Received: 27 March 2020 | Revised: 2 November 2020 | Accepted: 12 January 2021

DOI: 10.1002/isaf.1488

REVIEW

WILEY

Cryptocurrency price prediction using traditional statistical and machine-learning techniques: A survey

Ahmed M. Khedr[1,2] | Ifra Arif[1] | Pravija Raj P V[1] | Magdi El-Bannany[3] | Saadat M. Alhashmi[4] | Meenu Sreedharan[1]

[1]Department of Computer Science, University of Sharjah, Sharjah, 27272, UAE

[2]Mathematics Department, Zagazig University, Zagazig, Egypt

[3]Department of Accounting,University of Sharjah, Sharjah, 27272, UAE

[4]Department of Information Systems, University of Sharjah, Sharjah, 27272, UAE

Correspondence
Ahmed M. Khedr, Department of Computer Science, University of Sharjah, Sharjah, 27272, UAE.
Email: akhedr@sharjah.ac.ae

Abstract

Cryptocurrencies are decentralized electronic counterparts of government-issued money. The first and best-known cryptocurrency example is bitcoin. Cryptocurrencies are used to make transactions anonymously and securely over the internet. The decentralization behavior of a cryptocurrency has radically reduced central control over them, thereby influencing international trade and relations. Wide fluctuations in cryptocurrency prices motivate the urgent requirement for an accurate model to predict its price. Cryptocurrency price prediction is one of the trending areas among researchers. Research work in this field uses traditional statistical and machine-learning techniques, such as Bayesian regression, logistic regression, linear regression, support vector machine, artificial neural network, deep learning, and reinforcement learning. No seasonal effects exist in cryptocurrency, making it hard to predict using a statistical approach. Traditional statistical methods, although simple to implement and interpret, require a lot of statistical assumptions that could be unrealistic, leaving machine learning as the best technology in this field, being capable of predicting price based on experience. This article provides a comprehensive summary of the previous studies in the field of cryptocurrency price prediction from 2010 to 2020. The discussion presented in this article will help researchers to fill the gap in existing studies and gain more future insight.

KEYWORDS

cryptocurrency price prediction, bitcoin (BTC), machine learning (ML), reinforcement learning (RL), deep learning (DL)

1 | INTRODUCTION

A cryptocurrency is a digital or virtual currency used as a mode of exchange and transfer of assets digitally. It uses cryptography to transfer assets securely, to control and regulate the addition of cryptocurrencies, and secure their transactions (Garcia, Tessone, Mavrodiev, & Perony, 2014), hence the name cryptocurrency. Cryptocurrencies are established on the principle of decentralized control when compared with standard currencies, which rely on central banking systems. Thus, a cryptocurrency is used to electronically transfer funds without the interference of a central entity, such as a bank. The cryptocurrency market has evolved exponentially in a short period owing to its uncontrolled and untraceable nature. Digital currencies are used for financial transactions worldwide and are becoming popular. It is now an exciting research area, and many researchers are finding ways to analyze cryptocurrency features, such as stock market and market price prediction, and analyze their impact on real life. The increased interest of cryptocurrencies in economics and the financial world has attracted researchers to this area. However, the applications of cryptocurrency and its associated technologies are

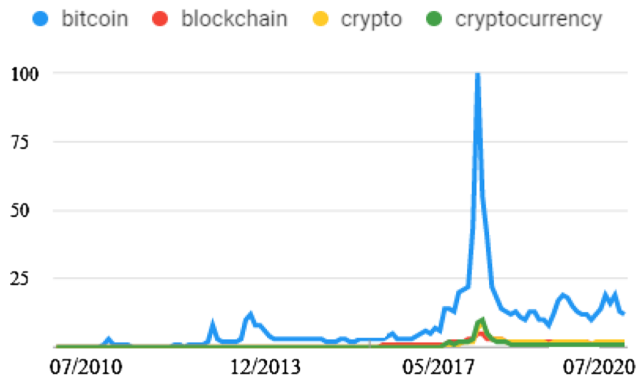


FIGURE 1 Global search trends in the field of cryptocurrency

not necessarily finance related; substantial computer science literature exists on the supporting cryptocurrency technologies that can contribute toward new and efficient approaches appropriate for handling bitcoin and other cryptocurrencies, their price volatility, and other associated technologies. This paper provides a comprehensive survey of cryptocurrency price prediction research in the period 2010–2020, by covering the significant research studies on various aspects of cryptocurrency price prediction, which include both statistical and machine-learning (ML) approaches. This paper also gives an insight into the data sets, research trends and techniques, and prediction approaches, concluding with some promising opportunities that remain open in cryptocurrency price prediction research.

An integrated overview of the characteristics of cryptocurrencies was provided in the seminal review paper of Corbet, Lucey, Urquhart, and Yarovaya (2019). Kyriazis (2019) introduced survey papers on the efficiencies of cryptocurrency markets in 2019 and the bubble characteristics of cryptocurrencies in 2020 (Kyriazis, Papadamou, & Corbet, 2020). A survey on anticipating the prices of cryptocurrencies using deep learning (DL) was introduced by Akshaya, Eswari, Dharani, and Lalitha (2019). The survey gives a good introduction to the existing ML approaches used in this field, but it does not provide details on each technique and how they were used in the prediction. Moreover, there is no comparison between the ML and DL approaches used in this field. In this survey, we tried to provide a state-of-the-art snapshot of the statistical and ML models developed for cryptocurrency price prediction. To the best of our knowledge, this is the first comprehensive study on cryptocurrency price prediction detailing the different traditional and ML approaches used in this field. All searchable articles on cryptocurrency price prediction related to traditional, ML, DL, and reinforcement-learning (RL) approaches are reviewed in our study. This survey gives an insight into the current status of cryptocurrency price prediction. We have categorized and analyzed the studies according to the different techniques used, which can provide researchers with knowledge related to the data sets, research trends and techniques, and prediction approaches, and also some promising opportunities that remain open in cryptocurrency price prediction research.

1.1 | Background of cryptocurrency

The traditional monetary system involves a set of rules, policies, frameworks, and institutions that a government uses to create money in the economy. The main participants involved in this system include the central bank, the national treasury, the mint, and commercial banks. Commodities, commodity-backed assets, and fiat currency are three traditional types of monetary systems. In a commodities-type monetary system, some precious metals or commodities with intrinsic value are exchanged physically as a currency. The obvious examples of this are gold and silver coins that have been used widely throughout history. These type of systems are not divisible, which means that they are inconvenient to use for purchases. They may also suffer from the bandwagon effect, in which prices may fluctuate based on the buying behavior of the general population. In a commodity-backed monetary system, money draws its value from a commodity having no actual physical value, such as paper notes. The gold standard is a famous example of this system. In this system, the more valuable commodity will gradually disappear from circulation and so people will not prefer this for regular transactions. Fiat money is the most common and widespread monetary system in which governments guarantee the value of a currency. In this system, people use notes or bank balances as the mode of exchange and store. In this system, risk arises at the moment the funds are deposited with the bank. Central banks may cause inflation by printing and devaluing money.

From the limitations of the traditional systems, the idea of using currency in digital form has emerged. One of the main reasons for introducing cryptocurrency is to deal with imperfections present in the existing monetary system, which relies on fiat money and leads to inflation. The main aim of such a digital monetary system is to solve the problem of inflation and negative yields for consumers and provide better financial stability. This might enhance the convenience, speed, and cost, and thereby provides economic benefit. The possible ways of implementing a digital monetary system include central-bank-based systems with or without decentralization, a fully decentralized digital monetary system that replaces the monopolistic power held by the central banks, and blockchain-based digital monetary systems. Among these, the blockchain-based systems are more successful in providing truly decentralized solutions to the public.

The first and most prominent cryptocurrency is bitcoin, introduced by pseudonymous developer Satoshi Nakamoto (Nakamoto & Bitcoin, 2008). In January 2009, Bitcoin was implemented and released to the public as an open-source code by Nakamoto. The top three cryptocurrencies circulating in the market are bitcoin, altcoin, and tokens. Civic and BitDegree are examples of tokens/dApps. Cryptocurrency technology moved the financial market one extra step toward the future by decentralizing the currency and releasing it from the hierarchical power structures. As a substitute, consumers and organizations perform transactions digitally on a peer-to-peer network. Within a short period of its existence, the cryptocurrency market has experienced exponential growth and widespread popularity. In recent years, cryptocurrencies have gained increasing popularity and have received worldwide attention from the media, attracting investors, academia, governments, regulators, and speculators. Figure 1 shows the global search trends in the field of cryptocurrency during the period 2010–2020 (https://trends.google.com). The use of cryptocurrency is growing, and it is necessary to study its impact on countries' monetary systems. The future of bitcoin, or any cryptocurrency, is not limited to any particular discipline; rather, it transcends every field (Holub & Johnson, 2018).

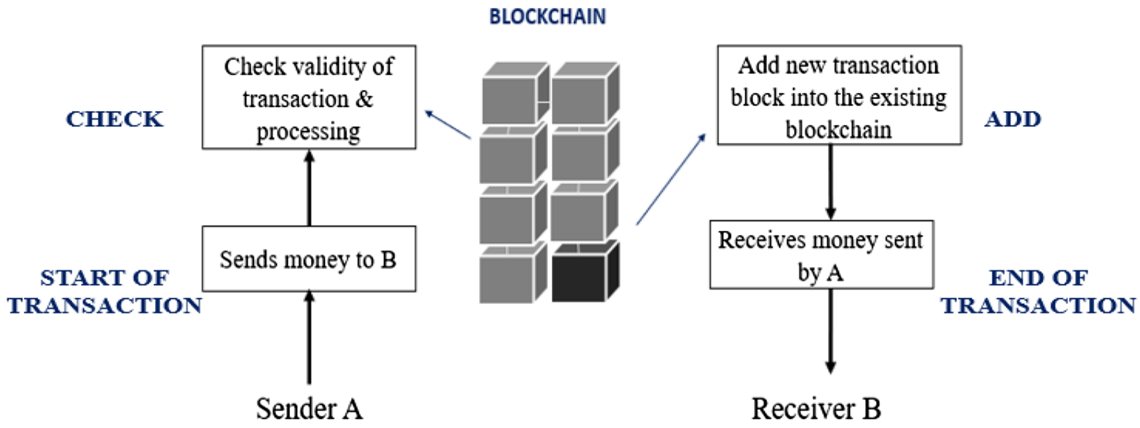


FIGURE 2 Cryptocurrency workflow using blockchain mechanism

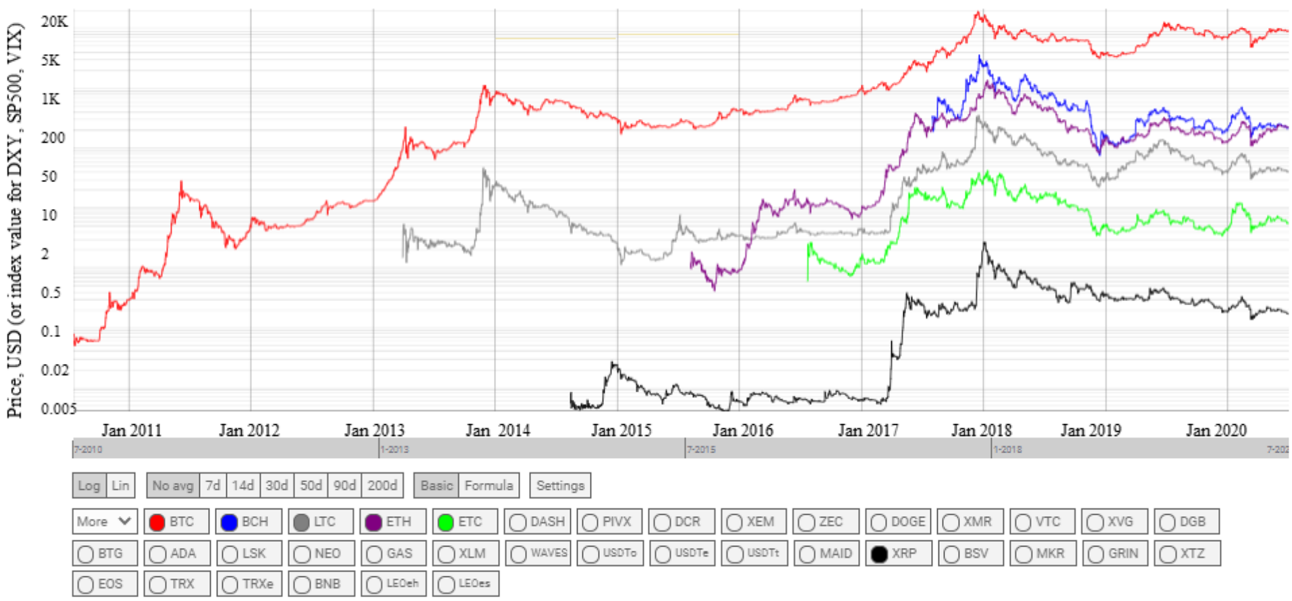


FIGURE 3 Volatility dynamics of top cryptocurrencies during 2010–2020

1.2 | Features of cryptocurrencies and need for price prediction

Cryptocurrencies are used for many useful purposes, such as online transaction systems, and their usage is increasing very fast. Cryptocurrency-based transactions involve a decentralized and distributed peer-to-peer system that allows the information to be recorded in an open transaction ledger, called the blockchain. It is not under the control of any company or the government (Mittal, Arora, & Bhatia, 2018). The blockchain supports transactional databases and offers greater transparency, which is unfamiliar in the world of classical financial markets. If the block refers to a financial transaction, then each transaction in the blockchain, by definition, includes information about previous transactions, and thus verifies the ownership of the financial asset being transferred. A sample blockchain database is shown in Figure 2, in which there are two clients, A and B, where client A wants to send money to client B. Authentication of each transaction from client A to client B can be verified through this distributed ledger or blockchain that is maintained by all participants. It checks the validity of the transaction at the sender and adds transaction information into the global storage at the receiver end, and then the transaction is closed.

The crypto industry has gained increasing popularity over the past years. Governments are now aiming to consider both taxation and regulation of cryptocurrencies. Accountants and auditors are now seeking business guidance from the standard setters, as they are now accepting cryptocurrencies as a payment form. Cryptocurrencies are primarily volatile and characterized by the number of transactions and the changes in their prices; this makes cryptocurrency price prediction challenging. The popular cryptocurrencies with the largest market capitalizations are bitcoin, ethereum, ripple, and litecoin. Cryptocurrencies operate differently and are distinguished from one another mainly due to their values, transaction speeds, usages, and volatility characteristics. For instance, by the end of 2013 there was no significant price fluctuation with bitcoin, whereas other cryptocurrencies, such as litecoin and ripple, have showed significant instability in price since the end of 2013 (Böhme, Christin, Edelman, & Moore, 2015). Though cryptocurrency prices have soared since 2016, with great fluctuation, people's interest in it has stayed more or less constant. Therefore, in recent years, various techniques have been proposed by

TABLE 1 Volatility trends of bitcoin in the past 10 years

Period	Change ($)	Change (%)	Trend
Last 30 days	−166.86	−1.79	↘
Last 6 months	+304.31	+3.43	↗
Last 1 year	−1,427.26	−13.47	↘
Last 2 years	+1,824.36	+24.84	↗
Last 4 years	+8,491.00	+1,255.43	↗
Last 6 years	+8,539.22	+1,359.49	↗
Last 8 years	+9,158.34	+101,759.33	↗
Last 10 years	+9,153.30	+65,194.44	↗

researchers to predict and model the price of cryptocurrencies and to analyze the volatility of the crypto market. The interest in cryptocurrencies increased even more after the great cryptocurrency crash (also called the Bitcoin Crash) at the beginning of 2018. The price reached a peak of nearly $20,000 per bitcoin in late 2017 and has since fluctuated quite a bit (2018–2019), averaging at about $7,000 as of April 2020. Crypto Research Report (2020) presents price predictions for several cryptocurrencies (bitcoin, bitcoin cash, ethereum, litecoin, and stellar); based on a new comprehensive analysis, the price of bitcoin was predicted to reach almost $20,000 in 2020 and to keep rising to almost $400,000 by 2030. Figure 3 shows the graph of cryptocurrency volatility for the top cryptocurrencies during the period 2010–2020 (https://coinmarketcap.com). A short summary on the volatility trends of bitcoin and other top cryptocurrencies currently circulating in the market can be found in Table 1 (https://coin.dance) and Table 2 (https://www.tradingview.com) respectively.

It is difficult to say exactly what drives the price of these cryptocurrencies over time. As with most price fluctuations in the cryptocurrency world, the exact causes are difficult to pinpoint. The entire crypto industry is known for its extreme volatility. The top 10 cryptocurrencies represent approximately 85% of the total market share, with bitcoin dominating with about 64% of the market capitalization (https://coinmarketcap.com). With the ever increasing interest in cryptocurrencies and their significance in the financial world, it is necessary to have a comprehensive analysis on and forecasting of the volatility dynamics of cryptocurrencies.

However, despite the growing interest, acceptance, and integration of cryptocurrencies in global financial markets, there is limited research on modeling the volatility dynamics of cryptocurrencies. Such unstable fluctuation are difficult to predict for users. Several factors may influence prices of cryptocurrencies over the years. This includes both internal and external factors. Factors related to the crypto market (e.g., trading volume, market beta, and volatility) can be regarded as one of the significant factors determining cryptocurrency price (during 2010–2020). Moreover, the attractiveness of cryptocurrencies can also influence their price (Sovbetov, 2018). Figure 4 provides an overview of the various types of factors that influence cryptocurrency prices.

Generating an accurate prediction model for such a complex problem is very challenging. The problem of cryptocurrency price prediction is still in its nascent stages and requires further research efforts to explore this area. So, the main objective of this work is to review all the techniques and methods used between 2010 and 2020 in cryptocurrency price prediction. Statistical and ML approaches are the two commonly used techniques in cryptocurrency price prediction. The motivation of this work is the speed of acceptance or recognition of cryptocurrencies as financial instruments and the increased utilization of ML techniques in predicting time-series problems to seek more accurate predictions. We aim to synthesize the collective knowledge from the first 10 years of cryptocurrency operation and development and to highlight the significant techniques adopted for cryptocurrency price prediction, ranging from traditional statistical methods to the recent ML approaches.

Figure 5 gives an outline of the survey theme and paper selection. This study also compares the different ML models to identify efficient and robust approaches for price prediction. This article can help researchers in developing robust models for cryptocurrency price prediction, using the conclusions derived at the end.

We divide the rest of this article into the following sections. In Section 2, we discuss the methodology used to conduct this survey. In Section 3, we review the common statistical and ML techniques. In Section 4, traditional statistical techniques for prediction of cryptocurrency prices are discussed. In Section 5, we discuss the ML techniques for prediction of cryptocurrency prices. Section 6 provides the discussion, and future directions are listed in Section 7. Finally, Section 8 concludes our work.

2 | METHODOLOGY

Research interest in the field of cryptocurrencies has started to increase considerably over the past few years. With the increasing popularity of cryptocurrencies and the worldwide attention as an emerging financial market, the number of publications in this area has been increasing since 2017, and especially in 2018 and 2019; and the trend continued in 2020. It is interesting that the research studies initially focused more on bitcoin than on the cryptocurrency topic in general. However, since 2018, the focus on the general cryptocurrency topic has started increasing (https://www.scopus.com, https://onlinelibrary.wiley.com/). The topic of cryptocurrency price prediction is still in its nascent stages and requires further research efforts to explore this area. When we consider only the finance-related literature (which includes the areas of finance, economics, business, management, and accounting), the number of publications is not very high. However, it is notable that the topic is being explored and is gaining attention in other disciplines, including mathematics, engineering, and computer science. The major milestones in cryptocurrency price prediction research are depicted in Figure 6.

In this paper, we present a systematic literature review of cryptocurrency price prediction using traditional statistical and ML techniques. We searched for topic-related keywords and short phrases such as "cryptocurrency price prediction," "ML techniques (and) cryptocurrency price prediction," "cryptocurrency price prediction using ML techniques," and "DL (and) cryptocurrency price prediction." After conducting the search process, the papers were screened and classified into different categories that included traditional statistical and ML techniques. We further classified the latter into different ML and DL techniques. The majority of the papers are related to use of ML techniques in predicting the price of a cryptocurrency, and some of the papers fall under DL and RL techniques. All of these papers

Name	Change (%)	Performance (%) Weekly	Monthly	6-Month	Year to date	Yearly	Volatility
Bitcoin	−0.25	−1.05	0.40	71.32	27.71	4.98	0.55
Ethereum	−0.55	−2.48	4.13	89.97	82.11	26.91	1.33
Tether	0.03	0.05	0.13	0.31	0.56	0.19	0.20
XRP	−0.83	−0.36	12.02	29.82	2.98	−27.65	3.10
Bitcoin cash	−0.05	−3.25	1.16	25.22	9.32	−21.73	0.61
Cardano	1.09	−6.16	54.73	363.11	279.22	188.38	2.81
Bitcoin SV	−0.87	−4.51	7.31	41.61	78.80	32.06	1.19
ChainLink	−0.04	−1.84	74.45	282.45	352.10	190.72	3.30
Binance coin	3.58	−1.71	15.39	71.53	29.96	−11.50	6.19
Litecoin	−0.15	−3.32	2.53	16.94	3.04	−31.32	1.20

TABLE 2 Volatility trends of top cryptocurrencies currently circulating in the market

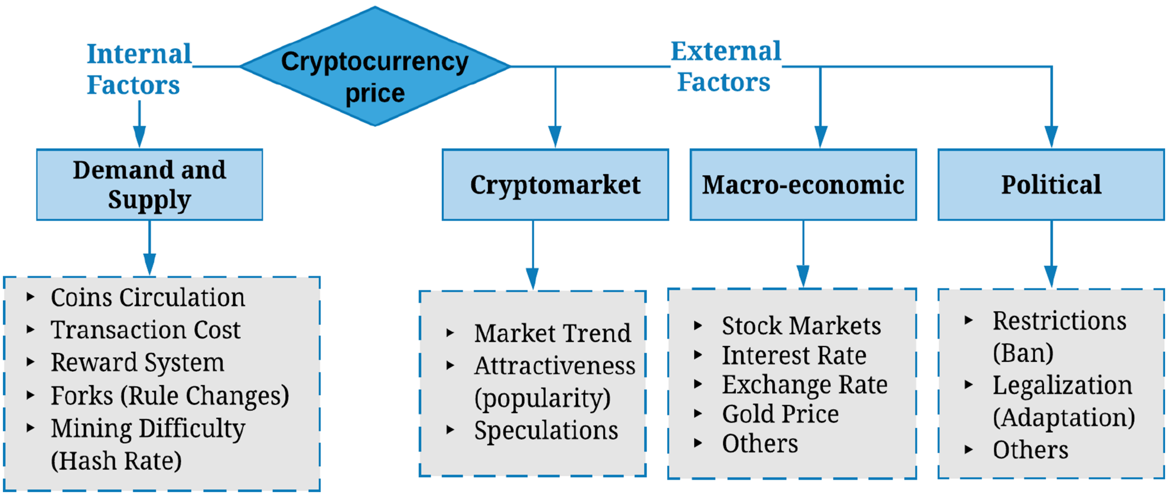


FIGURE 4 Overview of common factors influencing cryptocurrency price

talk about cryptocurrency price prediction using different methods. This area is still new, and the number of published papers is not very high; nevertheless, it is a very crucial area and deserves to be explored because of the impact it is having on the financial system. Figure 7 gives an illustration of the survey process and the scheme used.

3 | COMMON STATISTICAL AND ML TECHNIQUES: AN OVERVIEW

The emergent price prediction schemes for cryptocurrency include strategies that are based on statistical and ML technologies. Generally, the statistical models use mathematical equations to encode information extracted from the data. The traditional techniques for cryptocurrency price prediction usually adopted statistical and econometric models (Brooks, 2019). Econometric approaches apply an integration of statistical and economic theories to estimate and predict the values of various economic variables. In some cases, statistical-model-based techniques can quickly provide adequate models (Wang & Chen, 2020). A linear statistical-model-based approach evaluates the linear relationship between prices and an explanatory variable. If multiple explanatory variables exist, it is possible to model the linear relationship between explanatory variables (independent) and response variables (dependent) with the help of multiple linear models. The commonly used linear statistical model for time-series analysis is the autoregressive moving-average model (Choi, 1992). While investigating the cryptocurrency price fluctuations using econometrics, researchers usually utilize statistical models on time-series data. Among these models, the most widely used ones are the generalized autoregressive conditional heteroscedasticity (GARCH) model, multivariate linear regression, multivariate vector autoregressive model, and extended vector autoregressive model (Fang et al., 2020).

With the advancement of big data technology and artificial intelligence, numerous research studies have applied ML models to classification and prediction problems (El-Bannany, Sreedharan, & Khedr, 2020; Sreedharan, Khedr, & El-Bannany, 2020a; 2020b). Many researchers have focused their efforts on applying these new techniques on financial markets (Dixon, Halperin, & Bilokon, 2020; El-Bannany et al., 2020; Galeshchuk & Mukherjee, 2017; Hatefi Ghahfarrokhi & Shamsfard, 2020; Nikou, Mansourfar, & Bagherzadeh, 2019; Sarlin & Marghescu, 2011; Sreedharan et al., 2020a; 2020b). Hatefi Ghahfarrokhi and Shamsfard (2020) investigated the impact of social media data in predicting the Tehran Stock Exchange variables. Galeshchuk and Mukherjee (2017) investigated the ability of deep convolution neural networks (NNs) to predict the direction of change in forex rates. Sarlin and Marghescu (2011)) constructed a neuro-genetic model for predicting currency crises by using a genetic algorithm for specifying (1) the combination of inputs, (2) the network configuration, and (3) the training parameters for a back-propagation artificial NN (ANN). Nikou et al., (2019) evaluated

FIGURE 5 Outline of the subject matter in this survey

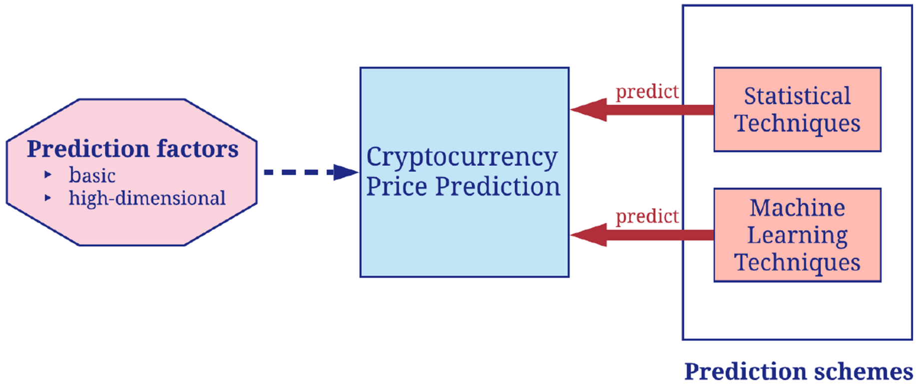
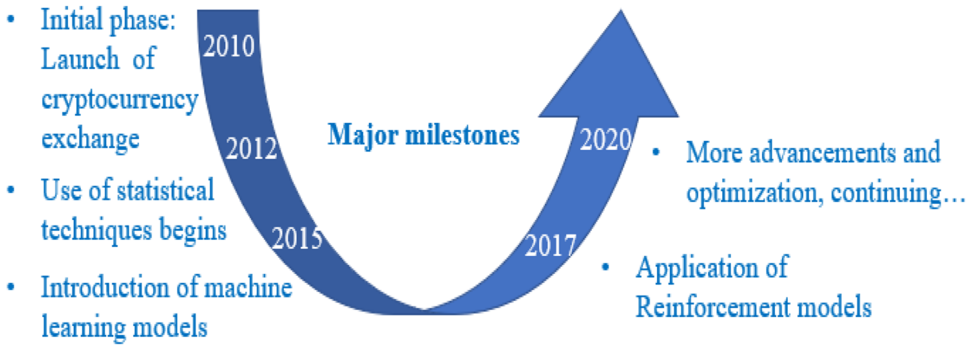


FIGURE 6 Major milestones in cryptocurrency price prediction research



the prediction power of ML models in a stock market. In this context, in the financial industry, the application and use of ML algorithms for cryptocurrency price prediction is increasing and is getting attention from researchers in various disciplines as well. ML is a technique that has used as a model or framework for predicting various aspects across the industry for over three decades. In recent years, ML and its associated techniques have made notable advances in various fields. ML is categorized into supervised and unsupervised learning. In supervised learning, labeled instances are present in a data set, whereas unsupervised learning has no such labels. Examples of supervised learning techniques include NNs and support vector machines (SVMs), whereas clustering techniques fall under unsupervised learning. ANNs and SVMs are the two most widely used algorithms for predicting price fluctuation (Patel, Shah, Thakkar, & Kotecha, 2015). Deriving a function from a training data set is the main task in supervised learning. Some of the main ML techniques discussed are as follows.

- **Logistic regression:** Logistic regression is one of the traditional variate regression methods, mainly used in binary classification problems (Kleinbaum & Klein, 2002). The output or response variable $y \in [0, 1]$ indicates a class label, which can be predicted using the input feature value x_i, where $i = 1, \ldots, k$. The logistic regression model can be represented as follows:

$$\text{logit}(P(y)) = \log \frac{P(y)}{1 - P(y)} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k \quad (1)$$

- **Support vector machine:** The SVM was introduced as an induction principle by Vapnik (1999) to avoid overfitting of data. SVMs are primarily used to identify the maximum margin hyperplane. It has a high-dimensional feature space and is also used for system learning. This technique is flexible as it creates explicit and accurate boundaries that lead to fast training results, besides being easy to use (Zhang & Wang, 2015). It performs well with a small data set and provides a nonlinear solution by applying a kernel function to map the input variables into high-dimensional space. It produces a classification hyperplane that is used to differentiate between two classes with maximum margin and is used to solve the pattern classification problem.
- **Artificial neural network:** An ANN is comprised of interconnected units called neurons that are activated depending on the input. It mimics the human brain in information processing and interacts with other processing features (Lu, 2010). The network consists of input and output neurons, where input neurons are triggered based on the sensing environment. Other neurons are activated using weighted connections from neurons. The sigmoid function is employed at each of the hidden layers as a transfer function. To adjust the weights, gradient descent with momentum is used so that the global minimum can be achieved.
- **Random forest:** Random forest (RF) is one of the popular approaches used for performing classification tasks. It uses an ensemble of decision trees for better classification results (Chen, Li, & Sun, 2020). The decision tree is one of the vital ML methods that uses a tree structure to iteratively partition the feature space (a node) until a single class sample is obtained. This pure node at the end is called a leaf node, and a class label is assigned to this node. By performing bootstrap aggregation and a random feature selection process, RF allows a random subset of the whole feature space to be assigned to the growth of each tree.
- **Deep learning:** DL is a subclass of ML algorithms that uses multiple layers to gradually extract high-level features based on raw data

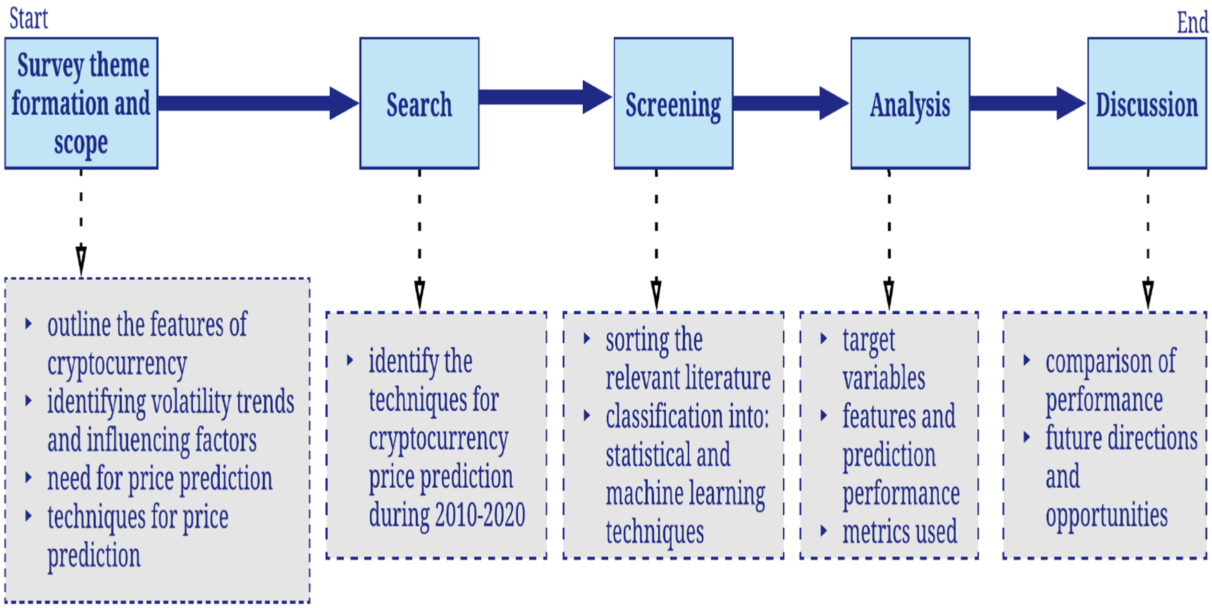


FIGURE 7 Survey scheme

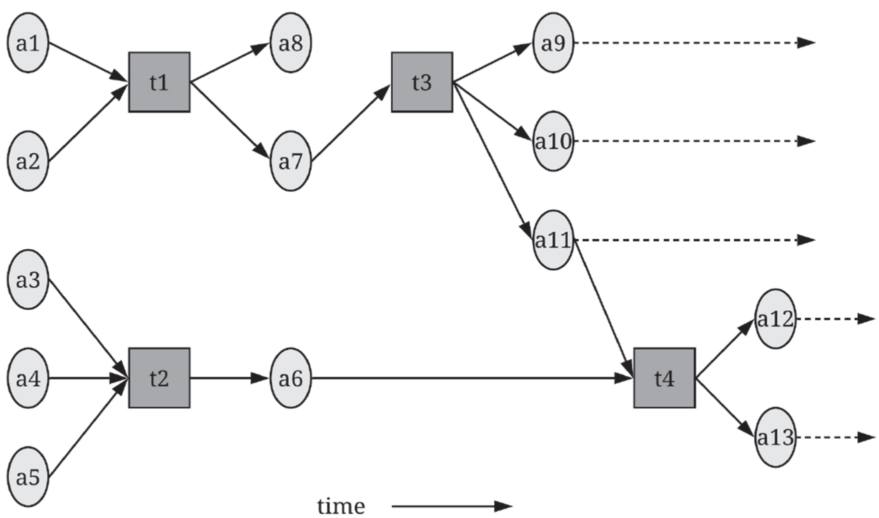


FIGURE 8 A transaction–address graph representation of the bitcoin network

input. The term "deep" refers to the number of processing layers through which the data transformation takes place. It is used to get more precise details using the layered approach. For example, in image processing, details such as an edge can be identified using lower level layers while all high-level information, such as letters, digits, or faces, can be extracted using higher level layers. Most of the modern DL models are based on ANNs (Längkvist, Karlsson, & Loutfi, 2014).

- **Reinforcement learning:** In the reinforcement learning (RL) process, learning is achieved through interaction between learning objects and its related environment. Objects try to learn using the trial-and-error method. RL consists of three components: a value function, an environment, and a reinforcement function. The environment for RL is often dynamic, with a set of probable states. For each state, there exist a set of possible actions at each time (Dixon et al., 2020).
- **Gradient boosting**: Gradient boosting (GB) is a technique for both regression and classification problems. Similar to RFs, GB is an ensemble learner. This means it will create a final model based on a collection of individual models. It produces a prediction model that is an ensemble of weak prediction models, such as decision trees (Li, Chamrajnagar, Fong, Rizik, & Fu, 2019; Sun, Liu, & Sima, 2020). It builds the model in a stage-wise fashion, and it generalizes them by allowing optimization of an arbitrary differentiable loss function. The implementations of this technique can have different names, among which the common ones are GB machines (GBMs) and XGBoost.

4 | TRADITIONAL STATISTICAL TECHNIQUES FOR CRYPTOCURRENCY PRICE PREDICTION

The traditional approaches for cryptocurrency price prediction usually applied statistical and econometric models. Econometric approaches apply an integration of statistical and economic theories to estimate and predict the values of various economic variables (Brooks, 2019). While examining the cryptocurrency price volatility and prediction using econometrics, researchers adopted statistical models on time-series data, generally. In this section, we review the existing statistical and econometric techniques for cryptocurrency price prediction.

Classical approaches were used previously, such as Holt–Winters exponential smoothing by Chatfield and Yar (1988), to forecast time-series data. This approach depends on linear assumptions. In this approach, input data are segregated into several trends and are used for predicting features that have seasonal effects, such as in sales. This approach cannot be used to predict cryptocurrency price accurately as there are no seasonal effects with cryptocurrency. Sovbetov (2018) employed the augmented Dickey–Fuller unit-root test and bound testing approach to examine factors influencing the prices of five cryptocurrencies (bitcoin, ethereum, dash, litecoin, and monero) over 2010–2018 using weekly data. The results revealed that market beta, trading volume, and volatility have an influential impact on the prices of all five cryptocurrencies both in the short and long run.

Roy, Nanjiba, and Chakrabarty (2018), using annual bitcoin data from 2013 to 2017, applied time-series models (autoregressive integrated moving-average [ARIMA] model, autoregressive model, and moving-average model) to forecast the bitcoin price. They found that the ARIMA model was the best model to predict the bitcoin price. Guo and Antulov-Fantulin (2018) collected data about volatility and order book related to bitcoin over the period September 2015 to April 2017 and proposed temporal mixture models to predict the changes in bitcoin price. The proposed models worked better than other models in predicting the changes in the price of bitcoin. Abu Bakar, Rosbi, and Uzaki (2019) used a moving-average method to predict the bitcoin price. Data collected from October 1 until December 20, 2019, was used in experiments. The moving-average forecasting method was implemented using 2-day, 3-day, 4-day, and 7-day calculations. The results revealed that the 2-day moving-average method is the better prediction method with the lowest mean absolute error (MAE) percentage for all observation periods.

Akcora, Dey, Gel, and Kantarcioglu (2018) extracted bitcoin price data over the period 2009–2018 to predict bitcoin prices. They introduced a new concept of *k*-chainlets on bitcoins that enlarges the ideas of motifs and graphlets to blockchain graphs. They used chainlets or bitcoin subgraphs to evaluate the topological structure over time. They developed an approach to comprehend chainlets and for local topological structure and used the techniques that have a greater effect on the price dynamics. These important chainlets are used for price prediction. The bitcoin transaction graph possesses three main components: addresses, blocks, and transactions. Bitcoins transferred from an input address to an output address are said to be one transaction. Figure 8 shows 13 addresses and four transactions in a network. Granger causality is used for bitcoin price prediction. Addresses are represented as circles, and transactions are shown as rectangles. An edge indicates a transfer of coins. The coin at address a8 represents unspent coin. Chainlet analysis provides a more in-depth insight into local topological properties of the blockchain and the role of those local higher order topologies in the bitcoin price formation. They found that specific types of chainlets have a high predictive utility for bitcoin prices. Moreover, extreme chainlets exhibit a vital role in the bitcoin price prediction.

Bhambhwani, Delikouras, and Korniotis (2019) extracted data from August 2015 to January 2019 to investigate the fundamental drivers of cryptocurrency (bitcoin, ethereum, monero, litecoin, and dash) prices using the dynamic ordinary least-squares method and found that the prices of these currencies depend on their computing power and network.

Bystrom and Krygier (2018) extracted daily, weekly, and monthly data covering from 2011 to 2017 to investigate correlations, regressions, vector autoregression (VAR), and impulse response functions techniques. The variables driving the changes in bitcoin were examined, and they found that these variables change with volatility of the trade-weighted USD currency index and search pressures on bitcoin-related words on Google. Kaya (2018) used correlation and regression analyses and extracted weekly data from August 8, 2014, to May 4, 2018, from the online cryptocurrency market Bitstamp to study the impact of public interest, volatility metric of S&P 500 Index options, and political and regulatory news on cryptocurrency prices. The results revealed that public interest is the most influential factor that drives cryptocurrency prices.

Kjærland, Khazal, Krogstad, Nordstrøm, and Oust (2018) used econometric methods represented by an autoregressive distributed lag model and the GARCH model to study the determinants of bitcoin (BTC) price dynamics. The data used were daily spot rates for BTC/USD for the period between January 1, 2013, and February 20, 2018. The results revealed that returns on the S&P 500 are essential in explaining the BTC price dynamics. Phillips and Gorse (2018a) used the wavelet coherence approach to investigate the co-movement between a cryptocurrency price and its related factors represented by social media factors, Google search volume, and Wikipedia. Cryptocurrency price data collected from a number of exchanges over the period 2010–2017, from social media factors derived from Reddit, from Google search volume from Google Trends service, and from Wikipedia to track the number of new users learning about a cryptocurrency were used for analysis. The results revealed a positive relationship between a cryptocurrency price and its related factors.

Wiedmer (2018) investigated the determinants of cryptocurrency price using a panel of 17 cross-sections. He employed unit-root and cointegration tests and estimated the effects with vector error correction models, dynamic ordinary least squares and fully modified ordinary least squares. Causality flows were tested by weak exogeneity and Granger causality tests. The results showed that Metcalfe's law, community factors, and search engine queries had an influential impact on the price of the cryptocurrency. Blau (2017) extracted price data from Bitcoin Charts and Bloomberg over the period July 17, 2010, to June 1, 2014, and used regression analysis to investigate the impact of speculative trading, the prior 5-day bitcoin return, the prior 5-day volume turnover, outstanding bitcoins, and the volatility estimate for the exchange rates of 51 other currencies variables on the price of bitcoin. The results revealed that speculative trading is irrelevant in explaining changes in the price of bitcoin. Hayes (2017) used cross-sectional data about 66 of the most widely used cryptocurrencies to investigate the impact of three factors on the cryptocurrency value: the level of competition in the network of producers, the rate of unit production, and the difficulty of the algorithm used to "mine" for the cryptocurrency. The results revealed that the three factors have a significant impact on the cryptocurrency value.

TABLE 3 Summary of traditional statistical and econometric techniques surveyed in Section 4

Technique	Cryptocurrency	Reference	Data type	Data frequency	Time range	Target variables	Performance metric	Data source
ADF unit-root test and bound testing approach	Bitcoin, ethereum, dash, litecoin, and monero	Sovbetov (2018)	Market data	Weekly data	2010–2018	Market beta, trading volume, and volatility	Crypto50 index price	BitInfoCharts, Finance, World Bank, and Google Trends
ARIMA, AR, and MA models	Bitcoin	Roy et al., (2018)	Market data	Daily	July 2013–August 2017	Price prediction	Accuracy	CoinDesk
Temporal mixture model	Bitcoin	Guo and Antulov-Fantulin (2018)	Order book data	Hourly volatility	Sep 2015–April 2017	Prediction of volatility	Accuracy	They are not mentioned
k-chainlets	Bitcoin	Akcora et al., (2018)	Market data	Daily	2009–2018	Prediction of price performance	RMSE, wallet gain	Bitcoin core
MA method	Bitcoin	Abu Bakar et al., (2019)	Cash data	Daily	October 1–December 20, 2019	Price prediction	Mean absolute error percentage	CoinDesk
Vector error correction	Bitcoin	Abbatemarco et al., (2018)	Market data	Daily	November 2013–Septemb 2017	Price prediction	Cost and revenues databases	Publicly
Dynamic ordinary least squares method	Bitcoin, ethereum, dash, litecoin, monero, and dash	Bhambhwani et al., (2019)	Market data	Weekly average	August 7, 2015–January 25, 2019	Prediction of price	Least squares method	Coinmetrics
Correlation network and VAR model	Bitcoin	Giudici and Abu-Hashish (2019)	Market data closing price	Daily	May 18, 2016–April 30, 2018	Price prediction	RMSE full, RMSE autoreg	cryptocoincharts.info
FCVAR model	Bitcoin	Dos Santos Maciel and Ballini (2019)	High and low bitcoin prices	Daily	January 2012–February 2018	Price prediction	Range (i.e., measure of realized volatility)	coindesk.com
MF-DCCA	Bitcoin	Kim et al., (2016)	Daily volume and price index data	Daily	July 17, 2010–May 2, 2018	Price–volume cross-correlatio	Cross-correlations coefficient	cryptocompare
Kapetanios unit-root test, Maki cointegration analysis and Markov regime switching regression analysis	Bitcoin, ethereum, litecoin, and ripple	Gunay (2019)	Market capitalization closing price, public information arrivals in bull and bear markets	Daily	August 7, 2015–January 13, 2018	Impact of public information arrivals on cryptocurrency market via Twitter posts (causality)	log price, Maki cointegration test, Granger causality	CoinMarketCap
ARIMA model	Bitcoin	Anupriya and Garg (2018)	Open, low, and high close price	Daily	January 1, 2015–Septemb 23, 2018	Seasonality and trend present in the price data	Accuracy	coindesk.com
Random matrix theory and minimum spanning trees	N = 119 cryptocurrencies ranked by market capitalization	Stosic et al., (2018)	Market data (closing price)	Daily	August 26, 2016–January 18, 2018	Cross-correlations between price changes of different cryptocurrencies	Cross-correlations	CoinMarketCap
GARCH–MIDAS framework	Bitcoin, etherium, litecoin, ripple, and stellar, the cryptocurrency index CRIX	Walther et al., (2019)	Explanatory variable (global real economic activity)	Daily, weekly, and monthly	May 1, 2013–July 31, 2019	Prediction of volatility in cryptocurrency markets	HMSE	CoinMarketCap, CRIX from thecrix.de (Trimborn & Härdle, 2019)
Correlation in DFA and DCCA	Bitcoin, etherium, ripple, and litecoin	Alvarez-Ramirez et al., (2018)	Market data (market capitalization and availability)	Sliding windows approach	August 2015–April 2019	Analyze major cryptocurrencies' prices and prediction	Confidence band	CoinMarketCap

Table 3 continued on next page

TABLE 3 Continued

Technique	Cryptocurrency	Reference	Data type	Data frequency	Time range	Target variables	Performance metric	Data source
Markov-switching GARCH models	Bitcoin, etherium, ripple, and litecoin	Caporale and Zekokh (2019)	Closing prices	Daily	Different periods with end date as April 30, 2018	Modeling volatility, prediction of VaR and ES	*P*-values	CoinMarketCap, CoinDesk price index
ARIMA	Bitcoin, XRP, and ethereum	Alahmari (2019)	Daily, weekly, and monthly	Time-series data	2013–2018	Price prediction	MAE, MSE, RMSE	CoinMarketCap
Natural language processing techniques, lexicon-based sentiment analyzer	Bitcoin	Karalevicius et al., (2018)	Bitcoin news media on investor sentiment	Daily	Till February 2016	Price movement prediction	Returns and Sharpe ratios	Expert news media, CoinDesk, Cointelegraph, NewsBTC
Correlations, regressions, VAR, and impulse response	Bitcoin	Bystrom and Krygier (2018)	Market data	Daily, weekly, and monthly data	2011–2017	Prediction of volatility	RMSE, PSE, QL loss function and the R^2LOG loss function	Luxembourg-based bitcoin exchange Bitstamp
Correlation and regression analysis	Bitcoin	Kaya (2018)	Financial data	Weekly data	August 8, 2014–May 4, 2018	Price movement	Coefficients, *SE*, *t* stat, *P*-values	Bitstamp
Autoregressive distributed lag model and generalized autoregressive conditional heteroscedasticity model	Bitcoin	Kjærland et al., (2018)	Sentiment-based data	Daily	January 1, 2013, and February 20, 2018	Determinants of BTC price dynamics	Variance inflation factors	Quandl
Wavelet coherence approach	Bitcoin, ethereum, monero	Phillips and Gorse (2018a)	Sentiment-based data	Daily	2010–2017	Wavelet coherence analysis of price	*P*-values	Brave New Coin
Vector error correction models, dynamic OLS, and fully modified OLS	Bitcoin and others	Wiedmer (2018)	Market data	Daily	September 2017–January 2018	Volume-weighted average of prices from each market	Unit root and cointegration tests	Marketcap
Regression analysis	Bitcoin	Blau (2017)	Financial and tecnical data	Daily	July 17, 2010–June 1, 2014	Volatility estimation	Correlation matrix	Bitcoincharts and Bloomberg
Regression model	Bitcoin	Hayes (2017)	Data of 66 cryptocurrencies	Daily	2010–2013	Cost of production	*SE* and *t*-statistic	CoinMarketCap, coinwarz.com, cryptsy.com, bitcoinwisdom.com, and blockchain.info
Correlations, regression, and ANOVA	Bitcoin	Vaddepalli and Antoney (2018)	Market data	Daily	2014–2016	Impact of macroeconomic factors on bitcoin price	*SE*	Fiatleak, Cryptocomp, and BitCoiny
Regression analysis	Bitcoin	Vieira (2017)	Market data	Daily	November 1, 2013–January 20, 2016	Asymmetrical impacts on price volatility	test-stat, *P*-value	Quandle and US Department of Treasury

Table 3 continued on next page

TABLE 3 Continued

Technique	Cryptocurrency	Reference	Data type	Data frequency	Time range	Target variables	Performance metric	Data source
Wavelet coherence analysis	Bitcoin	Krištoufek (2015)	Market data	Daily	2011–2014	Trade, money supply and price level	Total circulating bitcoins, transactions no., difficulty, est. output vol., trade vol. vs. transaction vol., rash rate	Bitfinex, Bitstamp, and BTC-e
GARCH–MIDAS model	Bitcoin	Conrad et al., (2018)	Financial and macroeconomic	Daily	May 2013–Decemb 2017	Long- and short-term volatility	Returns on the S&P 500	Bitcoinity
GARCH and GAS	Bitcoin	Troster et al., (2019)	Market data	Daily	July 19, 2010–April 16, 2018	Bitcoin returns and risk	RMSE	Coindesk
Copula quantile causality	Bitcoin, Ethereum, and other 5	Bouri et al., (2019)	Detrended volume data	Daily	January 1, 2013–Decemb 31, 2017	Prediction of price volatility, return		
Granger-causality method and ARCH	Bitcoin	Badenhorst (2018)	Market data	Daily	2014–2018	Volatility prediction, return	*SD*, skewness, min and max	CoinMarketCap
MSGARCH model	Bitcoin	Ardia et al., (2019)	Bitcoin mid prices	Daily	August 18, 2011–March 3, 2018	Volatility dynamics of bitcoin's log returns	Mean, median, *SD*, skewness, kurtosis, VaR forecasting	Datastream
Four GARCH-type models	Bitcoin, dash, litecoin, and ripple	Charles and Darné (2019)	Closing price	Daily	June 1, 2014–November 11, 2018	Breakpoints and sudden shifts in volatility	Jumps and structural breaks	CoinMarketCap
Markowitz mean–variance framework	500 most capitalized cryptocurrencies	Brauneis and Mestel (2019)	Market data	Daily	January 2015–Decemb 2017	Risk–return benefits	Mean and variance	CoinMarketCap
Fractional cointegration technique in VAR set-up	13 popular coins, including bitcoin	Yaya et al., (2018)	Price data	Daily	August 7, 2015–November 28, 2018	Volatility trends	Dependency of the paired variables	Coin Metrics
HMM and SIR method	Bitcoin and other 3	Phillips and Gorse (2017)	Daily closing price and volume	Daily	April 2015–Septemb 2016	Price prediction and bubbles	Sharpe ratio, Sortino ratio, returns	Crypto Compare
Dynamic topic model and Hawkes model	Bitcoin and ethereum	Phillips and Gorse (2018b)	Social media data and price data	Daily	August 8, 2016–August 30, 2017	Future price movement prediction	Event-based data analysis	Reddit
Sentiment analysis, ARIM	Cryptocurrencies	Zamuda et al., (2019)	Social media data	—	—	Crypto forecasting and investment	New sentiment analysis indicators	—
Granger-causality method	Bitcoin, ethereum	Bartolucci et al., (2020)	Sentiment and market data	—	December 2010–August 2017	Price prediction	Mean, *SD*, min–max values	Github, CoinMarketCap

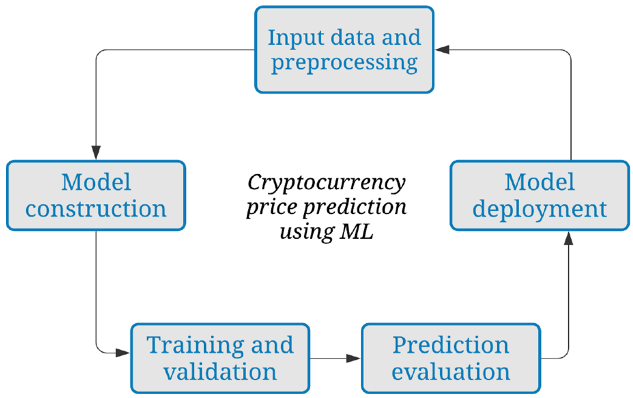


FIGURE 9 Illustration of ML model in cryptocurrency price prediction

Vaddepalli and Antoney (2018) examined the impact of financial openness, internet penetration, and inflation on the change of the price of bitcoin of United States, Canada, Russia, Brazil, China, and South Africa. Data collected from World Bank databases, Fiatleak, and Cryptocompare and BitCoiny over the period 2014–2016 and 15-years' average values for the explanatory variables. The results of correlations, regression, and analysis of variance indicated that the variables analyzed do not significantly affect the price of bitcoin. Vieira (2017) used regression analysis on a set of time-series data spanning the period from November 1, 2013, to January 20, 2016, to investigate the impact of the following variables on bitcoin price: S&P 500 index; daily Treasury real yield curve rates on "Treasury Inflation-Protected Securities" for a fixed maturity of 7 years; daily USD price per ounce of gold; the daily number of confirmed bitcoin transactions; the total number of unique addresses used on the bitcoin blockchain; total value of coinbase block rewards and transaction fees paid to miners and the daily number of the term 'Bitcoin' queries made in Wikipedia. The results showed that only the index for S&P 500 is irrelevant to explain the changes in bitcoin prices.

Kristoufek (2015), using data over the period from September 14, 2011, to February 28, 2014, employed wavelet coherence analysis to examine the impact of usage in trade, money supply, and price level on bitcoin price over the long term. The results showed that these factors have an influential impact on bitcoin price. Abbatemarco, Maria De Rossi, and Salviotti (2018) used an econometric model represented by the augmented Dicky–Fuller test, the Johansen tests for cointegration and a vector error correction model to estimate the bitcoin price based on the price data available over the period from November 11, 2013, to September 5, 2017. The results supported the validity of the suggested model to estimate the price of bitcoin.

Giudici and Abu-Hashish (2019) introduced partial correlations and correlation networks into VAR models. The model helped to determine the cryptocurrency prices dynamics in different crypto exchange markets and allowed one to understand its correlation with other traditional market prices. The application of VAR correlation networks also enabled building a model for predicting bitcoin price that leverages the information contained in different correlation patterns among various exchange prices. Dos Santos Maciel and Ballini (2019) analyzed the dynamic behavior and predictability of daily dynamics (high and low) of bitcoin prices. They applied the fractionally cointegrated vector autoregressive (FCVAR) model to understand the price patterns of bitcoin and dollar. The empirical analysis was performed for the period from January 2012 to February 2018. They made comparisons with various other algorithms, and the results indicated that fractionally cointegrated VAR performed better.

Kim et al., (2016) used multifractal detrended cross-correlations analysis to determine price–volume cross-correlation from July 2010 to May 2018 in the bitcoin market. A cross-correlation test was performed to evaluate the cross-correlations between the series. The level of cross-correlations was quantified using the detrended cross-correlations analysis coefficient. Changes in bitcoin price and trading volume mutually interact with each other in a nonlinear way that may assist the participants of the bitcoin market in boosting profit. Multifractality in cross-correlations presents the turbulent and dynamic characteristics of the bitcoin market. Public information arrival is the most essential asset of an efficient market.

Gunay (2019) explored the impact of public information arrival (official Twitter announcements) on the cryptocurrency market (bitcoin, ethereum, litecoin, and ripple) using various methods, such as Maki cointegration analysis, Markov regime-switching regression analysis and the Kapetanios unit-root test, to study its impact. The results indicated that positive public information arrival has a positive impact on ripple's value. All models exhibited same results with bitcoin as an independent variable, indicating its significant effect on ripple's value. The results were evaluated under bull and bear markets. In the bull market, public information arrival had a positive impact on ripple's value, whereas the bear market did not have enough power to divert the ripple value from a downward trend.

An ARIMA model was used by Anupriya and Garg (2018) for predicting bitcoin price. This model is recommended as it gives better results than predicting directly. This model shows good accuracy, with a mean percentage error of less than 6%. Cross-correlations between price fluctuations of various cryptocurrencies were analyzed using minimum spanning trees and random matrix theory by Stosic, Stosic, Ludermir, and Stosic (2018). To quantify correlations, change in price or cryptocurrency return value were calculated. The results showed the multiple collective behaviors in the crypto market that can be appropriate in constructing cryptocurrency investment portfolios.

Several econometrics methods in time-series research, such as GARCH and the Baba–Engle–Kraft–Kroner model, have been used in the literature on cryptocurrency research. Walther, Klein, and Bouri (2019) used the GARCH–mixed data sampling (MIDAS) framework was used to predict the volatility of cryptocurrencies (bitcoin, litecoin, etherium, stellar and ripple) and the cryptocurrency index CRIX. They determined most exogenous drivers and found that global real economic activity gave the most accurate predictions. Heteroscedasticity-adjusted mean squared error was used to analyze the forecast. Future work aimed to address the same issue with some other methodology or by using intra-day data to construct daily measures of cryptocurrency volatility.

Alvarez-Ramirez, Rodriguez, and Ibarra-Valdez (2018) adopted detrended fluctuation analysis over a sliding window to evaluate long-range correlations for bitcoin price returns, Caporale and Zekokh (2019) used Markov-switching GARCH (MSGARCH) models to estimate the changes in the cryptocurrencies (bitcoin, ethereum, ripple and litecoin). The result shows that using the standard GARCH model may yield incorrect predictions resulting in ineffective risk management. This can be improved by using regime-switching, which might be useful for both regulators and investors. An ARIMA model was presented by Alahmari (2019) to predict three major cryptocurrency prices (bitcoin, XRP, and ethereum) on the basis of daily, weekly, and monthly time series. The ARIMA model outperformed other models in terms of mean squared error (MSE), MAE, and root-mean-squared error (RMSE).

The influence on media sentiment and bitcoin price was studied by Karalevicius, Degrande, and De Weerdt (2018). Natural language-processing methods were used for data preprocessing. Lexicon-based sentiment analysis methods were integrated with Harvard Psychosocial and finance-industry-specific dictionaries to quantify sentiments and to inspect the reaction patterns. Reaction patterns were analyzed, which indicated that the price results following noteworthy news stories and then interday trading strategy is proposed to maximize the return. The return and Sharpe ratio values were examined, which showed that sentiment-driven technique outperformed other bitcoin strategies.

Conrad, Custovic, and Ghysels (2018) extracted the long and short-term volatility components of bitcoin using a GARCH–MIDAS model. This model converted the conditional variance into low- and high-frequency components. The results findings were that the volatility has a negative or significant impact on long-term bitcoin volatility and that volatility risk premium has a significant positive impact on long-term bitcoin volatility.

In Troster, Tiwari, Shahbaz, and Macedo (2019), GARCH and generalized autoregressive score (GAS) models were used to predict bitcoin returns and risks. Out-of-sample performance was compared for both of these models. The results showed that the GAS model with heavy-tailed distribution provided the best out-of-sample prediction and, owing to its flexibility, the GAS model was more robust. Bouri, Lau, Lucey, and Roubaud (2019) applied a copula-quantile causality approach on volatility of cryptocurrencies. The strategy of the experiment extended the copula–Granger-causality in distribution (CGCD) method of Lee and Yang (2014). The study used copula functions to construct two tests of CGCD. A parametric test employed six parametric copula functions to discover dependency density between variables. The performance matrix of these functions varied with independent copula density. The results provided remarkable evidence of Granger causality from trading volume to the returns of seven large cryptocurrencies.

The work of Badenhorst (2018) focused on revealing whether spot and derivative market volumes cause volatility in bitcoin price. He used the Granger-causality method and $ARCH(1,\ 1)$ for analysis. The result provided evidence that the spot trading volumes have a significant positive effect on price volatility, whereas the relationship between cryptocurrency volatility and the derivative market is uncertain.

An MSGARCH model was applied by Ardia, Bluteau, and Rüede (2019) to test the existence of institutional changes in the GARCH volatility dynamics of bitcoin's logarithmic returns. The results showed that MSGARCH models clearly outperform single-regime GARCH for value-at-risk forecasting.

The findings from Charles and Darné (2019), who examined four cryptocurrencies (bitcoin, dash, litecoin, and ripple), showed cryptocurrency returns are strongly characterized by the presence of jumps and structural breaks except for the dash market. Four GARCH-type models (i.e., GARCH, integrated GARCH, asymmetric power autoregressive conditional heteroscedasticity, and fractionally integrated GARCH) and three return types with structural breaks (original returns, jump-filtered returns, and jump-filtered returns with structural breaks) were considered. Their findings indicated the importance of jumps in cryptocurrency volatility and structural breakthroughs.

Brauneis and Mestel (2019) adopted the use of a Markowitz mean-variance framework to analyze the risk–return benefits of cryptocurrency portfolios. In an out-of-sample analysis accounting for transaction cost, they found that combining cryptocurrencies enriches the set of low-risk cryptocurrency investment opportunities.

Yaya, Ogbonna, and Olubusoye (2018) applied the fractional cointegration technique in VAR set-up to identify the dependence and persistence of bitcoin on other popular alternatives before and after the 2017–2018 crash in cryptocurrency markets. The research focus was necessitated since market players fouind bitcoin to drive other cryptocurrencies, though bitcoin is the most valuable and highly capitalized coin, taking about 40% of the 2,074 cryptocurrency types market share (https://cointelegraph.com). They included 13 highly priced and data-available cryptocurrencies in their analysis. The results revealed that higher persistence of shocks is expected after the crash due to speculation in the minds of cryptocurrency traders, and more evidences of non-mean reversions, implying chances of further price falls in cryptocurrencies.

Phillips and Gorse (2017) applied a hidden Markov model and the superiority and inferiority ranking approach to analyze the bubble-like behavior in cryptocurrency time series. Considering the hidden Markov model and the superiority and inferiority ranking method, an epidemic detection mechanism was used in social media to predict cryptocurrency price bubbles, which classify bubbles through epidemic and nonepidemic labels. Experiments demonstrated a strong relationship between Reddit usage and cryptocurrency prices.

Phillips and Gorse (2018b) applied a dynamic topic model and Hawkes model to decipher relationships between topics and cryptocurrency price movements. The authors used a latent Dirichlet allocation model for topic modeling, which assumes each document contains multiple topics to different extents. The experiment showed that particular topics tend to precede certain types of price movements in the cryptocurrency market.

Zamuda et al., (2019) used new sentiment analysis indicators to analyze cryptocurrency trends. A general model evaluating the influence between a user's network action–reaction influence model is mentioned in this research. The research covered the different aspects of necessary perspectives needed when preparing forecasting and investment, supported by cryptocurrency social media sentiment analysis.

Bartolucci et al., (2020) examined cryptocurrency prices with the "butterfly effect," which means "issues" of open-source project provides insights to improve prediction of cryptocurrency prices. Sentiment, politeness, emotions analysis of GitHub

TABLE 4 Summary of ML techniques surveyed in Section 5

Technique	Cryptocurrency	Reference	Data type	Data frequency	Time range	Target variables	Performance metric	Data source
Binary autoregressive tree	Bitcoin, ripple, and ethereum	Derbentsev et al., (2019)	Time-series data in log return	Daily	January 1, 2017–January 3, 2019	Short-term forecasting of cryptocurrencies' prices	RMSE	Yahoo! Finance
NN, switching regression model	STORJ	Chakraborty and Roy (2019)	STORJ token prices and related clustering coefficients of the transaction network	Daily	July 2, 2017–April 6, 2018	Forecast performance of various time-series models	MAE, ME, MAPE, RMSE, and RMPE	Not mentioned
Linear regression, logistic regression, ANN, SVM	Bitcoin	Greaves and Au (2015)	Bitcoin transactions	Bitcoin price at 15 s intervals	Prior to April 7, 2013	Bitcoin blockchain analysis	MSE, accuracy	CS224W website, bitcoincharts.com
Extreme gradient boosting regression tree model (XGBoost)	ZClassic, ZCash, and bitcoin private	Li et al., (2019)	Sentiment-based on Twitter and trading volume	Hourly pricing data	For a period of 3.5 weeks	Predicting	Correlation coefficients, adjusted *SD*	RStudio, rtweet package
Linear regression, polynomial regression, RNN, and LSTM-based analysis	Bitcoin	Mittal et al., (2019)	Social media and web search data	Average daily price	April 9, 2014–January 7, 2019	Short-term bitcoin price fluctuation	R^2 score, correlation coefficient	bitcoincharts
Bayesian structural time-series approach	Bitcoin	Poyser (2019)	Website data	Daily	January 2013–May 2017	The association between bitcoin's market price and a set of internal and external factors	RMSE	Block information service
Time-series analysis using bidirectional LSTM	Bitcoin and others	Mohanty et al., (2018)	Market sentiment and social sentiment	Daily data like price, additional 26 features about the blockchain of bitcoin and market, interval data	February 2016–December 2018	Prediction of fluctuation in the future price of cryptocurrencies	—	Users' comments and tweets from Twitter using Apache Flume, and price data were fetched from exchange (Blockchain Info)
Logistic regression	Bitcoin, ripple, ethereum, litecoin, nem, dash, and stellar	Bouri, Shahzad, and Roubaud (2019)	Market data	Daily	August 7, 2015–December 31, 2017	Date-stamped price explosiveness	—	CoinMarketCap
Linear regression, multiple linear regression, multilayer perceptron NN, and long short-term memory NN	Bitcoin	Uras et al., (2020)	Stock market data	Daily	November 2015–August 2018	Price prediction	Relative RMSE, MAPE	Yahoo! Finance website and CoinMarketCap
Logistic regression, naive Bayes, SVM, RF, ARIMA, and RNN	Ethereum	Chen et al., (2017)	Market data	Hourly	August 30, 2015, and December 2, 2017	Predict price changes	Prediction	Not mentioned
ARIMA, ARMA-GARCH, VAR, α-Sutte indicator, and NNAR models	Ethereum	Bush and Choi (2019)	Network transaction data	h = 5, 20, and 50 days	July 2, 2017–March 17, 2018	Impact of ethereum STORJ token clustering coefficients in forecasting	RMSE, MAE, MAPE	coingecko.com, coinmarketcap.com
ANFIS and ANN	Bitcoin and others	Atsalakis et al., (2019)	Historical time series data	Daily	September 13, 2011–October 12, 2017	Price forecast	RMSE, MSE, and MAE	bitcoincharts.com
Hybrid model of hidden Markov models and optimized LSTM networks	Bitcoin	Hashish et al., (2019)	Market data: orders and trades, technical indicators	2 min frequency	August 20–September 20, 2018	Price prediction	Mean, *SD*	Coinbase exchange market
BPNN, GANN, GABPNN, NEAT	Bitcoin	Radityo et al., (2017)	Historical bitcoin price	Daily	June 10, 2013–February 4, 2017	Predict the close value of Bitcoin in the next day	MAPE	cryptocompare.com
ANN with back-propagation	Bitcoin	Sovia et al., (2019)	Open, high, low, volume, and request	Daily close	Not mentioned	Bitcoin prices for the next hour	Not mentioned	Not mentioned

Table 4 continued on next page

TABLE 4 Summary of ML techniques surveyed in Section 5

Technique	Cryptocurrency	Reference	Data type	Data frequency	Time range	Target variables	Performance metric	Data source
Feedforward ANN with back-propagation	Bitcoin, bitcoin cash, dash	Almasri and Arslan (2018)	Market price	Hourly	—	Daily close price and hourly close price	—	—
NN, SVM RF	Bitcoin, ethereum, ripple, and litecoin	Valencia et al., (2019)	Market and social data	Daily	80 days of historical data	Daily price	Price movement	cryptocompare.com, Twitter
Bayesian NN	Bitcoin	Jang and Lee (2018)	Response variable, blockchain information	Daily	September 2011–August 2017	Bitcoin price	RMSE, MAPE	bitcoincharts.com, blockchain.info
PNN, SVM	Bitcoin, ripple, ethereum	Kim et al., (2016)	User comment data	Daily	December 1, 2013–November 10, 2015	Predict the price and the number of transactions of cryptocurrencies	Accuracy, F_1 score, MCC	CoinMarketCap, CoinDesk
ANN, SVM, ensemble techniques	Bitcoin	Mallqui and Fernandes (2019)	Internal (bitcoin behavior) and external (economic factors, demand)	Daily	August 19, 2013–July 19, 2016, April 2013–2017	Min–max and closing price	MAE, MAPE, RMSE	bitcoincharts.com, quandl.com, investing.com
ANN	Bitcoin	Almeida et al., (2015)	Historical data	Previous day's price and volume	Since 2009	Bitcoin trend	MSE and profit analysis	quandl.com
Linear regression, SVM	Ethereum	Poongodi et al., (2020)	Market price data	60 min intervals	—	Closing price and lowest price	Percentage accuracy	etherchain.org
Optimized SVM–PSO	Bitcoin, litecoin, ethereum, ripple, nem, and stellar	Hitam et al., (2019)	Market price data	Daily	Bitcoin, training: March 28, 2013–January 16, 2017; testing: January 17, 2017–January 16, 2018	Open price, close price, high price, low price	Accuracy	—
Multivariate linear regression	Bitcoin, exclusive coin, ripple, litecoin, lisk, monero, siacoin, cryptonex, AdEx, guldencoin	Mittal et al., (2018)	Cryptocoin historic price data set	Daily	2015, 2016, and 2017	Daily price	RSE, multiple R^2, F-statistic, adjusted R^2, P-value	Kaggle
Multiple linear regression model	Bitcoin and litecoin	Jain et al., (2018)	Concurrent price data, tweets	Per minute	Bitcoin: March 2018. Litecoin: February, March 2018	2 hr price	R^2 score	CoinDesk, Twitter
XGBoost, quadratic discriminant analysis, RF, LSTM, SVM	Bitcoin	Chen et al., (2020)	Aggregated price	Daily, 5 min interval	Dataset 1: February 2, 2017–February 1, 2018. Dataset 2: July 17, 2017–January 17, 2018	5 min interval price, daily price	Accuracy, precision, recall, F_1 score	CoinMarketCap
Bayesian regression	Bitcoin	Shah and Zhang (2014)	Market price	10 s intervals	February 2014–July 2014	Price variation	Profit	Okcoin.com
Binomial logistic regression, SVM, RF, binomial GLM	Bitcoin	Madan et al., (2015)	Market price data	Daily, 10 min, 10 s	—	Price change using 10 min time intervals	Sensitivity, specificity, precision, accuracy	CoinBase, Okcoin
SVM and ANN	Bitcoin	Silva de Souza et al., (2019)	Market price, gold and silver data	Daily	Bitcoin: May 2012–May 2017. Gold and silver: April 2012–May 2017	Gold, silver, and bitcoin prices	Risk-adjusted returns	Bloomberg
RF, SVM, GB, and linear regression	Bitcoin	Virk (2017)	Date, open, high, low, close, volume, market capital	Interval data	May 2013–May 2017	Prediction and correlaton	Accuracy, precision, recall, and F_1-score	Kaggle
SVM, naive Bayes, and RF	Bitcoin	Sun et al., (2019)	Historical prices	Daily	January 2011–December 2018	Next-day price trends, maximum and minimum daily variations	Prediction accuracy	—
RFs with factors in Alpha101	Bitcoin and others	Barnwal et al., (2019)	Market history data	1 min, 5 min, 30 min, 1 hr, and 1 day	August 2017–December 2018	Price prediction	Accuracy	Bitfinex, Binance

Table 4 continued on next page

comments are applied in ethereum and bitcoin markets. The results showed that these metrics have predictive power on cryptocurrency prices.

Table 3 summarizes the traditional statistical and econometric approaches for cryptocurrency price prediction. It is easily evident from this table that a lot of researchers within the

TABLE 4 Continued from previous page

Technique	Cryptocurrency	Reference	Data type	Data frequency	Time range	Target variables	Performance metric	Data source
Generative and discriminative classifiers combined using a one-layer NN	Bitcoin	Attanasio et al. (2019)	Volume, volatility, trend, and momentum	Daily	August 2017–July 2018	Cryptocurrencies	Accuracy	quandl

ANFIS: adaptive neuro-fuzzy inference system; ANN: artificial neural network; ARIMA: autoregressive integrated moving-average; ARMA: autoregressive moving-average; BPNN: back-propagation neural network; GABPNN: genetic algorithm back-propagation neural network; GANN: genetic algorithm neural network; GARCH: generalized autoregressive conditional heteroscedasticity; GB: gradient boosting; GLM: generalized linear model; LSTM: long short-term memory; MAE: mean absolute error; MAPE: mean absolute percentage error; ME: mean error; MSE: mean square error; NEAT: neuroevolution of augmenting topologies; NN: neural network; NNAR: neural network autoregression; PNN: probabilistic neural network; PSO: particle swarm optimization; RF: random forest; RMPE: relative mean percentage error; RMSE: root-mean-square error; RNN: recurrent neural network; RSE: residual standard error; VAR: vector autoregression; SVM: support vector machine.

statistics/econometrics field have applied various schemes for cryptocurrency price prediction. Financial data, sentiment analysis, and various internal and external factors are used in the studies. The most common and popular cryptocurrency studied by researchers was bitcoin.

5 | ML TECHNIQUES FOR CRYPTOCURRENCY PRICE PREDICTION

ML refers to the automated process of learning from experience (Alon, Lokshtanov, & Saurabh, 2009). Automatic learning and adaptation with exposed data without the need of human intervention are the main emphases of ML. Rather than writing an explicit program to solve a task, in ML the computer learns from provided example data and comes up with its program (Domingos, 2012). Based on that, the computers emulate human behavioral learning towards particular decision-making and reasoning. Cryptocurrency is volatile in nature. This motivated researchers to apply DL and ML paradigms to cryptocurrency concerns. Adopting the stock market price prediction techniques can help in increasing the precision rate (El-Bannany et al., 2020; Lahmiri, 2011; Sreedharan et al., 2020a; 2020b). In recent years, ML is one of the most researched approaches in cryptocurrency price prediction because of its ability to identify the general trend and fluctuation. Figure 9 gives an illustration of using ML in cryptocurrency price prediction. Several ML techniques are applied in cryptocurrency price prediction. This includes classification, regression, DL, and RL models. The ML techniques used in cryptocurrency price prediction research are as discussed in the following. We have distinguished these schemes and have added a separate section specifically on DL and RL models because of their inherent variation and wide adoption. Some researchers have focused on the comparison of different classification and regression ML methods.

A short-term forecasting model was presented by Derbentsev, Datsenko, Stepanenko, and Bezkorovainyi (2019) to predict cryptocurrency prices of ripple, bitcoin, and ethereum using an ML approach. A binary autoregression tree was implemented in that paper that combines classification and ARIMA. Simulation results proved that this algorithm is more accurate and efficient than other traditional approaches are. Various ML algorithms were compared by Rane and Dhage (2019) to select an optimal technique for predicting bitcoin price. A survey of different ML techniques was presented to show which method best suited the prediction of bitcoin price. The analysis conducted will be further extended by improving the precision of less accurate techniques. STORJ tokens were presented in Chakraborty and Roy (2019) to make transactions in a STORJ network for predicting future price. Different time-series models were presented, including Box–Jenkins models, NN, and a switching regression model. The key advantage of using these technologies is that we can obtain the estimated equation for each regime with flexibility.

Greaves and Au (2015) applied blockchain data for bitcoin price prediction using an SVM, an ANN, linear regression, and logistic regression. An NN classifier with two hidden layers marked the highest price accuracy of 55%, followed by logistic regression and SVM. In addition, the research also mentions the analysis with several tree-based models and *K*-nearest neighbors. Limited predictability was observed in this research using only blockchain data for training and prediction. The research concluded that by using features directly extracted from bitcoin exchanges, like financial flow features, would likely improve the bitcoin price prediction accuracy.

Li et al., (2019) analyzed Twitter signals to predict price fluctuations using ZClassic. Tweets were collected on an hourly basis for 3.5 weeks and each tweet was classified as positive, negative, or neutral. Tweets were then compiled to create a weighted or unweighted index. The model was trained using an extreme GB regression tree model and compared with historic price data.

Mittal, Dhiman, Singh, and Prakash (2019) used ML techniques such as linear regression, polynomial regression, recurrent NN (RNN), and long short-term memory (LSTM)-based analysis to identify the correlation among bitcoin price and Twitter and Google search patterns. Among Google Trends, tweet volumes, and tweet sentiments, tweet sentiment analysis gives the worst outcome. When LSTM, RNN, and polynomial regression were applied on Google Trends and tweet volume, an improved accuracy in performance was shown.

Atsalakis, Atsalaki, Pasiouras, and Zopounidis (2019) used a hybrid neuro-fuzzy controller called PATSOS to predict the daily price change trend of bitcoin. The scheme outperformed two other computational intelligence models, the first being developed with a simpler neuro-fuzzy approach and the second being developed with ANNs. They also stated that the performance of the PATSOS system was robust to use for other cryptocurrencies.

TABLE 5 Summary of deep- and reinforcement-learning techniques surveyed in Section 5.1

Technique	Cryptocurrency	Reference	Data type	Data frequency	Time range	Target variables	Performance metric	Data source
MLP and LSTM	Bitcoin	Misnik et al., (2018)	Market data	Timestamp of given minute	Rate at 0 s and 59 s of given minute	Predict price	Precision	Kraken
LSTM and GRNN	Bitcoin, digital cash, and ripple	Lahmiri and Bekiros (2019)	Digital currencies	Daily	*Start:*bitcoin, July 16, 2010; digital cash, February 8, 2010; ripple, January 21, 2015. *End:*October 1, 2018	Price prediction	RMSE	Not mentioned
LSTM and RNN	Bitcoin	McNally et al., (2018)	Average of 5 major exchanges	Daily	August 19, 2013–July 19, 2016	Closing price	RMSE, accuracy, precision, sensitivity	Coindesk, Bitcoin Price Index
LSTM and RNN	Bitcoin and litecoin	Yao et al., (2018)	Market data	Minute-to-minute updates	January 2012–March 2018	Predict price	MASE, SMAPE, MAE, and RMSE	Marketcap
DNN, LSTM, CNN, deep residual network	Bitcoin	Ji et al., (2019)	Market data	Daily	29 November 29, 2011–December 31, 2018	Predict price	Accuracy, precision, sensitivity, recall, specificity, F_1 score	bitcoincharts
GBDT and LSTM	Bitcoin	Alessandretti et al., (2018)	Market data	Daily	November 2015–April 2018	Return on investment	Geometric mean and cumulative returns	CoinMarketCap
Multiple input LSTM combined with Black–Schole	Bitcoin	Li et al., (2019)	Social media data, market, blockchain statistics	Daily	August 8, 2015–November 17, 2018	Price volatility	RMSE	Yahoo! Finance
Theil–Sen, Huber regression, LSTM and GRU	Bitcoin	Phaladisailoed and Numnonda (2018)	Market data	Minute	January 1, 2012–January 8, 2018 to	Predict price	MSE and R^2	Bitstamp
LSTM, GB	Bitcoin, ethereum, ripple, bitcoincash, litecoin, dash, ethereum classic, fiat, Korean won	Kwon et al., (2019)	Time series data	10 min	June 9, 2017–May 8, 2018	Classification of the cryptocurrency price time-series data	F_1-score, recall, precision	Bithumb API
Neural network	Bitcoin	Jiang and Liang (2017)	Financial data	30 min	1 year in time span	Portfolio weights or portfolio management	Cumulative return	Poloniex
GBDT, SVM, and RF	42 different cryptocurrencies	Sun et al., (2020)	Market data	Daily	January 1, 2018–June 30, 2018	Price trend	Accuracy	Investing.com
Convolution LSTM	Bitcoin, dash, ether, litecoin, monero, ripple	Alonso-Monsalve et al., (2020)	Cryptocurrencies data	1 min	July 1, 2018–June 30, 2019	Price prediction of exchange rates	Accuracy	Cryptocompare.
Multiple linear regression, RF, and LSTM ML	Bitcoin	Snihovyi et al., (2018)	Cryptocurrency data	Daily	January 25, 2017–January 22, 2018	Prices variation	MSE and R^2	Open source
LSTM	Bitcoin, ripple, dash, and ltc	Altan et al., (2019)	Digital currency data	Daily	*Start:*BTC prediction, July 18, 2010; XRP, January 22, 2015; dash, February 14, 2014; ltc, August 24, 2016.*End:*March 28, 2019	Currency	MAE, RMSE, MAPE	Not mentioned

CNN: convolutional neural network; DNN: deep neural network; GB: gradient boosting; GBDT: gradient boosting decision tree; GRNN: generalized regression neural network; GRU: gated recurrent unit; LSTM: long short-term memory; MAE: mean absolute error; MAPE: mean absolute percentage error; MASE: mean absolute scaled error; ML: machine learning; MLP: multilayer perceptron; MSE: mean square error; RF: random forest; RMSE: root-mean-square error; RNN: recurrent neural network; SMAPE: symmetric mean absolute percentage error; SVM: support vector machine.

Almeida, Tata, Moser, and Smit (2015) predicted bitcoin's trend for the next day based on the previous day's price and volume using an ANN. They used the historical data from 2009 onwards to learn the behavior. Mohanty, Patel, Patel, and Roy (2018) used LSTM for bitcoin future price prediction, and Twitter data was used to predict public mood. In this method, some salient features from the blockchain were selected that had a major effect on demand and supply of bitcoin and then they used this to train a model that enhanced the predictive power of bitcoin price in the future. The model showed high accuracy and good precision.

Hashish, Forni, Andreotti, Facchinetti, and Darjani (2019) used hidden Markov models to describe historical movements of cryptocur-

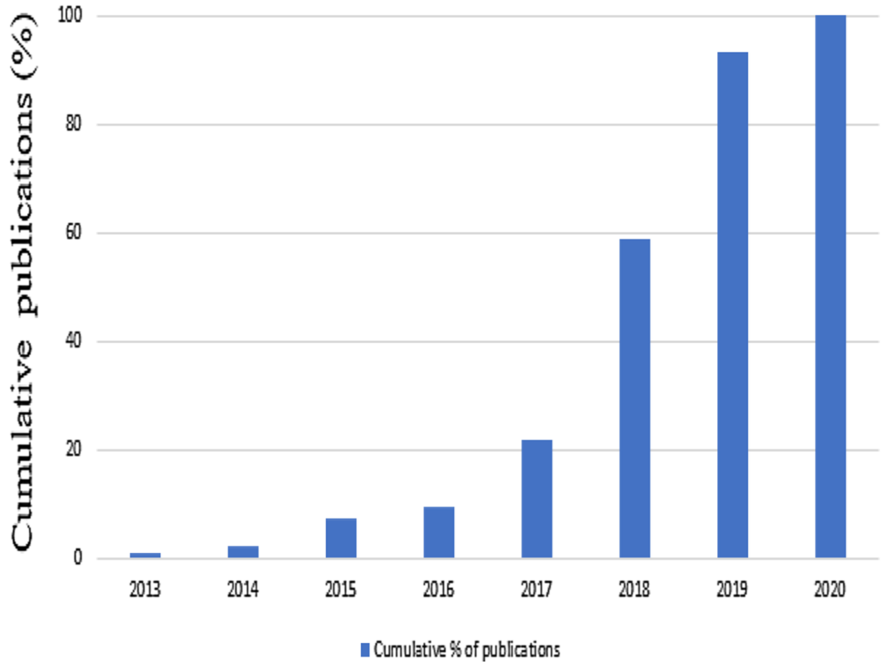


FIGURE 10 Percentage of publications (cumulative) versus year

TABLE 6 Paper distribution among categories

Paper category	Distribution (%)
Statistical/econometric	45.97
ML	54.03
	(ML: 70.22)
	(DL/RL: 29.78)

DL: deep learning; ML: machine learning; RL: reinforcement learning.

rencies and LSTM networks to predict future movements. A genetic algorithm was applied to further optimize the parameters of the hybrid approach. The simulation results showed the lowest MSE, RMSE and MAE compared with traditional models of time-series prediction, ARIMA, and conventional LSTM, which proved the effectiveness of the proposed approach. Internal details of bitcoin transactions were not accounted for in this model. So, for future work, they planned to consider additional features to provide more information about the blockchain.

Bush and Choi (2019) used ARIMA, ARMA-GARCH, VAR, alpha-Sutte indicator, and NN autoregression (NNAR) to forecast the ethereum STORJ token price. The dynamics of the model were evaluated using three time horizons (h = 5, 20, and 50 days). The performance of the model was analyzed using RMSE, mean absolute percentage error (MAPE), and mean absolute error (MAE). The simulation results showed that the VAR exceeded the other models in short- and mid-term horizons, whereas NNAR outperformeds for long-term forecasting. Radityo, Munajat, and Budi (2017) used the bitcoin exchange rate (closing price) on the American dollar for the next-day prediction using four variants of ANN. The methods selected for comparison were genetic algorithm NN, back-propagation NN (BPNN), genetic algorithm BPNN (GABPNN), and neuroevolution of augmenting topologies (NEAT). The performance of this method was compared depending on the accuracy (using MAPE) and complexity in terms of time required to construct a model. For each technique, generation of the model was done 30 times and the average value of MAPE and training time were calculated. Nine features were selected (open, low, high, volume, close, EMA 12, EMA 26, %R 5, and %R 14, 12, 26), for all these methods and for selection, a greedy forward-selection approach was used. The results showed that the GABPNN had greater accuracy, whereas the genetic algorithm NN was the worst, but the training time for the GABPNN is very long, which is why we cannot use this approach in real practice. So the next best candidate was the BPNN with 300 times faster accuracy time and slightly less accuracy compared with the GABPNN. The price movement of bitcoin was predicted using ANNs based on back-propagation algorithm in Sovia, Yanto, Budiman, Mayola, and Saputra (2019) using graph movements: open, low, and high, bitcoin requests, volumes, and next hour prediction prices. Prediction variables with target values can be predicted using prior bitcoin price. The best network patterns can be obtained that can be useful for the prediction process.

Almasri and Arslan (2018) applied a feedforward ANN with a back-propagation learning algorithm for training the network. Data were collected from Crypto compare API for prediction of day and hour close prices with the help of the Encog framework for ML. Bitcoin showed good results, but other coins showed fewer variations, which shows that a specific cryptocurrency needs a particular model to show good results. Hour close-price was more stable than the day close-price. Data-mining techniques might help in future. Various ML algorithms were also compared by Valencia, Gómez-Espinosa, and Valdés-Aguirre (2019), such as NN, RF, and SVM, to predict price movements of cryptocurrencies with highest market capitalization, which included bitcoin, ripple, ethereum, and litecoin. These. Three approaches were used to train a model. The first approach used social data to train a model, the second approach used market data, and the third one utilized both market and social data for training. To evaluate the robustness of each model, accuracy, precision, recall, and F_1 scores were used. Comparison results showed that the NN outperformed the other models.

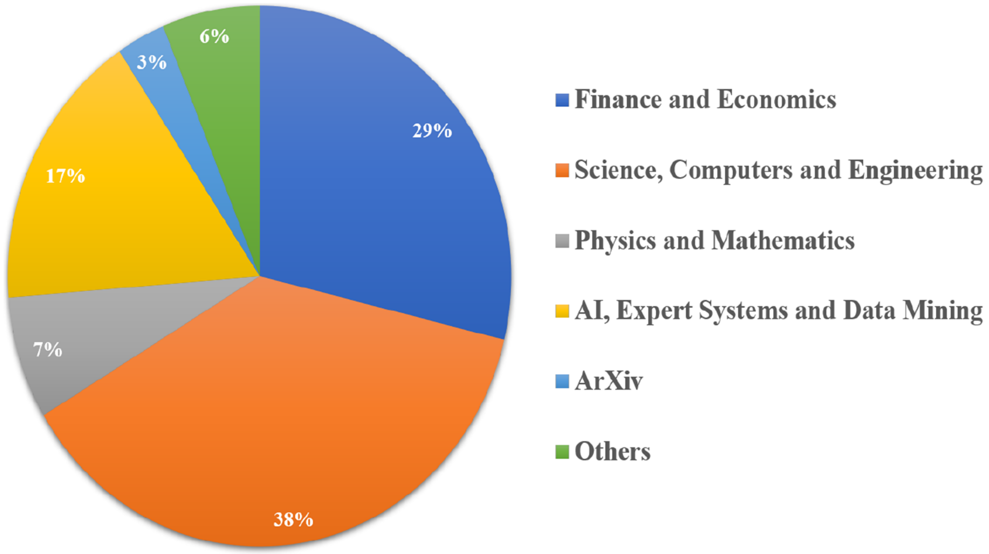


FIGURE 11 Distribution of papers among various disciplines

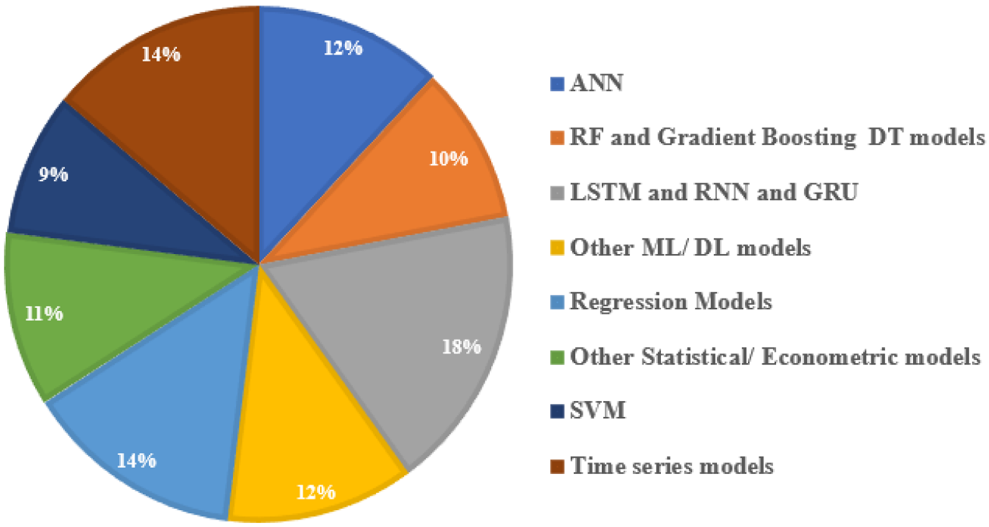


FIGURE 12 Distribution of papers among different categories

Uras, Marchesi, Marchesi, and Tonelli (2020) used four techniques, namely linear regression, multiple linear regression, multilayer perceptron (MLP) NN, and LSTM NN, to forecast the changes in the price of bitcoin based on the daily bitcoin price series, from November 11, 2015 to August 8, 2018. The results showed that all models performed well in predicting the changes in the price of bitcoin. Chen and Narwal (2017) collected data about the price of ether sampled at approximately 1 hr intervals between August 30, 2015, and December 2, 2017, to predict the price changes in ethereum using six techniques, namely logistic regression, naive Bayes, SVM, RF, ARIMA, and RNN.

Poyser (2019) studied using a Bayesian structural time-series approach for the determinants of bitcoin's price. He extracted the data from different websites over the period from January 2013 to May 2017 for 27 different countries to show the variables affecting bitcoin's price level provided by search trends as a proxy for blockchain. Macro-financial statistics and public interest were explored using a Bayesian structural time-series method. The experiments concluded that gold price and exchange rate negatively affected the price of bitcoin. The results showed that the bitcoin price was negatively associated with a neutral investor's sentiment, gold's price, and yuan-to-USD exchange rate, whereas it was positively related to the stock market index, USD-to-EUR exchange rate, and various signs among the different countries' search trends. Bouri, Shahzad, and Roubaud (2019) covered a daily data set on seven cryptocurrencies (bitcoin, ripple, ethereum, litecoin, nem, dash, and stellar) from August 7, 2015, to December 31, 2017. They used logistic regression to study the likelihood of change in the price of one cryptocurrency in response to a change in the price of other cryptocurrencies. They showed that change in the price of one cryptocurrency led to changes in the price of other cryptocurrencies.

Jang and Lee (2018) applied a Bayesian NN (BNN) model, with 10-fold cross-validation, to predict the fluctuations in bitcoin price, as it can naturally deal with the increased number of input variables. An additional regularization factor was included in the objective function of the BNN to prevent the problem of overfitting, which is crucial to any framework. The input features contained blockchain information in addition to macroeconomic variables. The experiment applied two input sets, one comprising of all 26 variables collected and the other containing 16 abridged input variables by extracting important variables and eliminating the redundant ones from the linear correlation analysis. The results confirmed that the BNN is better than the benchmarked models of SVR and linear regression. The error value of the BNN was comparatively small with all the 26 input variables, compared with the reduced variable data set. This indicates that the ignored variables explained the nonlinear relationships that accounted for the output variables. The study showed the fluctuations using the BNN model in bitcoin up to August 2017.

Kim et al., (2016) analyzed cryptocurrency communities for social activities to determine whether these attributes affected the prices of bitcoin, ripple, or ethereum. User comments and replies from these communities were extracted and modeled using ML models like the probabilistic NN and SVM to predict the price fluctuations and the transactions count. The paper concluded that the method predicted the low-cost cryptocurrencies' price variability.

Mallqui and Fernandes (2019) proposed that a selected set of attributes combined with the best data-mining model can improve the accuracy rate of cryptocurrency price prediction. In phase 1, different feature-selection algorithms, including relief technique, correlation analysis, information gain method, correlation-based feature subset selection, and principal component analysis, were evaluated to determine the most relevant attribute for prediction. In phase 2, ML models like ANN, SVM, ensemble techniques (based on RNN and *k*-means clustering) were evaluated. The proposed model was used for the prediction of maximum, minimum, and closing price of bitcoin. The SVM model with relief technique for attribute selection had the highest and most consistent accuracy rate in all time intervals. The proposed model showed a prediction accuracy improvement of 10% compared with the models proposed in previous research works. Whereas most of the research work focused on predicting the price of bitcoin, Poongodi et al., (2020) used ML models of linear regression and SVM for ethereum price prediction. SVM provided a higher accuracy of 96.06% over the linear regression method, which generated an accuracy of 85.46%.

Hitam, Ismail, and Saeed (2019) proposed an optimized SVM based on particle swarm optimization (PSO) in predicting price of selected cryptocurrencies, such as bitcoin, ethereum, litecoin, nem, ripple, and stellar. PSO iwas used for optimizing the SVM in the cryptocurrency market. Five years' daily price from 2013 to 2018 were used to predict the future price. Various classifiers were trained with the same feature set and were evaluated based on classification accuracy. The results showed that SVM–PSO exceeded others and is considered a reliable forecasting model.

Daily price changes in multiple cryptocurrencies, including bitcoin, exclusive coin, ripple, litecoin, lisk, monero, siacoin, cryptonex, AdEx, and guldencoin, were predicted in Mittal et al., (2018) using a multivariate linear regression model. In the initial stage, the independent features in the data set were examined, followed by figuring out the correlation between independent and dependent attributes. Later, the lowest and highest prices of cryptocurrencies were predicted using linear regression. Though the proposed approach yielded high efficiency, the experiments were limited to smaller data sets.

Jain, Tripathi, Dwivedi, and Saxena (2018) introduced a novel method to predict the price of two of the most widely used cryptocurrencies, namely litecoin and bitcoin, based on sentiments of users' tweets. Useful features from the tweets were analyzed and extracted in a multiple linear regression model for price prediction of the cryptocurrencies using R^2 score. This approach worked in two phases, in which one is the training phase and the other is the detection phase. The training phase was a single-time process in which the concurrent cryptocurrency prices and twitter data were collected. The collected data, if not in the same format, were converted into the same format and then tweets were evaluated for sentiment polarity. Tweets were then tagged as positive (if polarity above zero), negative (if polarity less than zero), and neutral (if polarity equal to zero). The tagged tweets were then stored and broken into chunks and the number of tweets in each chunk was counted. The counted number was then mapped with the average price in time duration of 2 hr. These tags represented the features of the data set, with mapped average price as its label. The model was then evaluated using actual labels, and if the labels matched with the original labels then the model was acceptable and was ready for prediction by validating real-time tweets, otherwise a new model was generated. The same process was done until an acceptable model was generated. After that, the detection phase started and the average price was predicted for the next 2 hr. Litecoin price was found to be highly affected by the tweet sentiments. Social factors can also influence the price and its prediction; these were not considered fully in this paper and were to be considered in future studies.

Chen et al., (2020) compared the statistical methods of linear discriminant analysis and logistic regression with more complicated ML approaches, including quadratic discriminant analysis, XGBoost, RF, LSTM, and SVM, on predicting daily bitcoin price using high-dimensional features. ML models outperformed statistical models with an average prediction accuracy of 62.2% over 53.05%. The best performance was observed with the LSTM model, generating an accuracy of 67.2%.

Shah and Zhang (2014) applied Bayesian regression to predict bitcoin price and compared this with earlier researches on bitcoin price prediction. The two major governing features of bitcoin price prediction are bitcoin mining speed and market capitalization. Though this method achieved high profitability, the relationship between features in space and bitcoin price was not discussed.

Madan, Saluja, and Zhao (2015) explored the link in the problem space to understand 16 additional features surrounding the bitcoin network, while also implementing different ML algorithms. In phase 1, daily bitcoin data were analyzed using three binomial classification algorithms: binomial logistic regression, SVM, and RF algorithms. In phase 2, the same set of ML algorithms was applied on bitcoin price interval data. The binomial generalized linear model outperformed in daily bitcoin price prediction with a higher percentage of true positives compared with true negatives. RF yielded a result fairly close to the original data. In phase 2, a 10 min window data yielded better results compared with a 10 s interval. RF showed higher accuracy than the generalized linear model using interval data because RF uses nonparametric decision trees, so linear separability of the data and outliers were not a concern. The error rates were higher for SVM using daily and interval-based bitcoin data because there was a necessity to generate artificial separations between data points in high-dimensional space for the classification of points. Though a prediction accuracy of 97% is noted in the research, the models were not cross-validated, limiting the generalizability of the results.

Silva de Souza, Almudhaf, Henrique, Negredo, Ramos, Sobreiro, and Kimura (2019) showed how an SVM and an ANN could generate abnormal risk-adjusted returns. They showed their approach to bitcoin. The procedure is more suited to practitioners, although the model ignored actual economic policy uncertainly and the relationship with bitcoin. Another interesting study is that of Valencia et al., (2019), who used social media data along with ML to predict price movement. However, the work acknowledged that social media data variables are limited.

Virk (2017) compared RF, SVM, GB, and linear regression to predict the price of bitcoin. The results indicated that the SVM achieved the highest accuracy of 62.31% and a precision value 0.77 among binomial classification ML algorithms.

Sun, Zhou, and Lin (2019) used RFs with factors in Alpha101 (Kakushadze, 2016) by representing features using cryptocurrency market history data obtained from Bitfinex and Binance to build a prediction model. They collected data from API in cryptocurrency exchanges and used 5 min frequency data for backtesting. The results indicated that the performances were proportional to the quantity of data (more data, more accurate) and the factors used in the RF model appeared to have distinct significance.

Barnwal, Bharti, Ali, and Singh (2019) used generative and discriminative classifiers to create a stacking model. The model included three generative and six discriminative classifiers combined using a one-layer NN to predict the cryptocurrency price trends. A discriminative classifier directly models the relationship between unknown and known data, whereas generative classifiers model the prediction indirectly through the data generation distribution. Technical indicators, such as trend, momentum, volume, and volatility, are collected as features of the model. The authors discussed how different classifiers and features affected the prediction.

Attanasio, Garza, Cagliero, and Baralis (2019) compared various classification algorithms, including SVM, naive Bayes, and RF, in forecasting the next-day price trends of a given cryptocurrency. The results indicated that forecasting models based on a series of forecasts appeared better than a single classification model, due to the volatility and heterogeneity of cryptocurrencies' financial instruments.

Table 4 summarizes the ML techniques surveyed in Section 5.

TABLE 7 Comparison of common machine-learning techniques used

Attribute	ANN	SVM	RF
Network model	Type of feed-forward neural network	Nonlinear model	Decision-tree-model
Layers	Single-layer and multilayer forms	N/A	It follows tree structure
Used for	Minimizing error functions	Time-series prediction task	Eliminates instability in a network
Accuracy	ANN has highest accuracy for predicting price	It eliminates irrelevant and scattered data so it shows improved precision and accuracy but not comparable to RF	Best feature with highest information efficiency is selected in this network so it outperformed SVM in predicting price

ANN: artificial neural network; N/A: not applicable; RF: random forest; SVM: support vector machine.

TABLE 8 Comparison of common deep-learning techniques used

Attributes	MLP	RNN	LSTM
Network model	Feedforward static neural network	Forward and backward dynamic network	Forward and back-propagated with forget/remember gates
Layers	Input, output, and hidden layers	Input, output, hidden, and context layers	Input, output, and forget gates
Used for	Classification tasks	Time-series prediction task	Recognizing long-term associations
Limitations	Vanishing gradient problem and signals only pass forward in static nature	Vanishing gradient is still an issue, but signals can pass forward as well as backward dynamically	Addresses both gradient and signal passing problem with long term dependency

LSTM: long short-term memory; MLP: multilayer perceptron; RNN: recurrent neural network.

5.1 | DL and RL techniques

Whereas ML-based models have achieved some success in predicting cryptocurrency prices, their chaotic and very complex nature and fluctuations in accordance with fast-paced technological developments, security, political, and economic factors have motivated researchers to investigate the application of DL and RL models toward providing accurate predictions of cryptocurrency prices. DL models have already found numerous applications in quantitative finance, such as forecasting volatility trends. In a supervised learning scheme, NNs are a useful tool for price prediction since no strong assumption is needed for their application, which contrasts with conventional time-series models, such as ARIMA and its extensions. DL models are based on ANNs. The key difference that distinguishes a DL model from an ANN is that, generally, the DL model corresponds to an ANN with multiple hidden layers. Moreover, DL models capture patterns with a significant generalization power. Most recent LSTM networks seem more suitable and convenient for sequential data, such as time series. The commonly applied DL models in cryptocurrency price prediction research include RNNs, gated recurrent units (GRUs), convolutional NNs (CNNs), LSTM networks, and MLPs. An RNN is a class of ANN in which the connections between nodes form a directed graph with possible loops, and this structure of recurrent connections with memory makes them suitable for processing time-series data. However, they face the vanishing gradients problem, and hence different variations have recently been proposed (Kwon, Kim, Heo, Kim, & Han, 2019). Another class of standard RNN used in cryptocurrency price prediction is a GRU (Phaladisailoed & Numnonda, 2018). An MLP is a type of feedforward ANN used in predicting cryptocurrency prices (Misnik, Krutalevich, Prakapenka, Borovykh, & Vasiliev, 2018). A CNN is another class of DL used for supervised learning that has been successfully applied in various image-processing and natural language-processing problems and is also adopted for cryptocurrency price prediction by researchers (Ji, Kim, & Im, 2019). Another widely used DL model, called LSTM, is a special RNN structure found to be superior to non-gated RNNs on financial time-series problems because of its ability in selectively remembering patterns for a long time (Chen et al., 2020; Mohanty et al., 2018). RL is another class of ML that enables software agents to learn and take appropriate action in an interactive environment to maximize a cumulative reward using feedback from its own actions and experiences. Deep and reinforcement models are also being considered by researchers to provide efficient prediction of cryptocurrency prices (Jiang & Liang, 2017; Längkvist et al., 2014). The deep and reinforcement models used in cryptocurrency price prediction research are as discussed in the following.

DL was used by Lahmiri and Bekiros (2019) for the first-time to predict digital currency prices of the three currencies used most: bitcoin, ripple, and digital cash. Long memory was used to access the market efficiency of cryptocurrencies. Inherent nonlinear dynamics were examined, which include inherent chaoticity and fractality, to measure the predictability of digital currency. Then they used DL to extract hidden information or patterns as an underlying dynamical system. They also focused on short-term predictability of cryptocurrency. The largest Lyapunov exponent and a detrended fluctuation analysis based on the extracted Hurst exponent were used to deal with the chaotic and fractal characteristics of digital currency. To make predictions accurate and fast, they developed a complex NN based on LSTM to extract hidden information. This process of extracting domain-specific patterns is time consuming and extremely expensive, but it introduces consistency and accuracy in predicting digital currency prices. An LSTM NN overcomes the problem present with RNNs. Nodes will be replaced with memory cells, and a forget/remember gating mechanism is used. Information can be stored for future processing. A well-known benchmark known as the generalized regression NN provides fast learning and optimal convergence. Owing to these unique characteristics of a generalized regression NN and LSTM, they efficiently applied it to solve various data analysis and data modeling problems. LSTM has higher predictive accuracy, but it takes more time to converge.

In the case of ANNs, the precision is highly dependent on an adequate set of inputs. Considering this, Misnik et al., (2018) applied various approaches to obtain additional data for an NN to analyze the high volatility trends exhibited by cryptocurrencies, and the impact on precision was analyzed.

McNally, Roche, and Caton (2018) used ML for predicting bitcoin price, and they compared this approach with a parallelization method that executes on multi-core and graphics processing unit environments. An ARIMA time-series model was introduced in that paper to compare the performance with DL models. An RNN and LSTM were used as a DL model with LSTM more capable of recognizing longer term associations. The closing price in USD for bitcoin was considered an independent variable for thes study and was collected from CoinDesk Bitcoin Price Index. The work considered five major bitcoin exchanges: Bitstamp, Bitfinex, Coinbase, OkCoin, and itBit. To assess the performance of the system, RMSE was used. The actual performance of the ARIMA model was worse thanthe the NN model. LSTM outperformed RNN but took a significantly longer time to train.

Yao et al., (2018) developed a framework using DL for cryptocurrency price prediction. The nonlinear nature of cryptocurrency was considered in this for prediction. In this paper, various factors are considered for cryptocurrency price prediction, such as volume, market cap, circulating supply, and maximum supply based on an RNN and LSTM. Different phases were involved in this method, including data analysis phase, data filtration phase, train–test split phase, data scaling phase, model building phase, model learning and evaluation phase, and prediction phase. Then, the model was executed and tested for benchmark data sets. The proposed approach achieved excellent accuracy, and this varied with reference to the size of the data set. Simulation results showed that the market open plays a significant role in predicting price.

Ji et al., (2019) discussed various DL approaches, including a deep NN, LSTM, a deep residual network, and a CNN, and compared them for bitcoin price prediction. Simulation results showed that the LSTM model outperformed the other models. Three methods were used to predict cryptocurrency prices of 1,681 currencies by Alessandretti, ElBahrawy, Aiello, and Baronchelli (2018). The two models used were based on GB decision trees, in which one uses the same model to predict the return on investment of all currencies and the second method uses a different model for each currency and uses information related to the whole market to predict a single currency. The third method was based on LSTM for RNN, in which prediction depends on previous currency prices. LSTM is very stable and is also capable of capturing long-term dependencies.

Multiple inputs using an LSTM-based prediction model was proposed by Li, Arab, Liu, Liu, and Han (2019) in conjunction with Black–Scholes to predict the price volatility for the next 30 days. They leveraged on the multiple inputs of the LSTM-RNN as a prediction model. They found out that it was more efficient in predicting bitcoin price. The data set iwas collected using traditional market statistics, such as daily price, trading volume, and historical volatility. Blockchain statistics included wallet address liveness of bitcoin exchange. In contrast, social media trends include an impact factor of Google/Reddit/Twitter. Using blockchain statistics, accurate results were obtained that also reduced the RMSE compared with the baseline approaches.

Phaladisailoed and Numnonda (2018) compared different ML models and selected features such as open, close, high, and low values to predict the future price. MSE and R^2 were used to measure the accuracy of the Theil–Sen regression model, Huber regression, LSTM, and GRU. The simulation results showed that the DL models, such as LSTM and GRU, showed better results than Theil–Sen and Huber regression, with an accuracy of 99.2% in the case of GRU. In contrast, the calculated time for Huber regression is much less than with LSTM and GRU.

An LSTM model is presented by Kwon et al., (2019) to classify cryptocurrency price time-series. Past data were collected and preprocessed to clean it for training and testing data. Then, data were encoded into a three-dimensional price tensor that represented the past price changes of cryptocurrencies. A grid-search-based *k*-cross-validation was applied to search for the most suitable parameters in Lthe STM model. Comparison results showed that the LSTM outperformed GB model and other ML models.

Jiang and Liang (2017) proposed using RL. The deep RL was carried out to address a portfolio management problem. A decision-making process in which some amount of funds is allocated to different financial products to achieve a maximum return while restraining from the risk. As an input, a portfolio vector was produced from raw data and historical prices. A model-less CNN was used in this approach, in which historic prices acted as an input, outputting portfolio weights. Training was done to reinforce and that maximizes the accumulative sums; that is, the reward function of the network. This network can be applied to any other financial market and is not limited to only one cryptocurrency. Its performance was tested after every 30 min with three recent portfolio-selection algorithms, achieving positive results. Simulation results showed that this method had less cumulative return than the other techniques. The major drawback of using this technique is that two assumptions are made that are not useful in the real market because history data is not involved in testing and training of the algorithm.

Sun et al., (2020) adopted a GB decision tree algorithm, light GB machine (LightGBM), to forecast the price trend of the cryptocurrency market. Though results were good, the study did not try a large data set. Alonso-Monsalve, Suárez-Cetrulo, Cervantes, and Quintana (2020) proposed a convolution LSTM NN model and tested it on six currencies: bitcoin, dash, ether, litecoin, monero, and ripple. The model was used for price prediction, and results were compared with a hybrid CNN–LSTM network, MLP, and radial basis function NN. According to the paper, the model could identify trends but may not lead to profitability.

Snihovyi, Ivanov, and Kobets (2018) developed a multiple line multiple linear regression, RF, and LSTM ML model and worked on bitcoin to predict price variation. The authors proposed a criterion for predicting cryptocurrency prices using the Python Anaconda Data Science tool combined with ML algorithms, and this combined criterion could explain more than 70% of the variation in cryptocurrency prices using either multiple regression, RF, or LSTM networks. According to the study, the data set was small. Altan, Karasu, and Bekiros (2019) used an LSTM model and applied it on four cryptocurrencies (namely, bitcoin, ripple, dash and ltc) for currency prediction according to the paper model. It showed a good prediction of currency. The paper did not highlight any limitations.

Table 5 summarizes the DL and RL techniques surveyed in Section 5.1.

6 | DISCUSSION

This section discusses and analyzes the findings from our survey of cryptocurrency price prediction research. We have included the most relevant papers related to cryptocurrency price prediction, published in the period 2010–2020. For this, we reviewed different types of research papers that included papers from categories such as journals, conferences, workshops, surveys, and other publications. The focus of researchers towards cryptocurrencies started to ascend considerably over the past few years. The research publications in this domain have been increasing since 2017, and especially in 2018 and 2019; and this trend continued in 2020 (see Figure 10 for cumulative percentage of publications per year). Bitcoin is the most common currency when it comes to implementation and research. However, since 2018, researchers have started considering the topic of cryptocurrency price prediction as a whole.

Each of these papers focuses on cryptocurrency price prediction using various approaches. The papers surveyed are classified into different categories that include traditional statistical and ML techniques. The primary category contains various traditional statistical and econometric techniques and the secondary category contains different ML and DL techniques used in cryptocurrency price prediction research.

Table 6 presents the distribution of papers analyzed in this survey versus different categories.

The majority of them are related to ML models, though conventional statistical and econometric models are still significant. Among the papers studied in this survey, 45.97% of papers belong to the category of conventional statistical or econometric approaches and the remaining 54.03% belong to the ML category. Among the ML category of papers related to cryptocurrency price prediction research, though 70.22% of papers concern ML models, 29.78% of research applies the DL/RL category of models. The distribution of papers among different publication sources is illustrated in Figure 11. Among the published papers, finance-related literature includes the finance and economics area and the business, management, and accounting area; other disciplines include Physics and mathematics, computers, science and engineering, artificial intelligence, expert systems and data mining, ArXiv, and others. It is evident that the topic is being explored and is gaining attention from multiple disciplines.

From the analysis conducted in this survey, it is clear that the researchers are now focusing more on models using ML for cryptocurrency price prediction, as ML and DL approaches play a significant role in cryptocurrency price prediction research. It is clear from the papers that the accuracy of prediction depends highly on the input attributes and ML technique used. Also, the accuracy of an ML algorithm is highly dependent on the problem and the integrity and complexity of the training data set. The majority of the studies focused on daily and interval-based price data. In contrast, some studies also focused on the impact of socio-economic factors, user trends, and macro attributes on the price prediction. Irrespective of the input attributes being focused on, SVM, Bayesian network, linear regression, ANN, logistic regression, RF, and LSTM are some of the main techniques highlighted in cryptocurrency price prediction. Figure 12 gives an illustration of research paper distribution among different categories of models used in the cryptocurrency price prediction domain. Moreover, among the papers analyzed, it is noteworthy that 26.92% of researchers applied and compared different models in their work.

6.1 | Analysis: Research contribution among different categories

The issue of developing suitable models for predicting prices of cryptocurrencies is relevant and significant for the scientific community, investors, financial analysts, and traders. Researchers have adopted different approaches to analyze the volatility dynamics of digital currencies and to predict their prices. One such approach is to build a cause-and-effect casual model that describes the relationship between exchange rates and other macroeconomic variables (especially economic growth rate, trade and balance of payments, inflation rates, etc.)

incorporating various economic concepts (e.g., Akcora et al., 2018; Gunay, 2019; Wiedmer, 2018). Another approach is to study the time series and make a prediction based on the processing and analysis of past observations. The most common models are the Box–Jenkins ARIMA time-series models and their modifications, GARCH models, or ANNs (Alahmari, 2019; Derbentsev et al., 2019; Ho, Xie, & Goh, 2002; Lu, 2010; Radityo et al., 2017; Sovia et al., 2019; Vapnik, 1999). Generally, the models for predicting cryptocurrency prices depend on an analyst's perception of the causal relationships in the pricing process. For example, consider a prediction model specified as a price formation model, taking into account different aspects as discussed in the following. It can be based on the interaction of market players that make economic decisions based on some indicators, taking into consideration the objective economic laws or behavioral finance laws (econometric and balance models) or by considering the production and other technological possibilities of creating the corresponding asset, such as commodity markets, fundamental, mining cryptocurrency, and so on (Alessandretti et al., 2018; Alvarez-Ramirez et al., 2018; Bouri, Shahzad, & Roubaud, 2019; Gunay, 2019; Kaya, 2018; Lahmiri, 2011; Matta, Lunesu, & Marchesi, 2015; Wang & Chen, 2020). Moreover, given the past dynamics and volatility trends, time-series models and autoregressive models are also adopted by researchers (e.g., Anupriya & Garg, 2018; Conrad et al., 2018; Derbentsev et al., 2019; Längkvist et al., 2014; Mohanty et al., 2018; Poyser, 2019; Roy et al., 2018; Troster et al., 2019).

We analyzed and compared the research papers of different categories, highlighting the contribution made in each work. Paper summaries of different categories can be found in Tables 3,4, and 5. Our survey highlights that ML and DL play a vital role in predicting the price of cryptocurrencies, but traditional statistical and econometric approaches are still significant. The volatility forecasts can be used to gauge the cryptocurrency price fluctuations, which is also advantageous in the development and analysis of quantitative financial trading practice (Fang et al., 2020).

Traditional statistical (or econometric) approaches: Among this category of papers, the commonly used techniques include basic regression schemes and correlations (linear statistical model, function estimation, CGCD model using copula functions, etc. Blau, 2017; Gunay, 2019; Hayes, 2017; Kaya, 2018; Kjærland et al., 2018; Vaddepalli & Antoney, 2018; Vieira, 2017; Wiedmer, 2018). Schemes based on time-series analysis include the GARCH model and its derivatives, ARIMA model, wavelet coherence ananlysis, and so on (Caporale & Zekokh, 2019; Conrad et al., 2018; Roy et al., 2018; Troster et al., 2019; Walther et al., 2019). In addition, multivariate linear regression, the multivariate vector autoregressive model and extended vector autoregressive model, the value at-risk (VaR) model, the least square method, chainlets, the temporal mixture model, the augmented Dicky–Fuller test, the Johansen tests for cointegration, the vector error correction model, the autoregressive distributed lag model, and the GARCH model, and so on, were also adopted (Abbatemarco et al., 2018; Akcora et al., 2018; Bhambhwani et al., 2019; Choi, 1992; Dos Santos Maciel & Ballini, 2019; Giudici & Abu-Hashish, 2019; Guo & Antulov-Fantulin, 2018; Kjærland et al., 2018; Wang & Chen, 2020).

The empirical analysis of the three most capitalized cryptocurrencies (bitcoin, ripple, and ethereum) did not reveal a static relationship between the yield of cryptocurrencies and the complexity of their extraction (Liu & Tsyvinski, 2020). Also, the macroeconomic factors, which usually determine the dynamics of currency, stock, and commodity markets, have no significant effect on the dynamics of the cryptocurrencies market. The results of Conrad et al., (2018) indicate that the influence of the US stock market (S&P 500 index) and the global stock market index (Nikkei 225 index) on bitcoin's volatility was not significant. The studies conducted by Baek and Elbeck (2015), Kaya (2018), Blau (2017), and Ciaian, Rajcaniova, and Kancs (2016) analyzed the price dynamics of cryptocurrencies and its impact in the cryptocurrency market using classical log-periodic models of price bubbles and their modifications. In the same way, a number of recent cryptocurrency market researches reveal that, unlike other financial assets, cryptocurrency prices are influenced by a number of specific factors that influence their demand, such as the number of Google Trends searches, and the number of posts in social networks and other mass media (Karalevicius et al., 2018; Lahmiri, 2011; Matta et al., 2015; Mittal et al., 2019; Mohanty et al., 2018). These studies substantiated the feasibility of using nontypical factors as predictors. All of these factors complicate the development of casual econometric models of cryptocurrency price dynamics.

As a consequence, in recent years, ML-based techniques have receiving increased attention in the study and analysis of cryptocurrencies and predicting their future price.

ML approaches: These techniques usually exploit supervised models trained on historical data in order to automatically generate and predict the price volatility trends of the cryptocurrencies in the financial markets (Mittal et al., 2018). In the past decade, the ML community has explored the use of ML techniques (e.g., classification, regression, time-series forecasting, DL, and RL) for cryptocurrency price prediction.

Recently, nonparametric methods based on ML and DL have gained popularity for the analysis and forecasting of financial and economic time series and received increased attention from researchers for use in cryptocurrency price prediction (Derbentsev et al., 2019). ML-based models allow one to solve the problem of prediction and classification by utilizing learning sequences in the data. The effectiveness of such models depends on the training speed and the degree of universality of approximating functions. This includes a collection of techniques, such as SVMs, ANNs, fuzzy logic, genetic algorithms, linear and nonlinear statistical models, DL and RL models, and so on (Atsalakis et al., 2019; Galeshchuk & Mukherjee, 2017; Hitam et al., 2019; Jiang & Liang, 2017; Längkvist et al., 2014; Lahmiri, 2011; Lahmiri & Bekiros, 2019; Nikou et al., 2019; Peng, Albuquerque, de Sá, Padula, & Montenegro, 2018; Radityo et al., 2017; Sarlin & Marghescu, 2011; Sin & Wang, 2017; Tupinambás, Cadence, & Lemos, 2018; Uras et al., 2020).

Though some ML models (classification and regression) were applied in predicting price volatility trends of cryptocurrencies, some researchers focused on the comparison of different statistical and ML methods, and also classification and regression-based ML schemes (Abbatemarco et al., 2018; Bush & Choi, 2019; Chen et al., 2017; Ciaian et al., 2016; Hashish et al., 2019; Mittal et al., 2019; Mohanty et al., 2018; Poongodi et al., 2020; Sovbetov, 2018; Uras et al., 2020; Vaddepalli & Antoney, 2018; Valencia et al., 2019). Furthermore, some schemes integrated various prediction models, including some of the popular classification techniques as well as some popular time-series forecasting techniques, while considering multiple aspects (Roy et al., 2018; Längkvist et al., 2014; Chakraborty & Roy, 2019; Derbentsev et al., 2019; Wang & Chen, 2020; Poyser, 2019). For example, some researchers studied the results using classical ARIMA models and different ML techniques, such as RF, linear discriminant analysis, logistic regression, and LSTM (Amjad & Shah, 2017; McNally et al., 2018; Saxena, Sukumar, Nadu, & Nadu, 2018). Their analyses indicated that the models that relied on training proved to be more appropriate for predicting both the prices of cryptocurrencies and their volatility. Another comparative analysis of the ARIMA forecasting properties with RNNs for cryptocurrencies, such as ethereum (ETH), dash, litecoin (LTC), stellar (STR), siacoin (SC), nem (XEM), monero (XMR), and ripple (XRP), showed that RNNs had better prediction capabilities than ARIMA models (Bush & Choi, 2019; Hashish et al., 2019; Mohanty et al., 2018; Rebane, Karlsson, Papapetrou, & Denic, 2018).

Most of the researches on cryptocurrency price prediction during 2010 to 2020 aimed at predicting the price of bitcoin, which is the trending and most capitalized cryptocurrency. For example, Hayes (2017), Jang and Lee (2018), Madan et al., (2015), McNally et al., (2018), Sin and Wang (2017), and Wu, Lu, Ma, and Lu (2018) addressed the prediction of the next-day trend of bitcoin (up or down) by adopting binary classification models trained on historical data. Various models, such as logistic regression, RF (Attanasio et al., 2019; Sun et al., 2019; Virk, 2017), SVMs (Silva de Souza et al., 2019; Madan et al., 2015), MLPs and genetic algorithms (Sin & Wang, 2017), Bayesian NNs (Jang & Lee, 2018), and LSTM and RNNs (Hashish et al., 2019; Kwon et al., 2019; Li et al., 2019; McNally et al., 2018; Rebane et al., 2018; Wu et al., 2018). Parallel attempts to perform intra-day price forecasting of bitcoin have also been made (e.g., Shah & Zhang, 2014; Tupinambás et al., 2018). Since bitcoin is also a distributed network that enables users to store and transfer digital currency, particular attention has been paid to the enrichment of time-series data with ad hoc features related to bitcoin trading and the bitcoin network, such as the average transactions count per block. The feature engineering process is aimed at including new variables that describe potentially discriminating factors in the prediction models, such as user activities, the level of attractiveness for investors, and global macrofinancial factors (Attanasio et al., 2019; Ciaian et al., 2016; Mohanty et al., 2018). Moreover, the correlation between the distribution of the bitcoin's price and the volumes of the related tweets or media published on the Web have been investigated as well (Matta et al., 2015).

As bitcoin and blockchain technology have begun to shape and define new aspects, researchers have started focusing on the matter of blockchain and cryptocurrencies. The public nature of blockchain technology opens the door for new price prediction challenges. Investigations on the relationship between a given currency's transaction network and its price have increased rapidly in recent years; the growing attention on user identification also strongly supports this direction. In-depth knowledge of these networks can help researchers in future to better identify new features in price prediction. In the past few years, there has been a rapid growth of numerous cryptocurrencies, hashing algorithms, and consensus agreements in the networks (Brooks, 2019). This grabbed the attention of researchers. Moreover, as blockchain has potential applications far beyond bitcoin, some researchers have examined the applications based on blockchain and their benefits in the crypto market in providing decentralized systems. The work of O'Leary (2017) investigated the alternative configurations of different blockchain architectures that can be used for gathering and processing transactions in a range of different settings, including accounting, auditing, supply chain, and other types of transaction information. Although there has been substantial focus on the peer-to-peer and public versions of blockchain, this paper focuses primarily on cloud-based and private configuration versions of blockchains and investigates use configurations, advantages and limitations as firms bring blockchain-based market mechanisms into their organizations. In addition, this paper investigates some emerging issues associated with blockchain use in consortium settings. Finally, this paper relates some proposed uses of blockchain for transaction processing to other technologies, such as data warehouses and databases. O'Leary (2018) investigated "open information transactions." Such transactions are in contrast to traditional transactions, where typically two parties to a transaction are the only ones with information about the transaction. For example, in a sale, the seller and the purchaser typically are the only ones with information about the transaction. However, some emerging technologies, such as blockchain accounting, supply chain social media, and hashtag commerce, are making information about the transactions potentially openly available to others. They investigated some of the implications and strategies that include the use of that open information. For example, open information in accounting and supply chain transactions provides the potential for both business intelligence analysis of the information and possibly misleading and illusory transactions, analogous to those that have garnered the recent attention of the Justice Department in cryptocurrencies. Finally, this paper suggests that blockchain transaction processing will provide reliable information in those settings where there is a "single truth" feed of information flow for the phenomena of interest, no ability to do off-blockchain transactions (or a large penalty cost), and limitation to a single identity for each enterprise on the blockchain.

An interesting research problem in our age of big data is that of determining provenance. Granular evaluation of provenance of physical goods (e.g., tracking ingredients of a pharmaceutical or demonstrating authenticity of luxury goods) has often not been possible with today's items that are produced and transported in complex, interorganizational, often internationally spanning supply chains. Recent adoptions of the Internet of Things and blockchain technologies give promise at better supply-chain provenance. The work of Kim and Laskowski (2018) evaluated how ontologies can contribute to blockchain design. They were particularly interested in the blockchain, as many favored use cases of blockchain are for provenance tracking, and they examined the application of ontologies on blockchain. To support this case, they analyzed a traceability ontology and translated some of its representations to smart contracts that execute a provenance trace and enforce traceability constraints on the ethereum blockchain platform.

O'Leary (2019) reviewed some recent blockchain-based applications for information capture, distribution, and preservation. As part of that review, this paper examines two key concerns with current blockchain designs for accounting and supply-chain transactions: data independence and multiple semantic models for the same information distribution problem. Blockchain applications typically integrate database, application, and presentation tiers all in the same ledger. This results in a general inability to query information in the ledger and other concerns. Further, since most applications appear to be private blockchain applications, there is a concern of agents needing to accommodate multiple blockchains depending on who their trading partners are and what they request. Finally, this paper uses a distributed database to design a "blockchain-like" system for virtual organizations. Money exchange is one of the most common day-to-day activities performed by humans in the daily market. Mohamed et al., (2019) presented an approach to money tracking through a blockchain. The proposed approach consisted of three main components: serial number localization, serial number recognition, and a blockchain to store all transactions and ownership transfers. The approach was tested with a total of 110 banknotes of different currency types and achieved an average accuracy of 91.17%. They conducted a user study in real time with 21 users, and the mean accuracy across all users was 86.42%. Each user gave feedback on the proposed approach, and most of them welcomed the idea.

Behavioral science states that emotions, principles, and the manner of thinking can affect the behavior of individuals and even investors in their decision-making on financial markets. Mnif, Jarboui, Hassan, and Mouakhar (2020) tried to measure the investor sentiment by three means of big data. The first was based on a search query of a list of words related to Islamic context. The second was inferred from the engagement degree on social media. The last measure of sentiment was built, based on the Twitter API classified into positive and negative directions by an ML algorithm based on the naive Bayes method. Then, they investigated whether these sensations and emotions had an impact on the market sentiment and the price fluctuations by means of a vector autoregression model and Granger causality analysis. In the final step, they applied the agent-based simulation by means of the sequential Monte Carlo method with the control of the Twitter measure on Islamic index returns. They showed that the three social media sentiment measures presented a remarkable impact on the contemporaneous and lagged returns of the different Islamic assets studied. They also gave an estimation of the parameters of the latent variables relative to the agent model studied.

In Rivas, Parras-Gutiérrez, Merelo, Arenas, and García-Fernández (2017), a time-series forecasting method (jsEvRBF) based on a genetic algorithm and neural nets written in JavaScript language that can be executed in most web browsers was proposed. Consequently, everybody can participate in the experiments, and scientists can nowadays take advantage of the available browsers and devices as computation environments. This is also a great challenge, as the language support and performance varys from one browser to another. jsEvRBF has been tested in a volunteer computing experiment, and also in a single-browser one. Both experiments were related to forecasting currencies exchange, and the results showed the viability of the proposal.

In financial trading, technical and quantitative analysis tools are used for the development of decision-support systems. Although these traditional tools are useful, new techniques in the field of ML have been developed for time-series forecasting. The work of Aloud (2020) analyzed the role of attribute selection on the development of a simple DL-ANN multiagent framework to accomplish a profitable trading strategy in the course of a series of trading simulations in the foreign exchange market. The paper evaluated the performance of the DL-ANN multiagent framework over different time spans of high-frequency intraday asset time-series data and determined how a set of the framework attributes produced effective forecasting for profitable trading. The paper showed the existence of predictable short-term price trends in the market time series, and an understanding of the probability of price movements may be useful to high-frequency traders. The results of that paper can be used to further develop financial decision-support systems and autonomous trading strategies for the financial market.

There are also several studies that analyze the contagion risks and spillover effects in cryptocurrencies, such as the work of Catania and Sandholdt (2019), Zhang, Chan, Chu, and Nadarajah (2019), and Eross, McGroarty, Urquhart, and Wolfe (2019). The predictability of cryptocurrencies' returns and volatility has been studied by Catania and Sandholdt (2019). These authors modeled the predictability at high frequencies up to 6 hr, but not at higher aggregation levels, while realized volatility was characterized by long memory and leverage effects. Such researches have either considered model-based estimates of bitcoin volatility or other nonparametric measures to construct daily or higher frequency time series of volatility. Additional results about the time-dependence properties of cryptocurrencies are reported and analyzed in the model proposed by Zhang et al., (2019) to forecast the VaR for bitcoin. Moreover, some studies have indicated that the standard volatility models, like GARCH, are generally not suitable for cryptocurrency time-series and suggest to use a more sophisticated modeling technique based on the score-driven approach (Creal, Koopman, & Lucas, 2013). The studies performed by Eross et al., (2019) pointed out that the European and North American traders are the main drivers of bitcoin trading, and the trading volume is the highest during the morning and day time, which is consistent with the other currency markets. By employing GMT timestamped tick data aggregated to the frequency of 5 min, they found that bitcoin returns have increased over time, while trading volume, volatility, and liquidity varied substantially over time. Realized volatility is fairly consistent throughout the day, although it is highest during the opening times of the three major global stock markets. Overall, it is seen that the trend for research into bitcoin, and increasingly into other cryptocurrencies, continues to grow and attract further research. Table 7 summarizes some of the findings on common ML approaches applied for cryptocurrency price prediction.

Based on the analysis conducted in this survey, we highlight the following points that can help researchers to fill the gap in existing studies and gain more insight on cryptocurrency price prediction research.

- The price fluctuations of bitcoin were anticipated by these studies to different degrees and revealed that NN-based algorithms yielded the best results (Greaves & Au, 2015; Jang & Lee, 2018; Chen et al., 2020). This is because an ANN has the ability to learn the nonlinear relationship between input and output variables contained in real-time cryptocurrency data sets. Bayesian regression provided acceptable accuracy in terms of profit and daily price prediction.

- RF outperformed the SVM in predicting bitcoin price (Madan et al., 2015). This could be because RF uses nonparametric decision trees, so outliers and linear separability of the data are not a concern. SVM, on the other hand, needs to create artificial separations between data points in higher dimensional space to classify points. However, the SVM model, combined with the relief technique for attribute selection, yielded an acceptable prediction accuracy. The performance of the SVM depends greatly on the parameters set during training—mainly the kernel function. The settings that produced the best accuracy for a problem may result in reduced efficiency for others. The price of ethereum was predicted accurately by the SVM model with the given selected features (Poongodi et al., 2020).
- The linear regression model is most often the first to be used in many problems owing to its simplicity and widespread availability. Though not exceeding other ML models, linear regression provided a reasonable accuracy in cryptocurrency price prediction (Greaves & Au, 2015; Mittal et al., 2018; Poongodi et al., 2020).
- While considering the combined performance of all the ML classifiers on any cryptocurrency price prediction, ANN and RF models are best suited for cryptocurrency price prediction in common. The SVM provided the highest accuracy for ethereum.
- Ensemble techniques in ML are not explored that much in the field of price prediction. Also, there is not much focus on optimizing ML techniques to improve accuracy.
- Researchers are recommended to use a hybridized approach of an SVM combined with a genetic algorithm to gain more accuracy, as well as optimized results.

In the field of DL, the RNN and LSTM are the well-known approaches, and they have many potential advantages compared with the traditional MLP. Some of the findings on common DL approaches applied for cryptocurrency price prediction are summarized in Table 8.

- The MLP model is a simple feedforward NN comprised of input and output layers with a hidden layer in between them. Each output from one of these layers represents a unit that is similar to the neuron in a human brain (Yao et al., 2018). Connections between these units are considered as weights that are similar to synapses in the brain. The performance of the MLP is limited due to the vanishing gradient problem. This problem is the primary concern, as the network becomes too small to learn. Another limitation in using an MLP is that signals can only pass forward statically, so it becomes difficult for it to recognize the temporal element of time series effectively.
- The RNN is a dynamic NN that addresses some of the limitations present in an MLP. Its structure is similar to an MLP, but in this technique the signals can pass forward as well as backward in both directions in an iterative manner and make it easy to add another context layer in the RNN. In this new layer, at each time, the step state can be overwritten. In this approach, the temporal issue of an MLP can be solved by assigning weights to events that are occurring in series rather than assigning the same weights to all inputs, but the vanishing gradient is still an issue in this approach.
- LSTM addresses both the issues present in the MLP and the RNN. This approach allows the weight preservation of signals that are forwarded and backpropagated between layers. This approach recognizes long-term dependencies. An LSTM unit is composed of a cell and three gates. The cell remembers values over arbitrary time intervals, and the three gates (i.e., input, output, and forget gate) regulate the flow of information into and out of the cell. In this way, weak signals get blocked, avoiding the vanishing gradient problem. LSTM has three state dependencies (i.e., previous cell state, previous hidden state, and current time steps) for memorizing things and special gates for manipulating this memory. So this approach selects the most appropriate information from the cell state.
- From the aforementioned techniques, we can conclude that LSTM can be considered to be the best technique for solving time-series prediction problems and recognizing long-term associations by removing irrelevant information from the network.

7 | FUTURE DIRECTIONS

In the cryptocurrency world, prices are very volatile. Hence, it is essential to encourage research that incorporates new techniques, strategies, and alternative approaches, such as more sophisticated prediction algorithms, advanced ensemble methods, feature engineering techniques, and other validation metrics for gaining accurate cryptocurrency price prediction. This can assist cryptocurrency investors toward potential increased profits, support policy-makers and financial researchers in studying cryptocurrency markets behavior. In general, it is seen that the trend for research into bitcoin, and increasingly into other cryptocurrencies, continues to grow and attract further research. There is no guarantee of which cryptocurrency will dominate in the future, so research work in this area is of significant importance. The discussion here could be beneficial in exploring some promising opportunities that remain open in cryptocurrency price prediction research.

- The adoption rates are increasing all the time, which includes more and more people using the network and an increase in wallets and apps. So, this may help in increasing the price, as it gives the cryptocurrency real-world usage. There is plenty of further work to be done in this area to understand the influence of these factors.
- Further study would require inclusion of all relevant factors that would influence market moods, as well as tracking over a longer time period, to understand the anomalous behavior of cryptocurrencies and their prices. This is a field with a lot of potential for research in financial time-series problems because of their high data availability and accessibility. The evidence from this survey suggests that there is much room for further work and enhancements.
- Researchers are recommended to analyze the use of LSTM models in future study, such as CNN LSTMs, bidirectional LSTMs, encoder–decoder LSTMs, and also to compare their results to attain good future insight and improve price prediction results. The intraday behavior of the intraday variables does vary over time, and this indicates that researchers should be careful to guard against this when examining any aspects of cryptocurrency (Blau, 2017;

Eross et al., 2019; Karalevicius et al., 2018; Kjærland et al., 2018; Poyser, 2019).

- Another promising perspective in the study of cryptocurrencies includes analyzing the impact of public opinion, as measured through social media traces, on the market behavior and prices in an improved way (Mittal et al., 2019; Phillips & Gorse, 2017). Social media traces are found to be effective in predicting the stock market behavior, and this makes it an interesting direction for future work in cryptocurrency price prediction research by making use of larger volumes of media input in sentiment analysis, enhancing the baseline natural language-processing models to provide more robust text preprocessing, applying NNs in label training, extending samples in terms of holding period, transaction fees, and user reputation research.
- The application of sentiment analysis for collecting social signals can further be enhanced by improving the quality of the content, and through use of increased number of content sources. Elimination of duplicates and content filtering from advertising or bots could further improve the prediction performance.
- From past studies, it can be noted that the use of content from other social networks, such as Facebook or Reddit (Matta et al., 2015; Mittal et al., 2019; Phillips & Gorse, 2017), can enhance the prediction power and improve the model in this direction, which would also be beneficial. It will be good to take advantage of the best models for standard sentiment analysis and then tune them using transfer learning or any other approach to enhance the prediction power. The other important refinement is to analyze different neural architectures and different data input options integrating media data (e.g., Twitter) by adding other data types that include news and reddit posts from influential users and/or their tweets.
- Another area of opportunity would be the usage of more specialized models that have different types of approaches, such as LSTM networks and temporal MLPs. Recent work proves that the predictability of LSTMs is significantly higher when compared to the generalized regression neural architecture (Hashish et al., 2019; Kwon et al., 2019; Li et al., 2019; Saxena et al., 2018; Wu et al., 2018). Such networks may be able to capture the inherent market trends and easy adapt as needed. Future research can make use of separate models for Twitter and market data in order to improve accuracy and precision scores of models. Also, proving whether these predictive models can be used for creating trading strategies would be interesting.
- Considering the anomalous behavior of cryptocurrency, correlation between cryptocurrency and other assets is still waiting for further work and enhancements. Possible breakthroughs might be achieved with principal component analysis, relationship between cryptocurrency and other currencies in extreme conditions (e.g., financial collapse).
- The public nature of blockchain technology opens the door for new price prediction challenges. Investigations on the relationship between a given currency's transaction network and its price have increased rapidly in recent years; the growing attention on user identification also strongly supports this direction. In-depth knowledge of these networks can help researchers in future to better identify new features in price prediction.
- Another possible attempt is to identify new pricing methods by analyzing real-time market changes. Considering the proportion of informed traders increasing in the cryptocurrency market and the corresponding relation with the pricing process is another area to explore.
- When considering the price of a cryptocurrency, it is always better to look at real-world events. This can include improved technology, future road-map objectives, new partnerships or even regulations. The influence of Lightning-based financial infrastructure, Lightning Labs (which has received investments from senior individuals from Twitter, PayPal, and litecoin) in the performance of bitcoin transactions can be analyzed. Transactions will become much faster and cheaper. If this is successful then it should have a very positive effect on the price of bitcoin. Moreover, regulation policies also influence the bitcoin prices. If major nations follow common regulation policies, it will give bitcoin far more legitimacy. Ultimately, the price of bitcoin should increase. Research analysis in this direction is also recommended.
- Further research with respect to the interaction of market players that make economic decisions based on some indicators, taking into consideration the objective economic laws or behavioral finance laws (econometric and balance models) or by considering the production and other technological possibilities of creating the corresponding asset, such as commodity markets, fundamentals, mining cryptocurrency, and so on, can be encouraged in future.
- Moreover, the research works that involves the model implementation in a practical or real-time setting for predicting into the future can be encouraged.
- Another promising path for future studies could be the linguistic analysis of the coordination of pump-and-dumps in online chat groups, and the means by which misinformation about specific currencies is spread (e.g., on social media) and the influence on cryptocurrency price.

8 | CONCLUSION

Overall, the main contribution of this paper is a recent analysis that explores and summarizes the articles published in the domain of cryptocurrency price prediction, by applying models using traditional statistical and ML techniques. We extracted information from documents published during the periods from 2010 to 2020. To the best of our knowledge, this is the first study on this topic. In this study, we have also investigated various challenges present in traditional approaches of cryptocurrency price prediction. We indicate how we can solve these problems by moving to ML and DL paradigms. Despite the existence of many ML approaches, many issues and challenges still exist in predicting the price of cryptocurrency accurately. The majority of these approaches still require, and are under, further consideration. The discussion presented in this article could be beneficial in exploring future research problems and finding ways to solve them. This study will hopefully give researchers a unique insight for the future and fill the gap in existing studies. The cryptocurrency domain is still in its infancy, as can be seen from the published papers. The majority of the documents are published over the last 3 years or so. And based on the observations, one can safely say, over the next few years, we can expect a lot of studies in this domain. We believe that this paper has provided that foundation by bringing relevant papers and analyzing the contributions, to attract and support future research.

ORCID

Ahmed M. Khedr https://orcid.org/0000-0001-7957-7862

REFERENCES

Abbatemarco, N., De Rossi, L., & Salviotti, G. (2018). An econometric model to estimate the value of a cryptocurrency network. The Bitcoin case. Research Papers. 164. https://aisel.aisnet.org/ecis2018_rp/16

Abu Bakar, N., Rosbi, S., & Uzaki, K. (2019). Forecasting cryptocurrency price movement using moving average method: A case study of bitcoin cash. *International Journal of Advanced Research*, *7*(12), 609–614.

Akcora, C. G., Dey, A. K., Gel, Y. R., & Kantarcioglu, M. (2018). Forecasting bitcoin price with graph chainlets, *Advances in knowledge discovery and data mining. PAKDD 2018*, pp. 765–776.

Akshaya, R., Eswari, B., Dharani, S., & Lalitha, R. (2019). A survey on anticipation the prices of crypto currency using deep learning. *International Journal for Research in Applied Science & Engineering Technology*, *7*(3), 1639–1644.

Alahmari, S. A. (2019). Using machine learning ARIMA to predict the price of cryptocurrencies. *ISeCure—The ISC International Journal of Information Security*, *11*(3), 139–144.

Alessandretti, L., ElBahrawy, A., Aiello, L. M., & Baronchelli, A. (2018). Anticipating cryptocurrency prices using machine learning. *Complexity*, *2018*, 8983590.

Almasri, E., & Arslan, E. (2018). Predicting cryptocurrencies prices with neural networks. In *2018 6th International Conference on Control Engineering & Information Technology (CEIT)*, IEEE, pp. 1–5.

Almeida, J., Tata, S., Moser, A., & Smit, V. (2015). Bitcoin prediction using ANN. *Neural Networks*, *7*, 1–12.

Alon, N., Lokshtanov, D., & Saurabh, S. (2009). Fast FAST, *International colloquium on automata, languages, and programming*. pp. 49–58.

Alonso-Monsalve, S., Suárez-Cetrulo, A. L., Cervantes, A., & Quintana, D. (2020). Convolution on neural networks for high-frequency trend prediction of cryptocurrency exchange rates using technical indicators. *Expert Systems with Applications*, *149*, 113250.

Aloud, M. E. (2020). The role of attribute selection in deep ANNs learning framework for high-frequency financial trading. *Intelligent Systems in Accounting, Finance and Management*, *27*(2), 43–54.

Altan, A., Karasu, S., & Bekiros, S. (2019). Digital currency forecasting with chaotic meta-heuristic bio-inspired signal processing techniques. *Chaos, Solitons & Fractals*, *126*, 325–336.

Alvarez-Ramirez, J., Rodriguez, E., & Ibarra-Valdez, C. (2018). Long-range correlations and asymmetry in the bitcoin market. *Physica A: Statistical Mechanics and its Applications*, *492*, 948–955.

Amjad, M., & Shah, D. (2017). Trading bitcoin and online time series prediction. *Proceedings of Machine Learning Research*, *55*, 1–15.

Anupriya, & Garg, S. (2018). Autoregressive integrated moving average model based prediction of bitcoin close price. *International Conference on Smart Systems and Inventive Technology (ICSSIT)*, pp. 473–478. Available at https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8748423

Ardia, D., Bluteau, K., & Rüede, M. (2019). Regime changes in bitcoin GARCH volatility dynamics. *Finance Research Letters*, *29*, 266–271.

Atsalakis, G. S., Atsalaki, I. G., Pasiouras, F., & Zopounidis, C. (2019). Bitcoin price forecasting with neuro-fuzzy techniques. *European Journal of Operational Research*, *276*(2), 770–780.

Attanasio, G., Garza, P., Cagliero, L., & Baralis, E. (2019). Quantitative cryptocurrency trading: exploring the use of machine learning techniques. In *Proceedings of the 5th Workshop on Data Science for Macro-modelling with Financial and Economic Datasets (DSMM'19)*, pp. 1–6. Available at https://dl.acm.org/doi/pdf/10.1145/3336499.3338003

Badenhorst, J. J. (2018). Effect of bitcoin spot and derivative trading volumes on price volatility. PhD thesis, University of Pretoria, 2019. Available at https://repository.up.ac.za/bitstream/handle/2263/68877/Badenhorst_Effect_2019.pdf?sequence=1&isAllowed=y

Baek, C., & Elbeck, M. (2015). Bitcoins as an investment or speculative vehicle? A first look. *Applied Economics Letters*, *22*(1), 30–34.

Barnwal, A., Bharti, H. P., Ali, A., & Singh, V. (2019). Stacking with neural network for cryptocurrency investment, *2019 New York scientific data summit (NYSDS)*, pp. 1–5.

Bartolucci, S., Destefanis, G., Ortu, M., Uras, N., Marchesi, M., & Tonelli, R. (2020). The butterfly "affect": Impact of development practices on cryptocurrency prices. *EPJ Data Science*, *9*(1), 21.

Bhambhwani, S., Delikouras, S., & Korniotis, G. M. (2019). Do fundamentals drive cryptocurrency prices? Available at https://news.miami.edu/miamiherbert/_news-assets/files/2019/12/crypto.pdf

Blau, B. M. (2017). Price dynamics and speculative trading in bitcoin. *Research in International Business and Finance*, *41*, 493–499.

Böhme, R., Christin, N., Edelman, B., & Moore, T. (2015). Bitcoin: Economics, technology, and governance. *Journal of Economic Perspectives*, *29*(2), 213–238.

Bouri, E., Lau, C. K. M., Lucey, B., & Roubaud, D. (2019). Trading volume and the predictability of return and volatility in the cryptocurrency market. *Finance Research Letters*, *29*, 340–346.

Bouri, E., Shahzad, S. J. H., & Roubaud, D. (2019). Co-explosivity in the cryptocurrency market. *Finance Research Letters*, *29*, 178–183.

Brauneis, A., & Mestel, R. (2019). Cryptocurrency-portfolios in a mean–variance framework. *Finance Research Letters*, *28*, 259–264.

Brooks, C. (2019). *Introductory econometrics for finance*. Cambridge, UK: Cambridge University Press. Available at https://www.cambridge.org/ae/academic/subjects/economics/finance/introductory-econometrics-finance-4th-edition?format=PB&isbn=9781108436823

Bush, R., & Choi, S. (2019). Forecasting ethereum storj token prices: Comparative analyses of applied bitcoin models. In *International Conference on Data Mining Workshops (ICDMW)*. Available at: https://www.computer.org/csdl/proceedings/icdmw/2019/1gAwR0WCegU

Bystrom, H., & Krygier, D. (2018). What drives bitcoin volatility? Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3223368

Caporale, G. M., & Zekokh, T. (2019). Modelling volatility of cryptocurrencies using Markov-switching GARCH models. *Research in International Business and Finance*, *48*, 143–155.

Catania, L., & Sandholdt, M. (2019). Bitcoin at high frequency. *Journal of Risk and Financial Management*, *12*(1), 36.

Chakraborty, D., & Roy, A. (2019). Time series methodology in storj token prediction. In *2019 International Conference on Data Mining Workshops (ICDMW)*, pp. 224–231. Available at https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8955655

Charles, A., & Darné, O. (2019). Volatility estimation for cryptocurrencies: Further evidence with jumps and structural breaks. *Economics Bulletin*, *39*(2), 954–968.

Chatfield, C., & Yar, M. (1988). Holt-winters forecasting: some practical issues. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *37*(2), 129–140.

Chen, M., & Narwal, N. (2017). Predicting price changes in ethereum. Available at http://cs229.stanford.edu/proj2017/final-reports/5244039.pdf

Chen, Z., Li, C., & Sun, W. (2020). Bitcoin price prediction using machine learning: An approach to sample dimension engineering. *Journal of Computational and Applied Mathematics*, *365*, 112395.

Choi, B. (1992). *ARMA model identification*. New York, NY: Springer Science & Business Media. Available at https://www.springer.com/gp/book/9781461397472

Ciaian, P., Rajcaniova, M., & Kancs, A. (2016). The economics of bitcoin price formation. *Applied Economics*, *48*(19), 1799–1815.

Conrad, C., Custovic, A., & Ghysels, E. (2018). Long-and short-term cryptocurrency volatility components: A GARCH–MIDAS analysis. *Journal of Risk and Financial Management*, *11*(2), 23.

Corbet, S., Lucey, B., Urquhart, A., & Yarovaya, L. (2019). Cryptocurrencies as a financial asset: A systematic analysis. *International Review of Financial Analysis*, *62*, 182–199.

Creal, D., Koopman, S. J., & Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, *28*(5), 777–795.

Crypto Research Report (2020). An absolute valuation approach to crypto assets, *Crypto Research Report, June 2020, Edition X*, pp. 6–36.Retrieved from https://cryptoresearch.report/wp-content/uploads/2020/06/Crypto-Research-Report-June-2020-ENG.pdf

Derbentsev, V., Datsenko, N., Stepanenko, O., & Bezkorovainyi, V. (2019). Forecasting cryptocurrency prices time series using machine learning approach. *SHS Web of Conferences*, *65*, 02001.

Dixon, M. F., Halperin, I., & Bilokon, P. (2020). Frontiers of machine learning and finance, *Machine learning in finance: From theory to practice*, pp. 519–541.

Domingos, P. M. (2012). A few useful things to know about machine learning. *Communications of the ACM*, *55*(10), 78–87.

Dos Santos Maciel, L., & Ballini, R. (2019). On the predictability of high and low prices: The case of bitcoin. *Brazilian Review of Finance*, *17*(3), 66–84.

El-Bannany, M., Sreedharan, M., & Khedr, A. M. (2020). A robust deep learning model for financial distress prediction. *International Journal of Advanced Computer Science and Applications*, *11*, 170–175.

Eross, A., McGroarty, F., Urquhart, A., & Wolfe, S. (2019). The intraday dynamics of bitcoin. *Research in International Business and Finance*, *49*, 71–81.

Fang, F., Ventre, C., Basios, M., Kong, H., Kanthan, L., Li, L., & Wu, F. (2020). Cryptocurrency trading: A comprehensive survey. arXiv preprint arXiv:2003.11352.

Galeshchuk, S., & Mukherjee, S. (2017). Deep networks for predicting direction of change in foreign exchange rates. *Intelligent Systems in Accounting, Finance and Management*, *24*(4), 100–110.

Garcia, D., Tessone, C. J., Mavrodiev, P., & Perony, N. (2014). The digital traces of bubbles: feedback cycles between socio-economic signals in the bitcoin economy. *Journal of the Royal Society Interface*, *11*(99), 20140623.

Giudici, P., & Abu-Hashish, I. (2019). What determines bitcoin exchange prices? A network VAR approach. *Finance Research Letters*, *28*, 309–318.

Greaves, A., & Au, B. (2015). Using the bitcoin transaction graph to predict the price of bitcoin. Available via DIALOG. https://pdfs.semanticscholar.org/a0ce/864663c100582805ffa88918910da89add47.pdf. Accessed 8 Dec 2015.

Gunay, S. (2019). Impact of public information arrivals on cryptocurrency market: A case of Twitter posts on ripple. *East Asian Economic Review*, *23*(2), 149–168.

Guo, T., & Antulov-Fantulin, N. (2018). Predicting short-term Bitcoin price fluctuations from buy and sell orders. Preprint. arXiv:1802.04065.

Hashish, I. A., Forni, F., Andreotti, G., Facchinetti, T., & Darjani, S. (2019). September. A hybrid model for bitcoin prices prediction using hidden Markov models and optimized LSTM networks. In *2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, IEEE, pp. 721–728.

Hatefi Ghahfarrokhi, A., & Shamsfard, M. (2020). Tehran stock exchange prediction using sentiment analysis of online textual opinions. *Intelligent Systems in Accounting, Finance and Management*, *27*(1), 22–37.

Hayes, A. S. (2017). Cryptocurrency value formation: An empirical study leading to a cost of production model for valuing bitcoin. *Telematics and Informatics*, *34*(7), 1308–1321.

Hitam, N. A., Ismail, A. R., & Saeed, F. (2019). An optimized support vector machine (SVM) based on particle swarm optimization (PSO) for cryptocurrency forecasting. *Procedia Computer Science*, *163*, 427–433.

Ho, S.-L., Xie, M., & Goh, T. N. (2002). A comparative study of neural network and Box–Jenkins ARIMA modeling in time series prediction. *Computers & Industrial Engineering*, *42*(2–4), 371–375.

Holub, M., & Johnson, J. (2018). Bitcoin research across disciplines. *The Information Society*, *34*(2), 114–126.

Jain, A., Tripathi, S., Dwivedi, H. D., & Saxena, P. (2018). Forecasting price of cryptocurrencies using tweets sentiment analysis, 2018 eleventh international conference on contemporary computing (IC3), pp. 1–7.

Jang, H., & Lee, J. (2018). An empirical study on modeling and prediction of bitcoin prices with Bayesian neural networks based on blockchain information. *IEEE Access*, *6*, 5427–5437.

Ji, S., Kim, J., & Im, H. (2019). A comparative study of bitcoin price prediction using deep learning. *Mathematics*, *7*(10), 898.

Jiang, Z., & Liang, J. (2017). Cryptocurrency portfolio management with deep reinforcement learning. In *2017 Intelligent Systems Conference (IntelliSys)*, IEEE, pp. 905–913.

Kakushadze, Z. (2016). 101 formulaic alphas. *Wilmott*, *2016*(84), 72–81.

Karalevicius, V., Degrande, N., & De Weerdt, J. (2018). Using sentiment analysis to predict interday bitcoin price movements. *The Journal of Risk Finance*, *19*(1), 56–75.

Kaya, Y. (2018). Analysis of cryptocurrency market and drivers of the bitcoin price: Understanding the price drivers of bitcoin under speculative environment (Dissertation). KTH Royal Institute of Technology, Stockholm, Swedon. Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-246060

Kim, H. M., & Laskowski, M. (2018). Toward an ontology-driven blockchain design for supply-chain provenance. *Intelligent Systems in Accounting, Finance and Management*, *25*(1), 18–27.

Kim, Y. B., Kim, J. G., Kim, W., Im, J. H., Kim, T. H., Kang, S. J., & Kim, C. H. (2016). Predicting fluctuations in cryptocurrency transactions based on user comments and replies. *PloS ONE*, *11*(8), e0161197. https://doi.org/10.1371/journal.pone.0161197

Kjærland, F., Khazal, A., Krogstad, E. A., Nordstrøm, F. ransB. G., & Oust, A. (2018). An analysis of bitcoin's price dynamics. *Journal of Risk and Financial Management*, *11*(4), 63.

Kleinbaum, G., & Klein, M. (2002). *Logistic regression*. New York, NY: Springer.

Krištoufek, L. (2015). What are the main drivers of the bitcoin price? Evidence from wavelet coherence analysis. *PLoS ONE*, *10*(4), e0123923.

Kwon, D.-H., Kim, J.-B., Heo, J.-S., Kim, C.-M., & Han, Y.-H. (2019). Time series classification of cryptocurrency price trend based on a recurrent

LSTM neural network. *Journal of Information Processing Systems*, *15*(3), 694–706.

Kyriazis, N., Papadamou, S., & Corbet, S. (2020). A systematic review of the bubble dynamics of cryptocurrency prices. *Research in International Business and Finance*, *54*, 101254.

Kyriazis, N. A. (2019). A survey on efficiency and profitable trading opportunities in cryptocurrency markets. *Journal of Risk and Financial Management*, *12*(2), 67.

Lahmiri, S. (2011). A comparison of PNN and SVM for stock market trend prediction using economic and technical information. *International Journal of Computer Applications*, *29*(3), 24–30.

Lahmiri, S., & Bekiros, S. (2019). Cryptocurrency forecasting with deep learning chaotic neural networks. *Chaos, Solitons & Fractals*, *118*, 35–40.

Längkvist, M., Karlsson, L., & Loutfi, A. (2014). A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters*, *42*, 11–24.

Lee, T.-H., & Yang, W. (2014). Granger-causality in quantiles between financial markets: Using copula approach. *International Review of Financial Analysis*, *33*, 70–78.

Li, L., Arab, A., Liu, J., Liu, J., & Han, Z. (2019). Bitcoin options pricing using LSTM-based prediction model and blockchain statistics, *2019 IEEE international conference on blockchain (blockchain)*, pp. 67–74.

Li, T. R., Chamrajnagar, A., Fong, X., Rizik, N., & Fu, F. (2019). Sentiment-based prediction of alternative cryptocurrency price fluctuations using gradient boosting tree model. *Frontiers in Physics*, *7*, 98.

Liu, Y., & Tsyvinski, A. (2020). Risks and returns of cryptocurrency. *The Review of Financial Studies*, hhaa113. https://doi.org/10.1093/rfs/hhaa113h

Lu, C.-J. (2010). Integrating independent component analysis-based denoising scheme with neural network for stock price prediction. *Expert Systems with Applications*, *37*(10), 7056–7064.

Madan, I., Saluja, S., & Zhao, A. (2015). Automated bitcoin trading via machine learning algorithms.

Mallqui, D. C. A., & Fernandes, R. A. S. (2019). Predicting the direction, maximum, minimum and closing prices of daily bitcoin exchange rate using machine learning techniques. *Applied Soft Computing*, *75*, 596–606.

Matta, M., Lunesu, I., & Marchesi, M. (2015). Is bitcoin's market predictable? Analysis of web search and social media, *Knowledge discovery, knowledge engineering and knowledge management. IC3K 2015*, pp. 155–172.

McNally, S., Roche, J., & Caton, S. (2018). Predicting the price of bitcoin using machine learning, 2018 26th Euromicro international conference on parallel, distributed and network-based processing (PDP), pp. 339–343.

Misnik, A., Krutalevich, S., Prakapenka, S., Borovykh, P., & Vasiliev, M. (2018). Neural network approximation precision change analysis on cryptocurrency price prediction. In *Fuzzy Technologies in the Industry-FTI 2018*, pp. 96–101.

Mittal, A., Dhiman, V., Singh, A., & Prakash, C. (2019). Short-term bitcoin price fluctuation prediction using social media and web search data. In *2019 Twelfth International Conference on Contemporary Computing (IC3)*, IEEE, pp. 1–6.

Mittal, R., Arora, S., & Bhatia, M. P. S. (2018). Automated cryptocurrencies prices prediction using machine learning. *ICTACT Journal on Soft Computing*, *8*, 2229–6956.

Mnif, E., Jarboui, A., Hassan, M. K., & Mouakhar, K. (2020). Big data tools for Islamic financial analysis. *Intelligent Systems in Accounting, Finance and Management*, *27*(1), 10–21.

Mohamed, K., Aziz, A., Mohamed, B., Abdel-Hakeem, K., Mostafa, M., & Atia, A. (2019). Blockchain for tracking serial numbers in money exchanges. *Intelligent Systems in Accounting, Finance and Management*, *26*(4), 193–201.

Mohanty, P., Patel, D., Patel, P., & Roy, S. (2018). Predicting fluctuations in cryptocurrencies' price using users' comments and real-time prices. In *2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, IEEE, pp. 477–482.

Nakamoto, S., & Bitcoin, A. (2008). Bitcoin: A peer-to-peer electronic cash system. Bitcoin. Retrieved from https://bitcoin.org/bitcoin.pdf

Nikou, M., Mansourfar, G., & Bagherzadeh, J. (2019). Stock price prediction using deep learning algorithm and its comparison with machine learning algorithms. *Intelligent Systems in Accounting, Finance and Management*, *26*(4), 164–174.

O'Leary, D. E. (2017). Configuring blockchain architectures for transaction information in blockchain consortiums: The case of accounting and supply chain systems. *Intelligent Systems in Accounting, Finance and Management*, *24*(4), 138–147.

O'Leary, D. E. (2018). Open information enterprise transactions: Business intelligence and wash and spoof transactions in blockchain and social commerce. *Intelligent Systems in Accounting, Finance and Management*, *25*(3), 148–158.

O'Leary, D. E. (2019). Some issues in blockchain for accounting and the supply chain, with an application of distributed databases to virtual organizations. *Intelligent Systems in Accounting, Finance and Management*, *26*(3), 137–149.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, *42*(1), 259–268.

Peng, Y., Albuquerque, P. H. M., de Sá, J. M. C., Padula, A. J. A., & Montenegro, M. R. (2018). The best of two worlds: Forecasting high frequency volatility for cryptocurrencies and traditional currencies with support vector regression. *Expert Systems with Applications*, *97*, 177–192.

Phaladisailoed, T., & Numnonda, T. (2018). Machine learning models comparison for bitcoin price prediction. In *2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE)*, IEEE, pp. 506–511.

Phillips, R. C., & Gorse, D. (2017). Predicting cryptocurrency price bubbles using social media data and epidemic modelling, *2017 IEEE symposium series on computational intelligence (SSCI)*, pp. 1–7.

Phillips, R. C., & Gorse, D. (2018a). Cryptocurrency price drivers: Wavelet coherence analysis revisited. *PloS ONE*, *13*(4), e0195200.

Phillips, R. C., & Gorse, D. (2018b). Mutual-excitation of cryptocurrency market returns and social media topics, *ICFET '18: Proceedings of the 4th International conference on frontiers of educational technologies*, pp. 80–86.

Poongodi, M., Sharma, A., Vijayakumar, V., Bhardwaj, V., Sharma, A. P., Iqbal, R., & Kumar, R. (2020). Prediction of the price of ethereum blockchain cryptocurrency in an industrial finance system. *Computers & Electrical Engineering*, *81*, 106527.

Poyser, O. (2019). Exploring the dynamics of bitcoin's price: A Bayesian structural time series approach. *Eurasian Economic Review*, *9*(1), 29–60.

Radityo, A., Munajat, Q., & Budi, I. (2017). Prediction of bitcoin exchange rate to American dollar using artificial neural network methods. In *2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, IEEE, pp. 433–438.

Rane, P. V., & Dhage, S. N. (2019). Systematic erudition of bitcoin price prediction using machine learning techniques. In *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, IEEE, pp. 594–598.

Rebane, J., Karlsson, I., Papapetrou, P., & Denic, S. (2018). Seq2Seq RNNs and ARIMA models for cryptocurrency prediction: A comparative study. In *SIGKDD Fintech'18, London, UK*, August 19–23.

Rivas, V. M., Parras-Gutiérrez, E., Merelo, J. J., Arenas, M. G., & García-Fernández, P. (2017). Time series forecasting using evolutionary neural nets implemented in a volunteer computing system. *Intelligent Systems in Accounting, Finance and Management*, *24*(2-3), 87–95.

Roy, S., Nanjiba, S., & Chakrabarty, A. (2018). Bitcoin price forecasting using time series analysis. In *2018 21st International Conference of Computer and Information Technology (ICCIT)*, IEEE, pp. 1–5.

Sarlin, P., & Marghescu, D. (2011). Neuro-genetic predictions of currency crises. *Intelligent Systems in Accounting, Finance and Management*, *18*(4), 145–160.

Saxena, A., Sukumar, T., Nadu, T., & Nadu, T. (2018). Predicting bitcoin price using LSTM and compare its predictability with ARIMA model. *International Journal of Pure and Applied Mathematics*, *119*(17), 2591–2600.

Shah, D., & Zhang, K. (2014). Bayesian regression and bitcoin. In *2014 52nd annual Allerton conference on communication, control, and computing (Allerton)*, IEEE, pp. 409–414.

Silva de Souza, M. J., Almudhaf, F. W., Henrique, B. M., Negredo, A. B. S., Ramos, D. G. F., Sobreiro, V. A., & Kimura, H. (2019). Can artificial intelligence enhance the bitcoin bonanza. *The Journal of Finance and Data Science*, *5*(2), 83–98.

Sin, E., & Wang, L. (2017). Bitcoin price prediction using ensembles of neural networks. In *2017 13th International conference on natural computation, fuzzy systems and knowledge discovery (ICNC-FSKD)*, IEEE, pp. 666–671.

Snihovyi, O., Ivanov, O., & Kobets, V. (2018). Cryptocurrencies prices forecasting with anaconda tool using machine learning techniques. In *CEUR Workshop Proceedings*, Vol. *2105*, pp. 453–456.

Sovbetov, Y. (2018). Factors influencing cryptocurrency prices: Evidence from bitcoin, ethereum, dash, litcoin, and monero. *Journal of Economics and Financial Analysis*, *2*, 1–27.

Sovia, R., Yanto, M., Budiman, A., Mayola, L., & Saputra, D. (2019). Backpropagation neural network prediction for cryptocurrency bitcoin prices. *Journal of Physics: Conference Series*, *1339*, 012060.

Sreedharan, M., Khedr, A. M., & El-Bannany, M. (2020a). A multi-layer perceptron approach to financial distress prediction with genetic algorithm. *Automatic Control and Computer Sciences*, *57*, 475–482.

Sreedharan, M., Khedr, A. M., & El-Bannany, M. (2020b). A comparative analysis of machine learning classifiers and ensemble techniques in financial distress prediction. In *2020 17th International Multi-Conference on Systems, Signals & Devices (SSD'20)*, pp. 653-657.

Stosic, D., Stosic, D., Ludermir, T. B., & Stosic, T. (2018). Collective behavior of cryptocurrency price changes. *Physica A: Statistical Mechanics and its Applications*, *507*, 499–509.

Sun, J., Zhou, Y., & Lin, J. (2019). Using machine learning for cryptocurrency trading. In *2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS)*, IEEE, pp. 647–652.

Sun, X., Liu, M., & Sima, Z. (2020). A novel cryptocurrency price trend forecasting model based on LightGBM. *Finance Research Letters*, *32*, 101084.

Trimborn, S., & Härdle, W. K. (2019). CRIX an index for cryptocurrencies. *Journal of Empirical Finance*, *49*, 107–122.

Troster, V., Tiwari, A. K., Shahbaz, M., & Macedo, D. N. (2019). Bitcoin returns and risk: A general GARCH and GAS analysis. *Finance Research Letters*, *30*, 187–193.

Tupinambás, T. M., Cadence, R. A. L., & Lemos, A. P. (2018). Cryptocurrencies transactions advisor using a genetic Mamdani-type fuzzy rules based system. In *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, IEEE, pp. 1–7.

Uras, N., Marchesi, L., Marchesi, M., & Tonelli, R. (2020). Forecasting Bitcoin closing price series using linear regression and neural networks models. *Peer J Computer Science*, *6*, e279.

Vaddepalli, S., & Antoney, L. (2018). Are economic factors driving bitcoin transactions? An analysis of select economies. *Finance Research Letters*, *163*(12), 106–109.

Valencia, F., Gómez-Espinosa, A., & Valdés-Aguirre, B. (2019). Price movement prediction of cryptocurrencies using sentiment analysis and machine learning. *Entropy*, *21*(6), 589.

Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, *10*(5), 988–999.

Vieira, P. J. M. (2017). Price analysis of bitcoin: Volatility, key drivers and evolution.

Virk, D. S. (2017). Prediction of bitcoin price using data mining (Doctoral dissertation). National College of Ireland, Dublin.

Walther, T., Klein, T., & Bouri, E. (2019). Exogenous drivers of bitcoin and cryptocurrency volatility—A mixed data sampling approach to forecasting. *Journal of International Financial Markets, Institutions and Money*, *63*, 101133.

Wang, Y., & Chen, R. (2020). Cryptocurrency price prediction based on multiple market sentiment. In *Proceedings of the 53rd Hawaii International Conference on System Sciences*.

Wiedmer, J. (2018). The price of cryptocurrencies: an empirical analysis (MSc thesis). Univeritát Bern, Bern, Switzerland.

Wu, C. H., Lu, C. C., Ma, Y. F., & Lu, R. S. (2018). November. A new forecasting framework for bitcoin price with LSTM. In *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, IEEE, pp. 168–175.

Yao, Y., Yi, J., Zhai, S., Lin, Y., Zhang, T. K. G., & Yoonjae, L. (2018). Predictive analysis of cryptocurrency price using deep learning. *International Journal of Engineering & Technology*, *7*(3.27), 258–264.

Yaya, O. S., Ogbonna, E. A., & Olubusoye, O. E. (2018). How persistent and dependent are pricing of bitcoin to other cryptocurrencies before and after 2017/18 crash?.

Zamuda, A., Crescimanna, V., Burgillo, J. C., Dias, J. M., Wegrzyn-Wolska, K., Rached, I., … Bracciali, A. (2019). Forecasting cryptocurrency value by sentiment analysis: An HPC-oriented survey of the state-of-the-art in the cloud era. In J. Kołodziej & H. González-Vélez (Eds), *High-performance modelling and simulation for big data applications*, Lecture Notes in Computer Science (Vol. 11400), Cham: Springer. https://doi.org/10.1007/978-3-030-16272-6_12

Zhang, L., & Wang, J. (2015). Optimizing parameters of support vector machines using team-search-based particle swarm optimization. *Engineering Computations*, *32*(5), 1194–1213.

Zhang, Y., Chan, S., Chu, J., & Nadarajah, S. (2019). Stylised facts for high frequency cryptocurrency data. *Physica A: Statistical Mechanics and Its Applications*, *513*, 598–612.

How to cite this article: Khedr A, Arif I, P V PR, El-Bannany M, Alhashmi S, S M. Cryptocurrency price prediction using traditional statistical and machine learning techniques: A survey. *Intell Sys Acc Fin Mgmt*. 2021;28:3–34. https://doi.org/10.1002/isaf.1488