, the differences between the actual and predicted prices are rather small and the low error measures (MAE, MSE, RMSE) prove the efficiency of the model. What this visualization shows is that LSTM model is indeed an ideal tool for predicting the future trends of cryptos' prices.

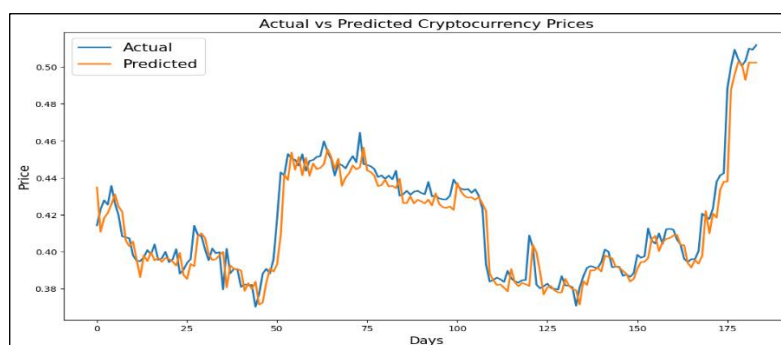


Figure 3: Visualization Actual vs. predicted prices.
Source: Authors, (2026).

VII.4 RESULTS DISCUSSION

Our analysis showed that the historical data can be employed to provide the price for the digital currency, with the focus on Bitcoin for this paper. The data can be considered as a factor to the model performance; for example, if the paper uses more data the prediction model performance may increase. There also has potential to generalize the method to other tasks such as parameter tuning for the model, testing various features, and using more complex deep learning architectures.

VIII. CONCLUSIONS

This paper shows how historical data and technical indicators can be employed for forecasting Bitcoin prices with reasonable accuracy. With the help of Yahoo Finance API, a detailed set of data from January 1, 2015, to June 1, 2024, was obtained and supplemented by Moving Averages, RSI, and Bollinger Bands. Using an LSTM-based deep learning model, we achieved robust results, with low error metrics (MAE: 0.000622, MSE: 8.39651, RMSE 0.00916 and close relationship between actual and predicted price movement also shows the robustness of the model. These results indicate that the model is capable of modelling both stable and unstableness of the market and therefore perhaps useful in the cryptocurrency prediction of the prices. We may venture into making future work improvements by adding extra attributes, adjusting the hyper parameters, or using the compound models to augment the predictions in the future. This work also emphasizes the fact that data quality and complex machine learning techniques are vital for solving the issues faced in regression model based cryptocurrency price prediction under a very noisy environment.

IX. AUTHOR'S CONTRIBUTION

Conceptualization: Ammar Ahmed Othman, Seddiq Hassan Al-Banna Ali and Mohammed Bakr Youssef.

Methodology: Ammar Ahmed Othman, Seddiq Hassan Al-Banna Ali and Mohammed Bakr Youssef.

Investigation: Ammar Ahmed Othman, Seddiq Hassan Al-Banna Ali and Mohammed Bakr Youssef.

Discussion of results: Ammar Ahmed Othman, Seddiq Hassan Al-Banna Ali and Mohammed Bakr Youssef.

Writing – Original Draft: Ammar Ahmed Othman, Seddiq Hassan Al-Banna Ali and Mohammed Bakr Youssef.

Writing – Review and Editing: Ammar Ahmed Othman, Seddiq Hassan Al-Banna Ali and Mohammed Bakr Youssef.

Resources: Ammar Ahmed Othman, Seddiq Hassan Al-Banna Ali and Mohammed Bakr Youssef.

Supervision: Ammar Ahmed Othman, Seddiq Hassan Al-Banna Ali and Mohammed Bakr Youssef.

Approval of the final text: Ammar Ahmed Othman, Seddiq Hassan Al-Banna Ali and Mohammed Bakr Youssef.

X. REFERENCES

- [1] U. Kayani and F. Hasan, "Unveiling Cryptocurrency Impact on Financial Markets and Traditional Banking Systems: Lessons for Sustainable Blockchain and Interdisciplinary Collaborations," *J. Risk Financ. Manag.*, vol. 17, no. 2, 2024, doi: 10.3390/jrfm17020058.
- [2] P. Rahman, Z. Zhang, and M. Musa, *Do technological innovation, foreign investment, trade and human capital have a symmetric effect on economic growth? Novel dynamic ARDL simulation study on Bangladesh*, vol. 56, no. 2. Springer US, 2023. doi: 10.1007/s10644-022-09478-1.
- [3] G. Soana and T. de Arruda, "Central Bank Digital Currencies and financial integrity: finding a new trade-off between privacy and traceability within a changing financial architecture," *J. Bank. Regul.*, vol. 25, no. 4, pp. 467–486, 2024, doi: 10.1057/s41261-024-00241-2.
- [4] P. K. Ozili, "Central Bank Digital Currency," pp. 94–107, 2024, doi: 10.4018/979-8-3693-1503-3.ch005.
- [5] G. Habib, S. Sharma, S. Ibrahim, I. Ahmad, S. Qureshi, and M. Ishfaq, "Blockchain Technology: Benefits, Challenges, Applications, and Integration of Blockchain Technology with Cloud Computing," *Futur. Internet*, vol. 14, no. 11, pp. 1–22, 2022, doi: 10.3390/fi14110341.
- [6] J. Zarrin, H. Wen Phang, L. Babu Saheer, and B. Zarrin, "Blockchain for decentralization of internet: prospects, trends, and challenges," *Cluster Comput.*, vol. 24, no. 4, pp. 2841–2866, 2021, doi: 10.1007/s10586-021-03301-8.
- [7] H. Li, "The Applications of Cryptocurrency: Evidence from Ethereum," *BCP Bus. Manag.*, vol. 34, pp. 237–242, 2022, doi: 10.54691/bcpbm.v34i.3019.
- [8] K. Shah, D. Lathiya, N. Lukhi, K. Parmar, and H. Sanghvi, "A systematic review of decentralized finance protocols," *Int. J. Intell. Networks*, vol. 4, no. July, pp. 171–181, 2023, doi: 10.1016/j.ijin.2023.07.002.
- [9] H. Bi *et al.*, "Not Just Learning From Others but Relying on Yourself: A New Perspective on Few-Shot Segmentation in Remote Sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, no. 8, pp. 1–21, 2023, doi: 10.1109/TGRS.2023.3326292.
- [10] D. L. John, S. Binnewies, and B. Stantic, "Cryptocurrency Price Prediction Algorithms: A Survey and Future Directions," *Forecasting*, vol. 6, no. 3, pp. 637–671, 2024, doi: 10.3390/forecast6030034.
- [11] M. Guindy, "Cryptocurrency price volatility and investor attention," *Int. Rev. Econ. Financ.*, vol. 76, Jun. 2021, doi: 10.1016/j.iref.2021.06.007.
- [12] P. Prabakaran and N. Panchalingam, "Factors Influencing the Trend of Cryptocurrency View project Factors Influencing the Trend of Cryptocurrency," no. January, 2022, [Online]. Available: <https://www.researchgate.net/publication/357982093>
- [13] Q. Li, F. Wang, J. Wang, and W. Li, "LSTM-Based SQL Injection Detection Method for Intelligent Transportation System," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4182–4191, 2019, doi: 10.1109/TVT.2019.2893675.
- [14] Y. Liu and A. Tsyvinski, "Risks and Returns of Cryptocurrency," *Rev. Financ. Stud.*, vol. 34, Sep. 2020, doi: 10.1093/rfs/hhaa113.
- [15] T. Bouazizi, E. Galarotis, K. Guesmi, and P. Makrychoriti, "Investigating the nature of interaction between crypto-currency and commodity markets," *Int. Rev. Financ. Anal.*, vol. 88, 2023, doi: 10.1016/j.irfa.2023.102690.

- [16] A. Ladhari and H. Boubaker, "Deep Learning Models for Bitcoin Prediction Using Hybrid Approaches with Gradient-Specific Optimization," *Forecasting*, vol. 6, no. 2, pp. 279–295, 2024, doi: 10.3390/forecast6020016.
- [17] M. Rizwan, S. Narejo, and M. Javed, *Bitcoin price prediction using Deep Learning Algorithm*. 2019. doi: 10.1109/MACS48846.2019.9024772.
- [18] X. Jiang, "Bitcoin Price Prediction Based on Deep Learning Methods," *J. Math. Financ.*, vol. 10, no. 01, pp. 132–139, 2020, doi: 10.4236/jmf.2020.101009.
- [19] J. Bajwa, U. Munir, A. Nori, and B. Williams, "Artificial intelligence in healthcare: transforming the practice of medicine.," *Futur. Healthc. J.*, vol. 8, no. 2, pp. e188–e194, Jul. 2021, doi: 10.7861/fhj.2021-0095.
- [20] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Comput. Sci.*, vol. 2, no. 3, p. 160, 2021, doi: 10.1007/s42979-021-00592-x.
- [21] D. Sambyal and A. Sarwar, "Recent developments in cervical cancer diagnosis using deep learning on whole slide images: An Overview of models, techniques, challenges and future directions," *Micron*, vol. 173, p. 103520, 2023, doi: <https://doi.org/10.1016/j.micron.2023.103520>.
- [22] A. Gatera, M. Kuradusenge, G. Bajpai, C. Mikeka, and S. Shrivastava, "Comparison of random forest and support vector machine regression models for forecasting road accidents," *Sci. African*, vol. 21, p. e01739, 2023, doi: 10.1016/j.sciaf.2023.e01739.
- [23] J. Wang and F. Biljecki, "Unsupervised machine learning in urban studies: A systematic review of applications," *Cities*, vol. 129, no. December 2021, p. 103925, 2022, doi: 10.1016/j.cities.2022.103925.
- [24] G. H. Rosenlund, K. W. Hoiem, B. N. Torsater, and C. A. Andresen, "Clustering and dimensionality-reduction techniques applied on power quality measurement data," *SEST 2020 - 3rd Int. Conf. Smart Energy Syst. Technol.*, no. October, 2020, doi: 10.1109/SEST48500.2020.9203294.
- [25] M. M. Taye, "Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions," *Computers*, vol. 12, no. 5. 2023. doi: 10.3390/computers12050091.
- [26] T. N. Sainath, O. Vinyals, A. Senior, and H. Sak, "Convolutional, Long Short-Term Memory, fully connected Deep Neural Networks," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 2015-August, pp. 4580–4584, 2015, doi: 10.1109/ICASSP.2015.7178838.
- [27] J. H. Park, J. H. Kim, and J. H. Huh, "Deep Reinforcement Learning Robots for Algorithmic Trading: Considering Stock Market Conditions and U.S. Interest Rates," *IEEE Access*, vol. 12, pp. 20705–20725, 2024, doi: 10.1109/ACCESS.2024.3361035.
- [28] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: state of the art, current trends and challenges," *Multimed. Tools Appl.*, vol. 82, no. 3, pp. 3713–3744, 2023, doi: 10.1007/s11042-022-13428-4.
- [29] A. Mohammed and R. Kora, "A comprehensive review on ensemble deep learning: Opportunities and challenges," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 35, no. 2, pp. 757–774, 2023, doi: 10.1016/j.jksuci.2023.01.014.
- [30] V. Kotu and B. Deshpande, "Chapter 12 - Time Series Forecasting," V. Kotu and B. B. T.-D. S. (Second E. Deshpande, Eds., Morgan Kaufmann, 2019, pp. 395–445. doi: <https://doi.org/10.1016/B978-0-12-814761-0.00012-5>.
- [31] N. ElKalaawy and A. Wassal, "Methodologies for the modeling and simulation of biochemical networks, illustrated for signal transduction pathways: A primer," *Biosystems*, vol. 129, pp. 1–18, 2015, doi: <https://doi.org/10.1016/j.biosystems.2015.01.008>.
- [32] M. A. Albadr, S. Tiun, M. Ayob, and F. Al-Dhief, "Genetic algorithm based on natural selection theory for optimization problems," *Symmetry (Basel)*, vol. 12, no. 11, pp. 1–31, 2020, doi: 10.3390/sym12111758.