IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 11, 2020, accepted September 11, 2020, date of publication September 21, 2020, date of current version October 6, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3025211

Cryptocurrencies and Artificial Intelligence: Challenges and Opportunities

FARIDA SABRY[1], (Member, IEEE), WADHA LABDA[1], (Associate Member, IEEE), AIMAN ERBAD[2], (Senior Member, IEEE), AND QUTAIBAH MALLUHI[1], (Member, IEEE)

[1]Department of Computer Science and Engineering, Qatar University, Doha, Qatar
[2]Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar

Corresponding author: Farida Sabry (faridasabry@ieee.org)

The research in this article was done for project number NPRP X-063-1-014 under a grant from Qatar National Research Fund (QNRF). However, the contents of the research are solely the responsibility of the authors and do not necessarily represent the official views of QNRF. Open access funding was provided by Qatar National Library.

ABSTRACT Decentralized cryptocurrencies have gained a lot of attention over the last decade. Bitcoin was introduced as the first cryptocurrency to allow direct online payments without relying on centralized financial entities. The use of Bitcoin has vastly grown as a financial asset rather than just a tool for online payments. A lot of cryptocurrencies have been created since 2011 with Bitcoin dominating the cryptocurrencies' market. With plenty of cryptocurrencies being used as financial assets and with millions of trades being executed through different exchange services, cryptocurrencies are susceptible to trading problems and challenges similar to those traditionally encountered in the financial domain. Price and trend prediction, volatility prediction, portfolio construction and fraud detection are some examples related to trading. In addition, there are other challenges that are specific to the domain of cryptocurrencies such as mining, cybersecurity, anonymity and privacy. In this paper, we survey the application of artificial intelligence techniques to address these challenges for cryptocurrencies with their vast amount of daily transactions, trades and news that are beyond human capabilities to analyze and learn from. This paper discusses the recent research work done in this emerging area and compares them in terms of used techniques and datasets. It also highlights possible research gaps and some potential areas for improvement.

INDEX TERMS Artificial intelligence, machine learning, deanonymization, price prediction, fraud detection, volatility prediction, anonymity, privacy, mining, security.

I. INTRODUCTION

This paper surveys research work applying artificial intelligence and machine learning techniques in the field of cryptocurrencies. Analyzing cryptocurrencies is considered a relatively recent domain that became active in the last decade. Bitcoin was announced at the end of 2008 as the first *decentralized* cryptocurrency that relies heavily on the field of cryptography for hashing and signing transactions. These transactions are committed to a distributed blockchain ledger to be synced and verified by nodes in a peer-to-peer network. The Bitcoin blockchain size reached over 280 GB in June, 2020.

With this big data representing transactions in the blockchain coupled with millions of trades being executed on different exchange websites, growing number of tweets, posts and articles related to Bitcoin and cryptocurrencies, there is a clear need for automated tools to process and analyze this big data. Artificial intelligence (AI) techniques can learn from this massive amount of data by analyzing and discovering patterns to ease and secure trading and mining. Discovering patterns in money-laundering transactions and other fraudulent transactions and trading schemes can help limit the crimes involving cryptocurrencies due to privacy and security threats they encounter. Artificial intelligence (AI) techniques are not limited to machine learning (ML) techniques (*supervised, unsupervised, semi-supervised,* and *reinforcement*), but also include evolutionary-based techniques and knowledge-based techniques [1].

The topic of data analytics for cryptocurrencies is gaining importance as more entities are becoming more reliant on cryptocurrencies. Our work differs from previous related surveys by focusing on the research of using artificial intelligence and machine learning techniques in cryptocurrencies as digital currencies or crypto-assets and the surrounding ecosystem. Hassani *et al.* [2] studied the interactions between

The associate editor coordinating the review of this manuscript and approving it for publication was Xiaochun Cheng.

Big Data and cryptocurrency focusing on two perspectives, namely, ‘‘security and privacy enhancement’’, and ‘‘prediction and analysis’’. They also used the word ‘‘cryptocurrencies’’ for its underlying *blockchain technology* and its applications and not for cryptocurrencies as digital currencies like in our work. Other authors in [3], [4] and [5] presented surveys on blockchain applications in AI and robotics. None of the mentioned surveys have explored the application of AI techniques to tackle cryptocurrencies’ challenges.

Our survey considers papers from reputable peer-reviewed conferences and journals. Papers that include mere descriptive statistics or discuss the use of AI in other blockchain applications like healthcare, smart contracts and IoT were excluded. The following exploratory questions form the basis for this paper:

- What are the problems in the cryptocurrencies domain that has been approached using AI techniques?
- Which AI techniques have been studied in the literature and employed in the field of cryptocurrencies, and what are the datasets used in this domain?
- What are the possible research gaps and areas of improvement that can be further studied?

The rest of this paper is organized as follows. Section II provides a background about Bitcoin and cryptocurrencies. Section III illustrates research problems related to the cryptocurrencies domain and reviews artificial intelligence and machine learning research addressing these problems. Section IV discusses possible research gaps and potential areas of improvement and conclusion of the paper is in Section V.

II. BACKGROUND

Bitcoin, the cryptocurrency with the highest market capitalization, has been circulating for more than a decade since January 3rd, 2009. Since then, Bitcoin has demonstrated tremendous success as a financial asset and there are currently around 18 million Bitcoins (BTCs) being traded and exchanged. Bitcoin, however, does not resemble other traditional assets from an econometric perspective [6].

On the technical level, Bitcoin depends on a *decentralized* peer-to-peer network that aims to replace the centralized financial system. In this decentralized network, a distributed ledger saves all the transactions taking place in a *blockchain* structure. This blockchain is synced by all nodes in the network to verify the transactions. Transactions are added to the blockchain in blocks by miners after they compete to solve a cryptographic puzzle.

Bitcoin uses *elliptic curve digital signature algorithm* (ECDSA) to sign a transaction between sender’s and receiver’s Bitcoin addresses. Bitcoin addresses are identifiers of 26-35 alphanumeric characters, formed by hashing of the public keys of the sender or receiver. Using these addresses made Bitcoin transactions look anonymous, however they are actually pseudo-anonymous. SHA-256 hashing is also used in Bitcoin’s proof-of-work consensus strategy. Proof-of-work (PoW) is needed by the blockchain to secure the distributed ledger against any tampering attempts. This is achieved through solving a computationally-intensive cryptographic problem that is hard to solve but easy to verify. The nodes compete to find a nonce that results in a block hash with a certain number of leading zeroes. The winner node adds a block of transactions to the blockchain and receives a block reward set by the Bitcoin protocol in addition to transaction fees from senders. The block reward was initially set to 50 BTC and it halves after every 210,000 blocks are added to the chain. This halving process happens nearly every four years. The block reward, at the time of writing this paper is 6.25 BTC. This process is called *mining* new Bitcoins. In the early days of Bitcoin, the difficulty of the PoW problem was relatively easy and Bitcoin mining was done using a personal computer with good CPU. As the difficulty increases, higher hash rates are needed to solve the crypto-puzzle. Machines with higher computing power that use graphics processing units (GPUs) and field-programmable gate array (FPGAs) were used for mining new coins. Currently, mining pools are of very high processing power and most of them use application-specific integrated circuits (ASICs), which are special hardware circuits designed specifically for Bitcoin mining.

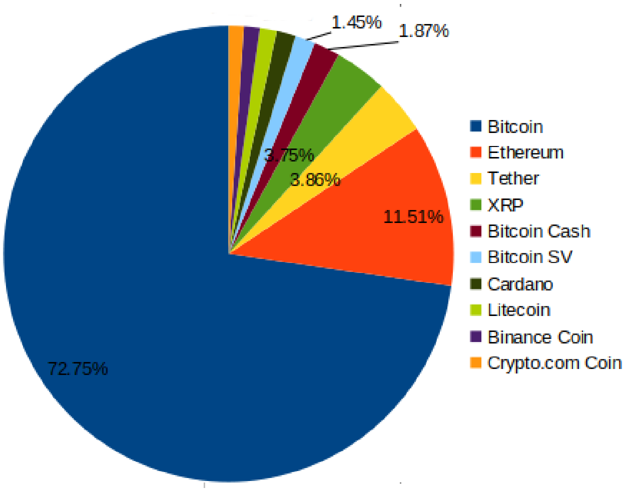

FIGURE 1. Market capitalization of cryptocurrencies as per CoinMarketCap in July 2020.

There are other alternative cryptocurrencies (altcoins) that offer greater speed, anonymity or some other advantages over Bitcoin. Namecoin was the first altcoin created as an attempt to improve some features like decentralization, security, privacy, and DNS speed, etc. Other examples are Litecoin, Ripple, Ethereum, Zcash, Monero, BCH, Dash, etc. Currently, there are more than 2000 cryptocurrencies as per CoinMarketCap.[1] However, Bitcoin dominates the market with more than 70% of the market share as shown in Figure 1 as per the market capitalizations reported by CoinMarketCap in July, 2020.

[1] https://coinmarketcap.com/

Decentralization allows cryptocurrencies to be more immune to government control and interference. On the other hand, as a disadvantage for the lack of mature regulations in the cryptocurrencies market; cryptocurrencies have been used in the dark web for money-laundering, weapons and drug dealings, and other criminal activities. Many countries have recently been forcing know-your-customer (KYC) and anti-money laundering (AML) policies on exchanges while conducting research to have their government-regulated cryptocurrencies and waiting for regulations acceptance. Similarly, Facebook Libra Coin is waiting for the necessary regulatory acceptance and was planned to launch in 2020. The Libra coin, governed by an ''*independent*'' association, is designed to overcome the problems of volatility and lack of scalability in existing cryptocurrencies.

This noticeably growing market of cryptocurrencies brings about the problem of analyzing the tremendous amount of trades and transactions taking place for different cryptocurrencies through different exchanges and over different blockchains. Artificial intelligence (AI) is a good candidate for solving problems involving this immense amount of data that humans can not analyze efficiently.

Techniques and models used in artificial intelligence tasks can be categorized into machine learning techniques, evolutionary-based techniques, knowledge-based techniques and other techniques in which machines think and act humanly and rationally as per the typical definition of AI [1]. There are many key players with various needs that can be inferred from the cryptocurrencies ecosystem (see Figure 2), e.g. traders (including criminals), regulatory agents, miners, security specialists, etc. Automated traders can analyze and learn from the cryptocurrencies' market prices and markers for taking trading decisions and achieving high returns for their owners. Regulatory institutions can use AI to learn from the data about possible financial frauds and potential threats. Miners can make use of AI techniques to increase their profit and save electricity for environmental considerations. Security specialists can use these techniques to analyze and assess the security and privacy level of cryptocurrencies and spot possible pitfalls and threats.

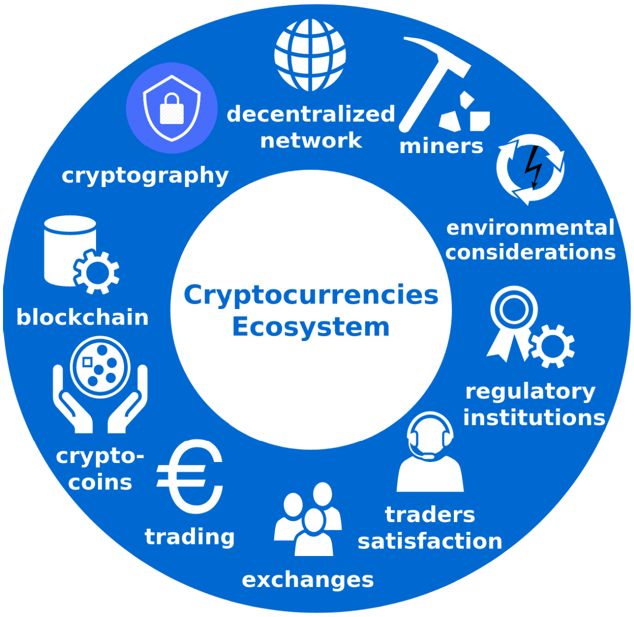

FIGURE 2. Cryptocurrencies ecosystem.

III. ARTIFICIAL INTELLIGENCE RESEARCH IN CRYPTOCURRENCIES

Cryptocurrencies face similar challenges to fiat currencies' and other financial market assets' challenges. According to Business Insider Report[2] in June 2019, there are three areas where AI techniques are used in banking, namely, conversational banking, anti-fraud detection, risk assessment and credit underwriting. Additionally, financial chatbots and voice assistants that mimic live employees, deepen customer relationships and provide personalized insights and recommendations, are examples of software systems using AI in the financial field. Moreover, AI is extensively used in intelligent trading systems to do stock market prediction and currency price prediction. This helps in taking decisions on when to buy, hold or sell a stock based on different markers that change over time. Furthermore, anti-fraud detection tasks make use of machine learning to learn from spending behaviors and patterns and detect suspicious patterns.

Trading cryptocurrencies shares with the financial market the aforementioned problems. Using artificial intelligence in cryptocurrencies, like in financial services, reduces the risk of human error and speeds up the process of trading by predicting the value of the currency or its rise and fall over time. In addition, there are some other challenges that are specific to cryptocurrencies that can be deduced from the cryptocurrencies ecosystem in Figure 2 and to which AI techniques can offer useful solutions.

From the reviewed papers in this domain, we can summarize the categories for challenges facing cryptocurrencies in a taxonomy shown in Figure 3. There are challenges that are related to the trading process like price and trend prediction, volatility prediction, portfolio construction, fraud detection and other analysis tasks to get insights and indicators about different cryptocurrencies. Trading bots do all of these tasks for trading cryptocurrencies. These challenges involve using machine learning techniques to learn from historical data of prices, other market indicators and social media interests to take profitable trading decisions. Additionally, natural language processing (NLP) -which involves using many AI techniques- is needed for sentiment analysis and processing of news, social media posts (e.g. Twitter, Facebook, Telegram [7], LinkedIn, Reddit, etc) and domain forums like BitcoinTalk[3] and Ethereum Community Forum.[4] The use of word2vec-based topic modeling and other NLP techniques

[2] https://www.businessinsider.com/the-ai-in-banking-report-2019-6
[3] http://bitcointalk.org
[4] https://forum.ethereum.org/

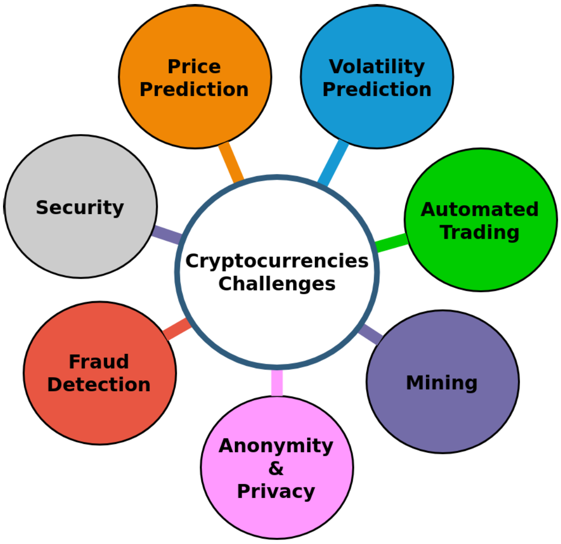

FIGURE 3. Challenges in cryptocurrencies.

was demonstrated in [8] to analyze Reddit submissions' topics temporally with the price shifts occurring during 2017 and 2018. NLP can be also useful if a trading bot is built to be a conversational bot to ease the trading experience. Moreover, NLP is a core component in the design of chatbots replying to queries and questions about cryptocurrencies as proposed in [9].

Other challenges include mining of crypto-coins, anonymity and privacy of cryptocurrency transactions, and security of the cryptocurrency peer-to-peer network, users' wallets and exchange services. Transactions' delay is a problem faced by cryptocurrencies, and specifically Bitcoin, as it takes some time for transactions to be approved on the various chains and confirmed. Transaction confirmation delay depends on many factors in the mining process so we will classify this problem among the mining challenges. Blockchain size and its high degree of replication is also considered as a challenge to cryptocurrencies but it is mainly considered as a challenge with the protocol definition and the data storage technology which is not typically tackled by AI techniques. The research on applying *AI techniques* in the field of cryptocurrencies has grown in the last four years especially after the price hype in 2017.

The next subsections briefly describe the cryptocurrencies' challenges identified in Figure 3 that can be tackled by AI techniques. Classification of the research work is done according to the challenge it addresses. The main research papers in each category are reviewed briefly. There exist a large amount of published papers particularly for tackling the price forecasting/prediction problem. Therefore, we could not be exhaustive in citing every piece of work that was done. Nevertheless, we chose to include papers that use diverse techniques of AI and that can contribute to answering the second question of this survey as mentioned in section I, related to the AI techniques employed in the field of cryptocurrencies and the datasets used in this domain.

A. PRICE PREDICTION/FORECASTING

The attention to Bitcoin and other cryptocurrencies as financial assets has grown dramatically in the last three years. This was after the hype occurred by the end of 2017 when the price of Bitcoin reached about 20,000 USD as shown in Figure 6.

For trading cryptocurrencies, domain observers and traders need to do Bitcoin/cryptocurrency analytics and to predict the cryptocurrency price. The terms ''*price prediction*'' and ''*price forecasting*'' are usually used in the same way to refer to the task of predicting an estimate for the price based on the history of past prices and other explanatory variables. The term ''prediction'' is more general as it refers to either prediction of the current or future prices while forecasting is used to refer to making estimates about the future prices or trends. The term ''*prediction*'' is widely used by researchers so it is adopted in the rest of this survey.

The price of Bitcoin can be affected by many factors (sometimes called indicators/markers/features or variables), among them are the interaction between supply, demand and attractiveness for investors. These factors are usually affected by trends in social networks, forums, search engines, declarations by leaders and political stability of countries. Past fluctuations in cryptocurrency's price or trades' growth/decline can be used to determine possible trends and predict what could happen in the future. Other factors that might affect cryptocurrency price are: other cryptocurrencies' prices, blockchain data, the gold price, silver price, oil price, stock market variables like S&P 500 index (Standard and Poor 500 index that measures the performance of the stocks of 500 big-size companies in the U.S. stock exchanges), and other *financial technical indicators* for cryptocurrency and other stock markets such as those used in [10]. *Online factors* that represent to an extent the public adoption and awareness of Bitcoin and other cryptocurrencies such as Reddit posts, Wikipedia views, Google trends, etc., were used by some researchers [11]–[16]. Table 1 shows the different factors used in literature indicating whether positive (+) or negative (-) correlation with the Bitcoin prices is reported and whether it is reported as a significant correlation (*).

The basic flow of most of the work done in this area starts with the collection of time-series data for different variables of concern. Analysis of the data and relationships between different variables and the cryptocurrency price are then deduced. A supervised machine learning technique is used to learn a model from data which can then be used for prediction. Using history of different variables makes price prediction a time-series prediction task. It can be modeled as a regression problem to predict the closing price based on a set of indicators. It can also be modeled as a classification problem to predict if there will be a rise/fall or no change in the price of a coin by encoding the cryptocurrency price time series output variable in terms of rise and fall. A linear

TABLE 1. Different factors used in cryptocurrency price prediction. (+: indicates positive correlation reported, -: indicates negative correlation reported, *: indicates significant correlation reported.)

	Factor	Used in
Cryptocurrency market	History of BTC prices	[11]* [17] [15] [18] [19] [13] [20]* [21] [22] [23] [24] [14] [25] [26] [16] [27]
	Other crypto prices	[17] [21] [22] [23] [14] [26] [19]
	Total number of coins	[20]
	Market capitalization	[20] [22] [23]
	Trade volume	[20]* [25] [26] [16]
	Market spread	[25]
	Price momentum	[25]
Other markets indicators	Exchange prices between USD and EUR	[11]-
	USD exchange prices for GBP, CNY,JPY,CHF	[26]
	Gold price	[17] [15] [19]
	Silver price	[22]
	Crude oil price	[15] [22] [26]
	S&P 500 index	[11]-* [12] [15] [19]* [22] [26]
	Volatility Index (VIX)	[17]- [19]*
	MSCI World Index (MXWO)	[17]+ [22]
	MSCI Emerging Markets Index(MXEF)	[17]- [22]
	USD index	[17]+ [19]
	Bloomberg Commodity Index (BCOM)	[17]+
	Value Weighted Index (VW)	[17]+
	S&P Bond Index	[19]
	GSCI Commodity Total Return	[19]
	Neikkei225	[26]
	Dow Jones	[22] [26]
	NASDAQ	[26]
	Eurozone stocks Index (Eurostoxx)	[26]
	Financial Times Stock Exchange Index (FTSE100)	[26]
Blockchain variables	Hashrate	[11]+ [15] [28] [20] [22] [26]
	BTC volume	[11]+* [15]-* [13] [20]
	No. of transactions	[11]-* [13] [28] [22] [25] [26] [16] [20]
	Difficulty	[28] [20] [22] [26]
	Miner’s revenue	[28] [20] [22] [26]
	Average block size	[28] [20] [26]
	Blockchain size	[20]
	Number of transactions per second	[20]
	Number of transactions per block	[20]
	Average transaction fees	[20] [22]
	Confirmation time	[26] [20]
	Cost per transaction	[20]*
	Total output value	[20]
	Unspent transaction output (UTXO) count	[20]
	Number of unique BTC addresses	[20]
Online variables	Negative Shock	[15]-**
	Positive Shock	[15]+*
	Wikipedia views	[11]+
	Twitter volume	[11] [12]+ [22]
	Twitter sentiment	[11]* [12] [16]
	Google search (SVI)	[12]+ [15]+* [13] [22]
	bitcointalk.org forum articles	[13] [14]

time series model for cryptocurrency price can be generalized by Equation 1, where y_t is the cryptocurrency price response variable, X is the input time series of explanatory variables and indicators such as those in Table 1, and k is the order of the auto regressive model to represent a dependency on the history of both the input variables and the history of the price.

$$y_t = \sum_{i=1}^{k} \theta_i y_{t-i} + \sum_{i=1}^{k} \omega_i X_{t-i+1} + \omega_0 \qquad (1)$$

Different regression models and statistical models have been tested by researchers to fit this linear model in different ways [11], [12], [15], [17]–[19]. Simple linear time series models sometimes leave certain aspects of economic and financial data unexplained [29]. That is why some authors tested nonlinear time series models to model nonlinear behavior in economic and financial time series data [13], [14], [16], [20]–[28].

Table 2 shows a summary for the research on price prediction of cryptocurrencies comparing the used AI techniques and datasets in each paper. They are ordered by the year of publication. Certainly, there are more papers published in this very active area of predicting and analyzing cryptocurrency prices. We only included recent publications with a set of diverse AI techniques being tested. As it can be seen from Table 2, each study depended on a different dataset with different features over different time periods for training and testing. Additionally, they used different evaluation metrics for evaluating the results. Geometric mean return and the Sharpe ratio were used in [23]. Young *et al.* [13] used accuracy, F1-Score and Matthews correlation coefficient (MCC). Mean square error (MSE), root mean square error (RMSE) and mean absolute error (MAE) were used in [21], [25]. Mean absolute percentage error (MAPE) was used in [20], [22], [26]. Accuracy, recall, precision and F1 Score were used in [16], [20], [27]. Moreover, the results can not be easily replicated due to the different models' parameters that are not thoroughly mentioned. For these reasons, it is a very difficult task to recommend a certain model over the others as a state-of-the-art for price prediction. However, at the end of this section we will give summarized observations based on the results reported in the referenced papers.

We have classified some of the state-of-art research done in this area according to the main model(s) type it uses; whether it is a statistical-based model, probabilistic-based model, neural network based model or a tree-based model (based on decision trees). Some papers tested different types of models and will be added in a separate subsection.

1) USING STATISTICAL BASED MODELS

There are many statistical techniques and models that have been used in Bitcoin price prediction and analysis. Multiple linear regression was used in [11], [12], [17] to model the relationship between the Bitcoin cryptocurrency price and some predictor variables. These variables included some economic and financial variables. Additionally, measurements of public interest in Bitcoin from Twitter feeds, Google Trends and Wikipedia views related to Bitcoin [11] were recorded. Twitter and Google Trends data were also used by [12] to predict prices of Bitcoin and Ethereum. They found that the tweet volume, rather than the tweet sentiment is a good predictor of price.

Ordinary least squares (OLS) criterion was used to model short-run impact of independent variables on the price of Bitcoins while long-run relationships between the co-integrated time series were captured using a vector error-correction model (VECM) in [11]. OLS was used in [15] to estimate the parameters of autoregressive distributed lag models (ARDL) depending on features describing how political incidents and statements mentioned in the news affect the price, the oil price, gold price and volatility variables, Google search volume, positive and negative shocks.

Logistic regression (LR) was compared to autoregressive integrated moving average (ARIMA) for Bitcoin Price prediction in [18] and its other variants in [19]. OLS regression and fractionally integrated ARMA (ARFIMA) were used in [19] to investigate the stochastic properties of the top six largest cryptocurrencies at that time (Bitcoin, Ethereum, Ripple, Litecoin, Stellar and Tether) and their relationship to six stock market indices.

2) USING NEURAL NETWORK BASED MODELS

Young *et al.* [13] used a deep learning model to predict Bitcoin price and extent of transactions fluctuation of the currency. Deep learning was also used in [28] to predict Bitcoin and Ethereum prices in Australian dollars (AUD). A comparative study of different *deep learning models* (deep neural network (DNN), long-short term memory (LSTM) and artificial neural network (ANN)) was done in [20] for Bitcoin price prediction. LSTM-based prediction models slightly outperformed other techniques for regression of Bitcoin price while DNN-based models performed better for classification of price changes whether up or down. They also showed that classification models were more effective than regression models for trading profitability.

In [21], Yiying *et al.* analyzed the price dynamics of Bitcoin, Ethereum and Ripple using ANN and LSTM. They surprisingly found that ANN relies more on long-term history while LSTM tends to rely more on short-term dynamics efficiently utilizing useful information hidden in them.

Many other researchers have tested neural networks and its variants, some of them are reviewed and compared in [22]. The authors of [22] proposed layer-wise randomness into the observed features' activations of multilayer perceptron (MLP) and LSTM to simulate market volatility. They achieved no more than 5% average improvement in MAPE in their best case experiment with 23 features with a window size of 7 days.

3) USING TREE-BASED MODELS

Alessandretti *et al.* [23] tested three forecasting models using daily cryptocurrency prices of 1,681 currencies. The first model is a decision tree model to predict the return on investment for all cryptocurrencies. The second model is a gradient boosting decision tree model for each cryptocurrency to make a prediction on each cryptocurrency depending on information about the behaviour of the whole market. In the third method, they used a different LSTM model for each currency, where the prediction is based on previous prices of the currency. For parameter optimization, they used two evaluation metrics; the geometric mean return and the Sharpe ratio. The three methods performed better than the baseline simple moving average strategy when applying the investment strategy for the whole period considered. The optimization of parameters based on the Sharpe ratio achieved higher returns. Methods based on gradient boosting decision trees performed best for short-term 5/10 days predictions. On the other hand, LSTM performed best for predictions based on $\approx$ 50 days of data, since they are able to also capture long-term dependencies and were very stable against price volatility.

4) USING PROBABILISTIC BASED MODELS

Shah and Zhang [24] used Bayesian regression for predicting Bitcoin prices. Kim *et al.* [14] used averaged one-dependence estimators (AODE) as a prediction model which is a probabilistic classification learning technique. They used the model to predict fluctuations in prices and number of transactions at different lags.

TABLE 2. Comparison between AI research work done for price prediction and analysis in cryptocurrency.

Study	Year	AI techniques	Datasets
[24]	2014	Bayesian regression	Price and order book from Okcoin.com (Yuan) between February 2014 to July 2014
[11]	2015	SVMs for sentiment analysis Linear regression for short-term dependency Vector error-correction model (VECM) for long-run dependency	Bitstamp data from 27/10/2014 to 12/1/2015 2,125,243 tweets about Bitcoin Google trends for daily searches bitcoinpulse.com for Wikipedia searches
[14]	2016	Averaged one-dependence estimators (AODE)	Bitcoin Forum from 12/2013 to 01/02/2016 Ethereum Forum from 7/8/2015 to 01/02/2016 Ripple Forum from 07/09/2015 to 21/01/2016 Bitcoin prices from Coindesk Ethereum prices via CoinMarketCap and Etherscan Ripple prices from RippleCharts
[18]	2016	Random Forest, LR, LDA	OKCoin Yuan exchange prices for Bitcoin for some months in 2014, 2015 and 2016
[13]	2017	Deep-learning model	Bitcoin Forum from 12/2013 to 21/09/2016 Bitcoin prices and transactions from Coindesk
[15]	2018	Autoregressive Distributed Lag (ARDL)	279 weekly observations from 99bitcoins.com, historyofbitcoin.org for political shock variables Quandl and Datasream for Bitcoin and economic indices Google Trends for Google search volume
[23]	2018	XGboost with two varying inputs LSTM	300 exchange markets prices between 11/11/2015 and 24/04/2018 (daily prices in USD for 1,681 cryptocurrencies)
[26]	2018	Bayesian neural network (BNN)	bitcoincharts.com from September 2011 to August 2017. Blockchain.info information Macroeconomic indexes Exchange rates between major fiat currencies
[17]	2019	LASSO, ElasticNet, k-NN, SVM, MLP, Stochastic Gradient Descent, Bayesian Regression, Decision Tree, Extra Tree Random Forest, AdaBoost	CoinMarketCap.com data of 100 cryptocurrencies ranging from 2015 to early 2018 prices of traditional assets like gold, USD and other stocks indices from Bloomberg
[19]	2019	ARFIMA OLS regression	cryptocompare.com of top six cryptocurrencies from 07/05/2015 to 05/10/2018 DataStream for six major financial indices
[16]	2019	Neural networks (NN) Support Vector Machines (SVM) Random Forest (RF)	Historical 80-days market data obtained from the top performing 65 cryptocurrency exchanges using cryptocompare.com public API
[27]	2019	Random Sampling Method MLP,LSTM	Proprietary commercial data provided by Kaiko for OkCoin BTC and LTC prices to CNY and USD time series data at a minute frequency from 13/06/2013 to 18/3/2017
[20]	2019	DNN, LSTM, CNN, deep residual network (ResNet), convolutional RNNs(CRNN), Ensemble, SVM and Base, Random as baselines	Bitstamp Bitcoin market (USD) prices 29/11/2011 to 31/12/2018 & Bitcoin blockchain
[21]	2019	ANN LSTM	historical prices data from Blockchain.com[5] from 7/8/2015 to 2/6/2018. features:opening, high, low,closing prices with different predictive lengths
[25]	2019	HMM-LSTM compared to ARIMA and LSTM	Bitcoin prices, orders and trades data from Coinbase[6] between 20/8/2018 and 20/9/2018
[22]	2020	MLP LSTM	bitinfocharts prices from mid of 2017 to the end of 2019. (23 features with window size of 7) Blockchain data features Tweets and Google Trends Bitcoin, Ethereum, Litcoin

Hidden Markov model (HMM) has been used in [25] to address the dynamics of Bitcoin prices and get hidden information that cannot be directly extracted. A total of 52 features were then fed to an LSTM to predict the Bitcoin prices. Genetic algorithm was used in optimizing the parameters of LSTM. Effectiveness of the resultant model HMM-LSTM was compared to ARIMA and conventional LSTM and was found to outperform them in terms of MSE, RMSE and MAE.

[5] https://www.blockchain.com/markets
[6] https://www.blockchain.com/markets

Bayesian neural network (BNN) was used in [26] for regularizing weights of input variables to a neural network to account for the high volatility of Bitcoin price. The input variables included blockchain variables, macroeconomics variables and international fiat currencies exchange rates. BNN was found to outperform support vector regression (SVR) in terms of RMSE and MAPE.

5) USING DIFFERENT MODEL TYPES

In [16], neural networks (NN), MLP, support vector machines (SVM) and random forest (RF) were used for the prediction task to predict the price of the Bitcoin, Ethereum, Ripple and Litecoin cryptocurrency market movements. MLP was found to outperform the other models in case of Bitcoin, Ethereum and Ripple, while SVM performed the best in case of Litecoin. Shintate and Pichl [27], proposed a random sampling method (RSM) for Bitcoin time series trend prediction and they used a walk forward optimization method. They used this method to account for non-stationarity of the high-frequency time-series trading data. They compared the proposed approach to LSTM and MLP. The profit rates based on RSM were higher than those based on LSTM and MLP, and not easily biased by class imbalance. Differently, designing automatic trading strategies has been tackled in [30] using reinforcement learning. Double and dueling double deep Q-learning networks were used with testing two reward functions: Sharpe ratio and profit reward function. The system based on Sharpe ratio achieved the highest profits for Bitcoin trading over a period of almost four years of simulation.

In [31], many statistical machine learning techniques like support vector regressions (SVR) and Gaussian Poisson regressions (GRP), algorithmic models such as regression trees (RT) and the k-nearest neighbours (k-NN) and finally artificial neural network topologies such as feedforward (FFNN), Bayesian regularization (BRNN) and radial basis function networks (RBFNN) to find BRNN achieving the best accuracy. They used Bitcoin intra-day price data sampled at 5 minutes intervals in the period from January 1st, 2016 to March 16th, 2018.

In [17], the authors compared different techniques and least absolute shrinkage and selection operator (LASSO) was found to dominate in predicting the 30-day returns of cryptocurrencies. In [20], LSTM slightly outperformed the other models under test when the prediction task was modeled as a regression problem, whereas DNN slightly outperformed the other models when perceived as a classification problem. There is no single model that can be said to be the best for the problem of cryptocurrency price prediction as the model depends on many factors such as the size of the dataset, the different indicators and features at different lags to be used for prediction. Furthermore, the uncertainty associated with the crypto-prices and the random black swan events that can happen at anytime, are hard to be predicted by any model with good accuracy. In Figures 4 and 5, box plots were plotted for the range of values reported for accuracy and MAPE for some of the machine learning models used in [13], [14], [16]–[18], [20], [22], [26], [27]. The most common evaluation metric for the models treating price prediction as a classification problem was accuracy as in Figure 4 and MAPE was used as the common evaluation metric for regression models as in Figure 5. Figure 4 shows that LR, linear discriminant analysis (LDA) and RF achieve the best results for classification models while MLP and stochastic MLP were the best for regression models as shown in Figure 5. The best reported accuracy is 72% and the best MAPE is 2.8% but each was based on different features and time periods. Price prediction models are away from perfect prediction due to the speculation involved in trading and the complexities involved in the market that affect the trading decisions but as it is well known all models are wrong, but some are useful.

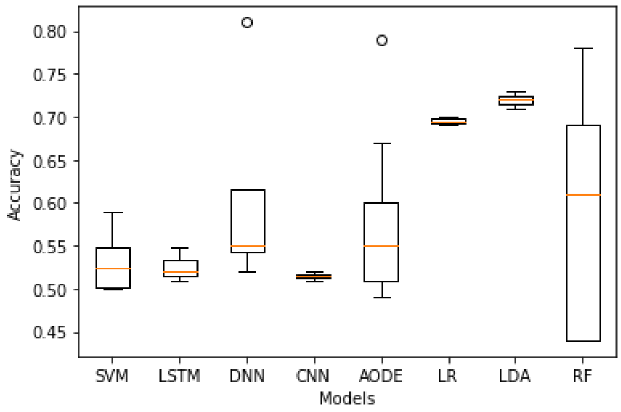

FIGURE 4. Box plot of accuracy for the machine learning techniques used in price prediction as a classification problem for papers cited in this section with accuracy as the evaluation metric (the higher the accuracy, the better prediction will be). On each box, the central mark is the median and the edges of the box are the 25th and 75th percentiles. The small circles represent outliers.

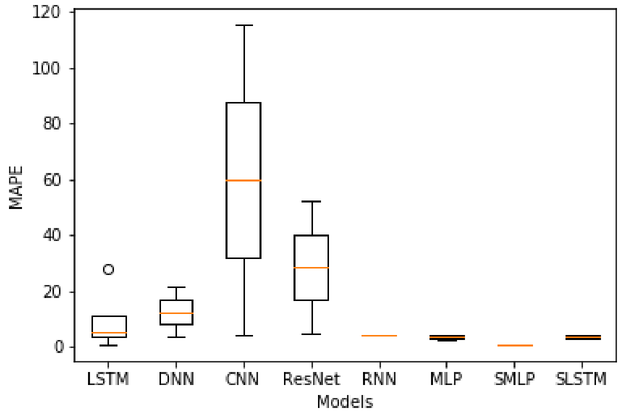

FIGURE 5. Box plot of MAPE for the machine learning techniques used in price prediction as a regression problem for papers cited in this section with MAPE used as the evaluation metric (the less MAPE, the better the prediction will be). On each box, the central mark is the median and the edges of the box are the 25th and 75th percentiles. The small circles represent outliers.

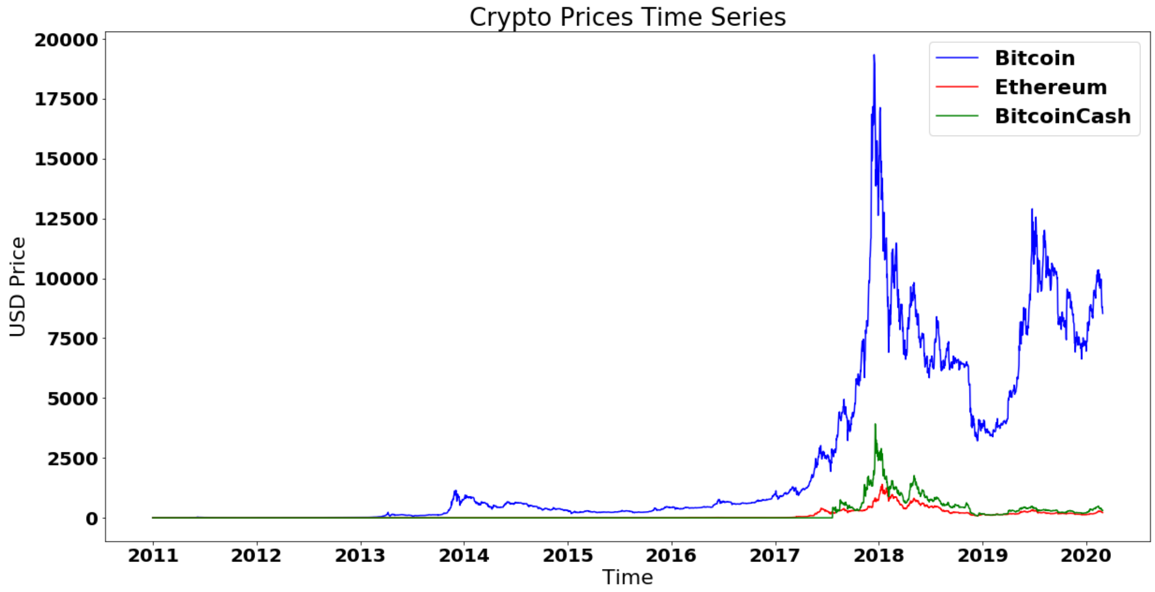

FIGURE 6. The top cryptocurrencies' prices until the start of March, 2020; prices are taken from Coindesk.[8]

Other studies [11], [12], [15], [19], [21], [23]–[25], [28] used different evaluation metrics so they were not included in the figures.

We believe that the best model highly depends on the selected features and the period of study. It can be a good approach to first test the time series for the prices and the different indicators for non-linearity using statistical tests and analytic techniques used in [32] to analyze the effect of different factors on the daily prices of cryptocurrencies. The work in [32] did not use any machine learning model to learn from these factors. However, the results presented can be analyzed and used to determine the best factors to use for building a cryptocurrency price prediction model. In [33], the authors tested for the importance of the economic and technology factors affecting the Bitcoin price over different periods of time using ANN and RF after dividing the Bitcoin price time series into four periods. It might also be useful to learn models based on segmented time periods. Some other nonlinear tests and techniques listed in [34] have not been explored in the field of cryptocurrencies price prediction.

B. VOLATILITY PREDICTION

Volatility is defined as the degree of variation of a trading price series over time. It represents the amount of uncertainty or risk about the size of changes in currency value. Bitcoin and other cryptocurrencies are considered to be volatile. Figure 6 shows a graph for cryptocurrencies' prices until the start of March 2020. Volatility of cryptocurrencies is mainly caused by their decentralized nature making their prices uncontrollable by any organization or government. Accordingly, cryptocurrencies can be considered as being traded in a free market where the price is solely determined by the supply and demand, however, there are other opinions that oversee the presence of a true long-run dependency [11], [13], [17] which invalidates the efficient market hypothesis. There are other factors affecting the price value, other than the interaction between supply and demand as mentioned in the last section. People investing in Bitcoin consider high volatility to be an indication of high-risk investment.

Volatility accounts for price movement away from its average value. A cryptocurrency price range could be estimated if volatility can be predicted or estimated for a day or a week based on historical data. Table 3 provides a summarized comparison of research papers using AI techniques to model cryptocurrency volatility and the datasets used.

Generalized autoregressive conditional heteroskedasticity (GARCH), which is a time-series statistical model, is used for modeling volatility [35]–[37]. Twelve different GARCH models were tested in [35] for modeling Bitcoin, Dash, Dogecoin, Litecoin, Maidsafecoin, Monero and Ripple. Integrated GARCH (IGARCH) and Glosten-Jagannathan-Runkle GARCH (GJRGARCH) models were found to be the best fit for most cryptocurrencies. Maximum likelihood was used to fit the different GARCH-type models. Peng *et al.* [36] combined the traditional GARCH model with Support Vector Regression (SVR), which can cover multi-variate and dynamic characteristics of financial series robustly. It was used to predict the volatility of three cryptocurrencies (Bitcoin, Ethereum and Dash) and three fiat currencies (Euro, British Pound and Japanese Yen) to evaluate the alternative risky investments and guide the investment decisions. They provided strong evidence that the SVR models significantly outperforms the traditional GARCH models.

[8] https://www.coindesk.com

TABLE 3. Comparison between AI research work done for prediction of cryptocurrency volatility.

Study	Year	AI techniques	Datasets
[35]	2017	Different GARCH models	BNC2database from Quandl from 22/6/2014 to 17/5/2017 for Bitcoin, Dash, LiteCoin, MaidSafeCoin, Monero, DogeCoin and Ripple.
[36]	2018	GARCH model combined with SVR	Altcoin Charts[9] for Bitcoin, Ethereum and Dash prices (in USD) Forex Historical Data[10] for Euro, British Pound and Japanese Yen (in USD) from 4/1/2016 to 31/7/2017
[37]	2019	12 GARCH models	Investing.com[11] daily closing prices of cryptocurrencies Bitcoin, Ethereum, Litecoin, Ripple, Moreno, Dash, Stellar and NEM from 7/8/2015 until 1/8/2018
[40]	2019	Markov-switching GARCH models (MSGARCH)	Coindesk Price Index for Bitcoin, from 18/7/2010 to 30/4/2018 CoinMarketCap for Ethereum, Ripple and Litecoin from 7/8/2015, 4/8/2013 and 28/4/2013 respectively till 30/4/2018
[38]	2019	Temporal mixture model against 12 baselines	Hourly volatility from OKCoin Order book data from OKCoin from September 2015 to April 2017
[39]	2019	ANN(MLP,GRU,LSTM) SVM, Ridge regression HAARV	BTC blockchain from 1/1/2012 to 11/11/2018

Using and comparing GARCH models was also investigated in [37] to model volatility in Bitcoin, Ethereum, Litecoin, Ripple, Moreno, Dash, Stellar and NEM. Then, they estimated the one-day-ahead *value-at-risk* (VaR) forecasts. The asymmetric GARCH models with long memory property achieved overall better performance for all cryptocurrencies.

Guo *et al.* [38] tested temporal mixture model using both incremental learning and rolling procedure for performance prediction to predict volatility of Bitcoin prices. They compared their model to different statistical (GARCH, Beta-GARCH, structural time series model, ARIMA) and machine learning baseline models (Random Forest, Gradient Boosting, Elastic-net regression, Gaussian process based regression and LSTM). Their temporal mixture model proved to be more accurate in most of the cases while being robust and adaptive with respect to time-varying data.

Other linear and nonlinear machine learning models were used for predicting realized volatility in Bitcoin prices based only on past values of realized volatility at different lags in [39]. Ridge regression was found to perform the best in terms of MSE and RMSE among the tested techniques (ANN in its different forms (MLP, GRU and LSTM), SVM, ridge regression, and heterogeneous autoregressive realized volatility (HAARV)).

C. AUTOMATED TRADING

Many cryptocurrencies' trading bots are currently available that implement trading strategies and offer customized customer's strategy. Trading bots are software products or websites that offer what is called ''algorithmic trading'' as they automatically analyze market actions and indicators, offer strategies to maximize trader's gains and improve her satisfaction. They can aggregate historical market data, calculate indicators, simulate order execution and even can be set up to execute strategies while the customer is asleep. Some bots use natural language processing techniques to communicate with the customer in a more natural and friendly way [9]. In the design of these trading bots, many algorithms and techniques similar to those used for price and volatility prediction mentioned in the last two subsections are used to maximize the profit and develop a strategy with maximum return. They differ in the number of exchanges they support and the features they offer. Additionally, they can offer portfolio construction and optimization to find an optimal weighing of financial assets which might include Bitcoin, other cryptocurrencies and other traditional financial assets like stocks and bonds. This optimization aims at maximizing the overall return while minimizing the variance of the return. The results in [41] suggest that investors should include Bitcoin in their portfolio as it generates substantial higher risk-adjusted returns through comparing eight different well-known portfolio optimization techniques widely used for traditional assets. They used the statistical-based GARCH model to learn the dynamic conditional correlations between Bitcoin prices and other bonds and indices.

Hierarchical risk parity approach was applied in [42] to a large portfolio of 61 cryptocurrencies which involves an unsupervised tree clustering, quasi-diagonalization and recursive bisection. They carried out-of-sample comparison with traditional risk-minimization methods for financial assets (Inverse Volatility (IV), Minimum Variance (MV), and Maximum Diversification (MD)). A thorough

[9] http://alt19.com
[10] http://fxhistoricaldata.com
[11] http://www.investing.com/

review of financial portfolio optimization techniques can be found in [43]. We are here only concerned with applying AI and ML techniques in portfolio construction and optimization for automated trading of cryptocurrencies. Alessandretti *et al.* [23] used the gradient boosting decision trees models and and LSTM prediction model to build investment portfolios for cryptocurrencies based on the predictions optimizing the geometric mean return and the Sharpe ratio. Nakano *et al.* [44] used a deep neural network for return prediction in Bitcoin intra-day trading based on technical variables and time-series data every 15 minutes. Irene *et al.* [45] relied on variants of GARCH models to determine the optimal portfolio weights for a minimum variance equity portfolio. Differently, Cocco *et al.* [46] viewed the trading optimization problem from an evolutionary perspective and proposed using genetic algorithms (GA). They simulated a Chartist trading agent that uses the best trading rules evolved through GA against random traders.

The area of applying artificial intelligence and machine learning in portfolio construction and optimization for cryptocurrencies is considered a recent research area. It still needs further investigation to find the best strategy that suits cryptocurrencies while getting benefit from machine learning models and techniques used for other financial assets [47].

Investors classification can help investors and automated trading bots get more insights about the cryptocurrencies market and its price dynamics in order to develop profitable trading strategies. In [48], the socio-demographic characteristics of cryptocurrency investors and the factors that affect their investment decisions (whether they invested, never invested or intend to invest in the future) for any cryptocurrency are investigated. With about 402 survey responses for Australian and Chinese blockchain and cryptocurrencies' followers, they built a multinomial Logit model to find the factors that affect the choice of investment in cryptocurrency coins versus other types of initial coin offering (ICO) tokens. They found age, gender, education, occupation, and investment experience significantly affect the decision. Having insights and clear analysis for the reasons behind people's investing decisions in an ICO is critical to decide on trading strategies. It can help in marketing ICOs of cryptocurrencies, but bigger and more diverse datasets are required to have a good outcome.

In a recent study [49], the authors used unsupervised clustering technique to group different types of investors. They based their clustering on similarities in trading behavior according to trade volume, average bid volume, average relative price and average duration to finish a trade from the time the offer is placed on a well-known exchange website. They were able to classify investors into 10 clusters (six types of investors offering Bitcoins and four types of investors ordering Bitcoins). They used ARDL model to identify the factors (macro-financial, technical trading indicators, technological measures and market sentiment from Twitter) that might influence the trading behavior of investor types (speculators, cryptocurrency miners, informed traders, large professional investors, USD-orientated investors, global traders, etc.). They showed that the exchange rate of Bitcoin is seen to be significantly driven by the number of orders placed by only one cluster of investors. The discussion in their work sheds more light on the cryptocurrency market investors and the herd behavior shown by one cluster which results in speculative price changes. We suggest that performing a similar behavior clustering analysis directly on blockchain data, although it will face many challenges, may give a more accurate clustering. It might give more evidence about how different indicators affect the investors' behaviors and accordingly affect the price.

D. FRAUD DETECTION

The use and reputation of Bitcoin and other cryptocurrencies in aiding illicit activities is a big concern, as it affects the stability and the trust in cryptocurrencies. Cryptocurrencies are known to attract cybercriminals for their pseudo-anonymity and for being operated outside the laws of governments and banks. However, regulators are continuously trying to enforce know-your-customer (KYC) and anti-money laundering (AML) laws for exchanges and escrow services. There are different types of scams and criminal activities that can occur in cryptocurrencies, such as digital theft, hacking, phishing, Ponzi-schemes, pump-and-dump schemes, purchasing illegal drugs and money laundering in the black market. A Ponzi-scheme is one of the fraudulent schemes that offer high rates of return for early investors as it generates returns for early investors by acquiring new customers.

Fraud detection is based on detecting anomalies and suspicious behaviour in the transactions and trades history [50], especially that Bitcoin transactions are transparently recorded on the blockchain public ledger.

With the scarcity of labeled incidents or examples for different fraud activities, Monamo *et al.* [50] used trimmed k-means and k-means clustering based on features from the transactions graph in a semi-supervised way to detect fraudulent activity in the Bitcoin transactions network. Both algorithms achieved optimal clustering. The trimmed algorithm was able to detect 5 from the 30 well-known anomalies such as the Mt Gox, Linode Hack, and 50 BTC Theft as examples for thefts and hacks. Using k-d trees, they were able to detect 2 more thefts. Based on the clustering labels for outliers, the authors employed some supervised classification models to understand the relation between the labels and predictor variables. Random forest achieved the best precision.

For estimating the proportion of possible cybercriminal entities in the Bitcoin ecosystem, Yin *et al.* [51] tested 13 different machine learning classifiers for this task. They used LR, LDA, k-NN, classification and regression tree (CART), Naive Bayes (NB), SVM, random forest, extremely randomized forests, bagging and gradient boosting. The last four achieved the best results as reported in Table 4. The dataset used by the authors and provided by Chainalysis[12] is limited and facing the same problem of substantial undersampling of

[12] https://www.chainalysis.com/

TABLE 4. **Comparison between AI research work done for fraud detection in cryptocurrency.**

Study	Year	AI techniques	Datasets
[50]	2016	K-means , Trimmed K-means clustering General linear model (GLM) General additive model (GAM) Random forest (RF)	Bitcoin blockchain data before 7/4/2013
[51]	2017	Random forest (RF) Extremely randomized forests Bagging (0.78) Gradient boosting	A dataset provided by Chainalysis: (854 labeled observations for 12 different clusters, 100000 observations for uncategorized cluster)
[52]	2018	RIPPER Bayes Net Random Forest	BitcoinTalk: 32 Ponzi-schemes addresses 24 features for 6432 bitcoin addresses 2 classes (32 Ponzi and 6400 non-Ponzi)
[53]	2019	J48, Random Forest Stochastic Gradient Descent	Opensource dataset for 184 Ponzi addresses[13] 3,203 addresses verified by Etherscan
[54]	2019	RF, XGBoost Neural network, SVM, k-NN	Bitcoin blockchain from 9/1/2009 to 9/2/2017 services labels from WalletExplorer and BlockchainInfo tags for nonHYIP addresses 2,134 HYIP owners' addresses from bitcointalk.org

some of the 12 categories like stolen-bitcoins, ransom-ware or mixing.

To detect *Ponzi schemes* automatically, Bartoletti *et al.* [52] constructed a dataset of real-world Ponzi schemes by analyzing the Bitcoin blockchain transactions used for scams through a manual search on Reddit and bitcointalk.org, they were able to find a list of 32 Ponzi schemes for which advertisers in the forums present them as high-yield investment programs (HYIP), or as gambling games. Then, they collected the deposit addresses for these schemes through manual search in the forum and on their websites. They extended the number of addresses through address clustering and multi-input heuristic. To address the well-known class-imbalance problem in this task, they used different under-sampled sets and different cost-matrix weights. Formalizing the problem as a binary classification problem for an address to be either Ponzi-related or non-Ponzi related, they evaluated three different types of classifiers: repeated incremental pruning to produce error reduction (RIPPER) which is a rule-based technique, Bayes network and random forest. The best classifier was random forest which correctly classified 31 Ponzi schemes out of 32. For Ponzi-schemes detection in Ethereum, Jung *et al.* [53] used J48 (decision tree algorithm), random forest and stochastic gradient descent (SGD) for classification. *HYIP* classification has been addressed in [54] using a dataset with a total of 2,134 HYIP addresses. Toyoda *et al.* [54] used random forest (RF), gradient boosting implementation (XGBoost), ANN, SVM and k-NN for binary classification of HYIP and non-HYIP addresses. Random forest achieved the best results based on 7 features characterizing the transactions.

A summarized comparison for the aforementioned reviewed research work is presented in Table 4. The common thing that can be obviously seen very challenging in this kind of problem is the lack of a reliable labeled dataset of reasonable size available for researchers. Even studies that used the Chainalysis dataset still face the problem of dealing with the high class-imbalance problem. Except for [50], which combined semi-supervised and supervised techniques, other efforts [51]–[54] solely relied on supervised techniques. The high costs of manual labeling as done in [52] and the small size of the labeled datasets for fraudulent addresses suggest we need to explore using semi-supervised learning techniques for fraud detection in a way similar to [55].

E. ANONYMITY AND PRIVACY

Privacy and anonymity are two necessary aspects for online financial trading. Anonymity is mostly favored by criminals to hide their identities when dealing illegally for drugs or weapons or being involved in money laundering transactions. However, it is also preferred by privacy-savvy people who want to keep their identities and transactions anonymous and private. Privacy means protecting the data of transacting users including the traded amount, the transacting parties, their balances and the timing of the transaction.

Trying to reveal the identities of Bitcoin users and linking their Bitcoin addresses and trades usually rely on using public data information from social media or other publicly available data in a process called ''deanonymization''. It is either based on heuristics to link this data to the blockchain transactions as in [56]–[59], or it can be based on AI techniques. Deanonymization has been approached using AI in two ways; clustering [60] or classification [61], [62] and [63]. Ermilov *et al.* [60] used off-chain information (e.g. twitter and walletexplorer.com) together with the blockchain

[13] https://goo.gl/CvdxBp

information based on common spending and one time change heuristics for address clustering. They used a probabilistic framework to assign an address to a cluster based on maximizing the log-likelihood function with six clusters of entities: mining pools, exchanges, darknet markets, mixers, gambling and other services. The authors in [61] and [62] considered a particular pattern of 2-motif and 3-motif features as a *potential* laundering pattern used in the classification model based on the pattern of acquisition and spending of Bitcoin in the dataset. They derived an entity-transaction graph from the address-transaction graph. Additionally, features like betweenness, closeness, in-degree, out-degree, PageRank and load centrality over a week, a month and a year were used. With other temporal features; they proposed a decision tree method with gradient boosting and compared the results against a logistic regression algorithm. This classification problem faces the entity-category class imbalance between *Service* category and the *Mining Pool* category for example.

Gradient boosting classifier (GBC) achieved the best results in [63] to classify a Bitcoin address into one of ten categories. They used synthetic minority over-sampling technique (SMOTE) to oversample the two minority classes *hosted-wallet* and *mixing* to overcome the class imbalance problem. Modeling the problem as a classification problem as well, Lin *et al.* [64] proposed a different set of features including moments of transaction time in higher order to train different supervised machine learning models (LR, SVM, RF, AdaBoost, LightGBM, ANN, etc.) on a labeled category data set. Researchers in [65], used cascading of machine learning models to enrich entities' information with data from previous classifications. This enhanced the overall classification performance using 34 features for initial classification. They utilized three different models; *Adaboost, random forest and gradient boosting*.

Different from all the above research work reviewed, Juhász *et al.* [66] did not rely on labeled data for Bitcoin addresses classification. They monitored Bitcoin network messages through 140 nodes distributed over the network to bind Bitcoin transactions to geographical locations. The monitoring agents watched out for clients relaying transactions during the first time segment; estimated to be 2 sec. The client originating the transaction is possibly one of those agents. A comparison between different studies, the datasets used and the techniques tested are listed in Table 5. It can be seen that most of the studies relied on supervised techniques, except for [60], which used a semi-supervised method combining clustering with labels from different public sources.

F. CRYPTOCURRENCY MINING

The mining process has the disadvantage of high electricity consumption used by mining pools for participating in the PoW computations. Only one miner succeeds to add a block of transactions, while other mining pools are left with the expenses of huge energy costs. This disadvantage threatens the decentralization of the cryptocurrency and makes it susceptible to monopolization, especially when the block reward will vanish over time due to Bitcoin block reward halving. Some researchers [68] theoretically proposed using deep-learning tasks as a PoW to let the electricity be consumed in useful tasks. A coin is rewarded when a miner exceeds a minimum threshold for performance. They also proposed a proof-of-storage mechanism to store the deep learning models on distributed nodes called keepers which are also to be rewarded as per the proposed model for keeping secure storage for the models. While the idea of saving the electricity for useful operations seems beneficial, yet the proposed model ignored some important aspects. Among them is the required security and protection for the training and validation data. The challenge is how to protect the data that must be shared with different untrusted miners in order to solve the problem, as proposed in the model. A second issue is the time for adding a block since some deep-learning tasks may take days for training a model. The Bitcoin block adding time is already criticized by people in the domain as it is greatly limiting the transactions' rate, but it is needed for properly synchronizing the blockchain. Also deep-learning problems and their training time depend on the training set size. The authors did not mention how their proposed model could handle these issues.

Confirmation time for transactions depends on various factors in the mining process and the network. Singh *et al.* [69] addressed the confirmation time prediction for Ethereum blockchain. For this task, they used the dataset for Ethereum transactions until November 2018 to get a set of features including the transactions count done by the sender, the sender's gas, gas price, gas used, the timestamp for sending the transaction (calculated from the pending transaction pool) and the transaction's block timestamp. They formulated the problem as a classification problem with eight categories depending on the duration of confirmation of the transaction within (15 s, 30 s, 1 min, 2 min, 5 min, 10 min, 15 min, 30 min and longer). They used SMOTE to address the class imbalance problem as most of the transactions are confirmed within 15s. They compared three models; *Naïve Bayes, random forest and MLP* according to accuracy, null-accuracy and Cohen's Kappa score to account for the imbalanced nature of the data. They found that MLP achieved the best accurate results.

Additionally, Feng *et al.* [70] proposed a *Markov model* to model and analyze a selfish mining strategy in Ethereum. They viewed the PoW mining process as a series of Bernoulli trails that independently search for a nonce in order to generate a new block. They found that the computational power threshold which makes selfish mining profitable in Ethereum is lower than that in Bitcoin mining. This makes Ethereum more vulnerable to 51% attack. They suggested designing new reward functions to allow miners maximizing their profits through a more secure and honest strategy.

Game-theoretic analysis has also been used in [71] to prove the importance of block reward in mining to keep the blockchain secure since the transaction fee model could encourage selfish mining. Another study [72] used game

TABLE 5. Comparison between AI research studies done in cryptocurrency deanonymization.

Study	Year	AI techniques	Datasets
[60]	2017	Maximum log-likelihood estimation (Clustering technique)	Bitcoin blockchain data (3/1/2009 to 9/3/2017 tags from 97 sources (twitter.com, walletexplorer..) >20M clean tags for six categories
[61]	2017	Random Forest , AdaBoost Linear SVM , Perceptron) Logistic Regression	Bitcoin blockchain data from January 2009 to April 2015
[62]	2018	Logistic regression Gradient Boosting Machine	Bitcoin blockchain data before 24th March 2018 addresses' labels from WalletExplorer 315 extracted features from entity-transaction graphs
[66]	2018	Naive Bayes classifier (0.952)	Bitcoin blockchain data "INV" messages to 140 Bitcoin network monitoring agents between 10/14/2013 and 12/20/2013
[63]	2019	k-nearest neighbors (k-NN)(0.429) CART decision trees (0.692) AdaBoost(0.613), Gradient Boosting (0.804) Random Forest (0.783), Extra Trees (0.75),Bagging (0.783)	A dataset from Chainalysis 957 entities, 385M transactions 98 features were extracted, 12 classes
[64]	2019	Logistic Regression, SVM, AdaBoost,Random Forest, XGBoost, Neural Network LightGBM (best F1 scores)	Bitcoin blockchain data from 3/1/2009 to 30/06/2018 services labels from WalletExplorer and BlockchainInfo HYIP operators' labels from BitcoinTalk 7 classes
[67]	2019	imECOC, CART, MC-HDDT, AdaBoost	Transaction network with 44,503,503 edges based on November 2018 transactions WalletExplorer tags for transactions' addresses
[65]	2019	Random Forest, Adaboost, Gradient Boosting	Bitcoin blockchain data until 5/2/2019 311 different entities from WalletExplorer 6 classes (Exchange, Gambling, Marketplace, Mining Pool, Mixer, Service)

theory for analysis of some malicious mining strategies like block withholding (BWH).

It can be concluded that ideas using AI techniques in the mining process to replace solving the crypto-puzzle in the PoW, or entirely replace the PoW by an AI-based consensus mechanism, have not been thoroughly investigated and evaluated. AI techniques can be employed by mining pools to choose which cryptocurrency to mine and which mining pool to join in order to reduce the electricity consumption and increase their profit based on historical data. A study done in [73] used prospect theory in comparison with utility theory to predict the profitability of a miner. They learned from data obtained from mining with 5 pools (AntPool, F2Pool, BTC.com, SlushPool and BatPool) for 40 consecutive days using AntMiner S5. The data included parameters related to the mining pools (e.g. hash distribution and reward sharing method) and other parameters specific to the miner (e.g. its hash rate and electricity cost) in addition to the current value of the currency and the value of the block reward. They concluded that their prospect theoretic approach predicted their profits more accurately than the expected utility approach.

G. SECURITY

Despite the security and privacy properties that exist in blockchain-based cryptocurrencies which were surveyed in [74], there are several security threats that are facing the cryptocurrency ecosystem [75]. They can be classified as attacks on the distributed network, mining process attacks, double spending and transaction malleability attacks. There are also client-side security attacks and privacy threats to wallet, exchange or escrow services [58]. In this paper, we only focus on research papers related to security that rely on AI techniques.

Johnson *et al.* [76] used game-theoretical models of competition between two mining pools of varying sizes to find the trade-off between mining strategies. Triggering a distributed denial-of-service (DDoS) attack to lower the chance that a competing mining pool wins in solving the crypto-puzzle and adds a block to the chain. They considered differences in costs of both investment and attack, as well as uncertainty of the success of a DDoS attack. By analyzing the game's equilibria, they found that pools have a greater incentive to attack large pools than small ones. It was also concluded that larger mining pools have a greater incentive to attack than smaller ones.

In a different context, DDoS attack detection in Bitcoin-related services (e.g. mining pools, currency exchanges, e-Wallet, gambling services) has been studied in [77]. Johnson *et al.* used MLP trained with features based on data collected in [78] and blockchain blocks and transactions to detect DDoS attacks. Blocks were labeled as either ''DDoS'' or ''nonDDoS'' to reflect whether the block was created on the day the DDoS attack occurred. The inadequacy of the tested model with its low accuracy and some limitations are reported in the paper.

A cryptocurrency network is more vulnerable to a 51% attack [75] if selfish miners controlled more than a certain threshold of computational hash power. Selfish mining behavior has been modeled by a Markov-model and analyzed in [70] for Ethereum as discussed in the last section. Selfish mining poses a serious threat to cryptocurrencies adopting PoW. Another type of attack called block with-holding (BWH) has been analyzed in [79] based on the work done in [72]. In a BWH attack, a pool of miners sends a share of its mining power to another pool and does not announce any PoW solutions it finds thus ''withholding'' them while still collecting the shared reward generated by other non-attacking miners. Consequently, the BWH attack lowers the effectiveness of the victim pool while increasing the effectiveness of the attacker's pool. Eyal *et al.* [72] considered BWH as a rational behavior that is profitable for the attacker. Profit analysis for a realistic and complex BWH attack scenario was done in [79] using game theory to derive a mathematical representation for BWH attacks. The analysis revealed that pools attacking each other can reach a nash equillibrium and that largest pools benefit from BWH while smaller pools lose if there is an attack so they have to attack to maximize their profits. The smallest pools, however, are better not to attack in any scenario. In a similar fashion, Li *et al.* [80] analyzed the nash equillibrium of the mining process through numerical simulations exploring the influence of the mining pools' power, the ratio of the power to be infiltrated, and the betrayed rate of dispatched miners.

Caporale *et al.* [81] used a Markov-switching non-linear specification to analyze the effects of cyber attacks on returns in the case of four cryptocurrencies (Bitcoin, Ethereum, Litecoin and Stellar) over the period August 2015 to February 2019. In addition to the daily prices data and the VIX data obtained from the Federal Reserve of St.Louis., they used the data source for cyber attacks from Hackmageddon.[14] They included not only the crypto attacks targeting cryptocurrencies, but also other cyber attacks as they see that the media coverage for these attacks could also affect the investors' perception of cryptocurrencies which rely on cyber security. The results suggest significant negative effects of cyber attacks on the probability of cryptocurrencies staying in the low volatility regime. This work can be seen as statistical learning from data which can further be used for price prediction after the occurrence of cyber attacks.

The aforementioned research for security analysis either used supervised machine learning models or game-theory techniques. Some other recent ML solutions and proposals to address the problem of detecting suspicious activities in Bitcoin and blockchain were surveyed in [82]. Some of the referenced work are better categorized as fraud detection or anonymity and deanonymization and we have already covered them in the previous subsections.

IV. DISCUSSION AND POSSIBLE FUTURE RESEARCH DIRECTIONS

Technology advances have impacted the cryptocurrencies evolution by creating new ways to mine new coins, store the blockchains over distributed nodes, secure the network and analyze the huge amount of trades and blockchain transactions that are beyond human capabilities. In this study, we presented a survey for the state-of-art research that makes use of artificial/machine intelligence techniques to address the challenges facing cryptocurrencies.

The AI research studies addressing Bitcoin are remarkably more than those researching other altcoins as seen in Table 6. The possible dependencies between cryptocurrencies' prices should be further identified. The possibility of using AI techniques to address security, anonymity and privacy level of other cryptocurrencies is recommended for further exploration as security and privacy are major and critical concerns for traders to gain more trust while trading.

A. NOTES ON PRICE PREDICTION

Although this topic is the most studied among other topics using AI to tackle cryptocurrency issues, there is still room for more research on applying AI techniques for cryptocurrency price predictions in new contexts. Most of the papers predicted Bitcoin prices in USD but few papers cover price prediction of other altcoins with other fiat currencies. Conducted research relied mostly on the cryptocurrency price history indicators such as open, high, low and closing prices, while few studies take advantage of different sources of social media, online metrics and other stock markets indicators. Among the used social media sources are tweets and their sentiments, Reddit posts, Wikipedia views, and Google Trends data. Few research efforts considered using BitcoinTalk forum posts. Other news sources, such as newspapers and news agencies have not been taken into account, although such sources can shape the opinion of some novice investors and influence users' buying or selling behaviour. Technical news about security breaches or instability of the

[14] https://www.hackmageddon.com

TABLE 6. AI research work for Bitcoin and other altcoins.

Challenge	Bitcoin	Altcoins
Price prediction	[11], [13], [15], [18], [20], [21], [24]–[26]	[12], [16], [17], [19], [22], [23], [28]
Volatility prediction	[38], [39]	[35]–[37], [40]
Automated trading	[9], [42], [44]–[46], [48], [49]	[23]
Fraud detection	[50]–[52], [54]	[53]
Mining	[68], [71]	[69], [70]
Security	[72], [76], [77], [82]	[70], [81]
Anonymity and privacy	[54], [60]–[63], [65]–[67]	None

cryptomarket might affect the price as well. Moreover, almost all the studies depend merely on posts and content in the English language for their analysis while investors from all over the world contribute to the cryptocurrency market.

The results of research efforts in price prediction depended on different datasets with features at different time periods. Therefore, the results can not be fairly compared to reach a conclusion for recommending or favoring one price prediction model over another. Tree-based models and probabilistic-based models were the least models to be tested. Probabilistic-based models are better to be investigated in more depth to model the uncertainty in the domain of cryptocurrencies. Introduction of some sort of trust or confidence score measure for the prediction accuracy or performance of the model to account for the uncertainty and missing factors or explanatory variables is recommended. Nonlinear tests and techniques listed in [34] have not yet been thoroughly explored for application on cryptocurrencies time series and need further investigation to capture the non-linear dependencies on the explanatory variables.

B. NOTES ON VOLATILITY PREDICTION

Volatility prediction has been mostly approached using GARCH model variants. GARCH models are preferred by financial experts because they provide a more real-world context to predict the prices and financial returns. Most of the volatility prediction research was conducted by finance researchers using GARCH variants and was based on historical prices. Only [39] explored other techniques but with only the past values of the realized volatility. No volatility prediction research took advantage of social media and news metrics that can give an estimate of the user adoption and interest in the cryptocurrency market. Investors classification [49] and estimation of the size of short-term traders cluster that shows herd behavior can be seen to directly affect the volatility of Bitcoin price and need to be further investigated. Additionally, [36] was able to combine a regression technique with GARCH and the results were promising. There is a need for additional research efforts to combine other regression techniques and to use different indicators.

C. NOTES ON AUTOMATED TRADING

There are many online commercial trading bots currently available that apply price and volatility prediction techniques and portfolio trading strategies for portfolio management that were reviewed in section III-B. Most of the portfolio trading strategies research is based on the portfolio construction strategies used for traditional financial assets. We believe there is still a room in this area to learn the best strategy for trading of cryptocurrencies that can adapt dynamically to new ICOs and take into account the behavior of different clusters of investors. The portfolio construction may be also approached using game theory approaches with the design of a proper payoff.

D. NOTES ON FRAUD DETECTION

Fraud detection research suffered from the unavailability of sufficient reliable datasets. The class imbalance problem is a challenge facing fraud detection since only a few addresses are marked as fraudulent. Different oversampling and undersampling approaches to address the class imbalance problem need to be investigated in the domain of cryptocurrencies. Semi-supervised techniques need to be investigated as they can be a way for learning from a small labeled dataset while having many unlabeled examples. AI solutions for other types of a scam like fake ICOs and pump-and-dump schemes have not been examined as well.

E. NOTES ON DEANONYMIZATION

Clustering techniques need to be further investigated for their ability to cluster addresses based on the patterns of exchanges between addresses. This can help to regulate authorities to spot addresses suspected for money-laundering or other illegal behavior. Additionally, this can help in investor classification and in gaining more insights about the cryptocurrency market dynamics to develop better trading strategies. Furthermore, most of the deanonymization research was done for Bitcoin [59]. Deanonymization of other cryptocurrencies remains to be explored as well.

F. NOTES ON MINING

The sharp drop in the Bitcoin price after the 2017 hype, together with the increased difficulty of mining a new block and the uncertainty about what might happen when reaching the maximum limit of Bitcoins, create an environment where mining becomes less profitable and unpredictable than in the past. This creates a serious risk that impacts the trust in cryptocurrencies. This urges miners to employ *data analysis for* *mining* in order to implement new mining strategies that predict profit based merely on transactions fees as income, and ignore block reward as a major source of income. The mining strategy was best modeled using game theoretic approaches. However, miners' data of total computation power and electricity consumption is not available. The mining research work reviewed for modifying the consensus protocol or the PoW puzzle to be solved is still immature and requires further investigation and more concrete evidence of effectiveness.

G. NOTES ON SECURITY

It is thought that the data of cryptocurrencies' network events can hold significant information about miners, their devices and their network joining and leaving patterns. AI techniques can help detect attacks or suspicious activities on the network based on this data. This direction has not been explored yet. One of the reasons for that is the unavailability of a dataset for attacks and suspicious activities to use for learning. Building big labeled datasets for security attacks is an expensive task. This suggests semi-supervised learning as a good candidate to solve this kind of problem as long as the assumptions for semi-supervised techniques are met.

V. CONCLUSION

In summary, this survey navigates through and organizes the vast amount of diverse research work that applies AI techniques in the field of cryptocurrencies. The state-of-art research efforts were classified into six classes. For each class, a comparison of different research work according to the used techniques and datasets was provided. We highlighted possible research gaps and open directions that require future development in this highly dynamic field. Although we have not cited all the research papers in the field, yet we did our best to cite recent papers that investigated a wide spectrum of different AI techniques to tackle different challenges. This survey can greatly help researchers interested in the application of AI and machine learning techniques in the field of cryptocurrencies. It gives them a quick, yet full, overview of this multidisciplinary area; through presenting simplified reviews of some of the research done in this area and the used techniques while listing some of the available datasets they used to address the different cryptocurrencies challenges.

ACKNOWLEDGMENT

The research in this article was done for project number NPRP X-063-1-014 under a grant from Qatar National Research Fund (QNRF). However, the contents of the research are solely the responsibility of the authors and do not necessarily represent the official views of QNRF. They would like also to acknowledge Qatar National Library (QNL) for its Open Access Fund to support the article processing charges.

REFERENCES

[1] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach* (Series in Artificial Intelligence), 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2010.

[2] H. Hassani, X. Huang, and E. Silva, "Big-crypto: Big data, blockchain and cryptocurrency," *Big Data Cognit. Comput., Open Access J.*, vol. 2, pp. 10–34, Dec. 2018.

[3] K. Salah, M. H. U. Rehman, N. Nizamuddin, and A. Al-Fuqaha, "Blockchain for AI: Review and open research challenges," *IEEE Access*, vol. 7, pp. 10127–10149, 2019.

[4] K. Sgantzos and I. Grigg, "Artificial intelligence implementations on the Blockchain. Use cases and future applications," *Future Internet*, vol. 11, no. 8, p. 170, Aug. 2019.

[5] V. Lopes and L. A. Alexandre, "An overview of blockchain integration with robotics and artificial intelligence," *Ledger*, vol. 4, Apr. 2019, doi: 10.5195/ledger.2019.171.

[6] T. Klein, H. Thu, and T. Walther, "Bitcoin is not the new gold—A comparison of volatility, correlation, and portfolio performance," *Int. Rev. Financial Anal.*, vol. 59, pp. 105–116, Oct. 2018.

[7] N. Smuts, "What drives cryptocurrency prices?: An investigation of Google trends and telegram sentiment," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 46, no. 3, pp. 131–134, Jan. 2019.

[8] A. Burnie and E. Yilmaz, "An analysis of the change in discussions on social media with bitcoin price," in *Proc. 42nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, New York, NY, USA, Jul. 2019, pp. 889–892.

[9] Q. Xie, D. Tan, T. Zhu, Q. Zhang, S. Xiao, J. Wang, B. Li, L. Sun, and P. Yi, "Chatbot application on cryptocurrency," in *Proc. IEEE Conf. Comput. Intell. Financial Eng. Econ. (CIFEr)*, May 2019, pp. 1–8.

[10] J.-Z. Huang, W. Huang, and J. Ni, "Predicting bitcoin returns using high-dimensional technical indicators," *J. Finance Data Sci.*, vol. 5, no. 3, pp. 140–155, Sep. 2019.

[11] I. Georgoula, D. Pournarakis, C. Bilanakos, N. D. Sotiropoulos, and M. G. Giaglis, "Using time-series and sentiment analysis to detect the determinants of bitcoin prices," in *Proc. Mediterranean Conf. Inf. Syst. (MCIS)*, 2015, p. 20. [Online]. Available: http://aisel.aisnet.org/mcis2015/20

[12] J. Abraham, D. Higdon, J. Nelson, and J. Ibarra, "Cryptocurrency price prediction using tweet vol. and, sentiment analysis," *SMU Data Sci. Rev.*, vol. 1, no. 3, pp. 37–73, 2018.

[13] Y. B. Kim, J. Lee, N. Park, J. Choo, J.-H. Kim, and C. H. Kim, "When bitcoin encounters information in an online forum: Using text mining to analyse user opinions and predict value fluctuation," *PLoS ONE*, vol. 12, no. 5, pp. 1–14, May 2017.

[14] Y. B. Kim, J. G. Kim, W. Kim, J. H. Im, T. H. Kim, S. J. Kang, and C. H. Kim, "Predicting fluctuations in cryptocurrency transactions based on user comments and replies," *PLoS ONE*, vol. 11, no. 8, pp. 1–17, Aug. 2016.

[15] F. Kjærland, M. Meland, A. Oust, and V. Øyen, "How can bitcoin price fluctuations be explained," *Int. J. Econ. Financial Issues*, vol. 8, no. 3, pp. 323–332, 2018.

[16] F. Valencia, A. Gómez-Espinosa, and B. Valdés-Aguirre, "Price movement prediction of cryptocurrencies using sentiment analysis and machine learning," *Entropy*, vol. 21, no. 6, p. 589, Jun. 2019.

[17] J. Liew, R. Li, T. Budavári, and A. Sharma, "Cryptocurrency investing examined," *J. Brit. Blockchain Assoc.*, vol. 2, no. 2, pp. 1–12, Nov. 2019.

[18] J. Muhammad Amjad and D. Shah, "Trading bitcoin and online time series prediction," in *Proc. NIPS Time Ser. Workshop*, 2016, pp. 1–15.

[19] L. A. Gil-Alana, E. J. A. Abakah, and M. F. R. Rojo, "Cryptocurrencies and stock market indices. Are they related?" *Res. Int. Bus. Finance*, vol. 51, Jan. 2020, Art. no. 101063.

[20] S. Ji, J. Kim, and H. Im, "A comparative study of bitcoin price prediction using deep learning," *Mathematics*, vol. 7, no. 10, p. 898, Sep. 2019.

[21] W. Yiying and Z. Yeze, "Cryptocurrency price analysis with artificial intelligence," in *Proc. 5th Int. Conf. Inf. Manage. (ICIM)*, Mar. 2019, pp. 97–101.

[22] P. Jay, V. Kalariya, P. Parmar, S. Tanwar, N. Kumar, and 9, "Stochastic neural networks for cryptocurrency price prediction," *IEEE Access*, vol. 8, pp. 82804–82818, 2020.

[23] L. Alessandretti, A. ElBahrawy, L. M. Aiello, and A. Baronchelli, "Anticipating cryptocurrency prices using machine learning," *Complexity*, vol. 2018, pp. 1–16, Nov. 2018.

[24] D. Shah and K. Zhang, "Bayesian regression and bitcoin," in *Proc. 52nd Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Sep. 2014, pp. 409–414.

[25] I. A. Hashish, F. Forni, G. Andreotti, T. Facchinetti, and S. Darjani, "A hybrid model for bitcoin prices prediction using hidden Markov models and optimized LSTM networks," in *Proc. 24th IEEE Int. Conf. Emerg. Technol. Factory Autom. (ETFA)*, Sep. 2019, pp. 721–728.

[26] H. Jang and J. Lee, "An empirical study on modeling and prediction of bitcoin prices with Bayesian neural networks based on blockchain information," *IEEE Access*, vol. 6, pp. 5427–5437, 2018.

[27] T. Shintate and L. Pichl, "Trend prediction classification for high frequency bitcoin time series with deep learning," *J. Risk Financial Manage.*, vol. 12, no. 1, p. 17, Jan. 2019.

[28] G. Senthuran and M. Halgamuge, *Prediction of Cryptocurrency Market Price Using Deep Learning and Blockchain Information: Bitcoin and Ethereum*. New York, NY, USA: Taylor & Francis, Sep. 2019, pp. 349–364.

[29] E. Zivot and J. Wang, "Nonlinear time series models," in *Modeling Financial Time Series With S-PLUS*. New York, NY, USA: Springer, 2006, pp. 653–712.

[30] G. Lucarelli and M. Borrotti, "A deep reinforcement learning approach for automated cryptocurrency trading," in *Artificial Intelligence Applications and Innovations*, J. MacIntyre, I. Maglogiannis, L. Iliadis, and E. Pimenidis, Eds. Cham, Switzerland: Springer, 2019, pp. 247–258.

[31] S. Lahmiri and S. Bekiros, "Intelligent forecasting with machine learning trading systems in chaotic intraday bitcoin market," *Chaos, Solitons Fractals*, vol. 133, Apr. 2020, Art. no. 109641.

[32] C. Ross Phillips and D. Gorse, "Cryptocurrency price drivers: Wavelet coherence analysis revisited," *PLoS ONE*, vol. 13, no. 4, pp. 1–21, Apr. 2018.

[33] W. Chen, H. Xu, L. Jia, and Y. Gao, "Machine learning model for bitcoin exchange rate prediction using economic and technology determinants," *Int. J. Forecasting*, to be published. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0169207020300431

[34] R. S. Tay and R. Chen, *Nonlinear Time Series Analysis*. Hoboken, NJ, USA: Wiley, Aug. 2018.

[35] J. Chu, S. Chan, S. Nadarajah, and J. Osterrieder, "GARCH modelling of cryptocurrencies," *J. Risk Financial Manage.*, vol. 10, no. 4, p. 17, Oct. 2017.

[36] Y. Peng, P. H. M. Albuquerque, J. M. C. de Sá, A. J. A. Padula, and M. R. Montenegro, "The best of two worlds: Forecasting high frequency volatility for cryptocurrencies and traditional currencies with support vector regression," *Expert Syst. Appl.*, vol. 97, pp. 177–192, May 2018.

[37] A. Ngunyi, S. Mundia, and C. Omari, "Modelling volatility dynamics of cryptocurrencies using GARCH models," *J. Math. Finance*, vol. 9, no. 4, pp. 591–615, 2019.

[38] T. Guo, A. Bifet, and N. Antulov-Fantulin, "Bitcoin volatility forecasting with a glimpse into buy and sell orders," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2018, pp. 989–994.

[39] R. Miura, L. Pichl, and T. Kaizoji, "Artificial neural networks for realized volatility prediction in cryptocurrency time series," in *Advances in Neural Networks—ISNN*, H. Lu, H. Tang, Z. Wang, Eds. Cham, Switzerland: Springer, 2019, pp. 165–172.

[40] G. M. Caporale and T. Zekokh, "Modelling volatility of cryptocurrencies using Markov-switching GARCH models," *Res. Int. Bus. Finance*, vol. 48, pp. 143–155, Apr. 2019.

[41] E. Platanakis and A. Urquhart, "Should investors include bitcoin in their portfolios? A portfolio theory approach," *Brit. Accounting Rev.*, vol. 52, no. 4, Jul. 2020, Art. no. 100837.

[42] T. Burggraf and A. Vyas, "Beyond risk parity—A machine learning-based hierarchical risk parity approach on cryptocurrencies," in *Proc. Finance Res. Lett.*, 2020, Art. no. 101523.

[43] P. Xidonas, R. Steuer, and C. Hassapis, "Robust portfolio optimization: A categorized bibliographic review," *Ann. Oper. Res.*, vol. 292, no. 1, pp. 533–552, 2020.

[44] M. Nakano, A. Takahashi, and S. Takahashi, "Bitcoin technical trading with artificial neural network," *Phys. A, Stat. Mech. Appl.*, vol. 510, pp. 587–609, Nov. 2018.

[45] I. Henriques and P. Sadorsky, "Can bitcoin replace gold in an investment portfolio?" *J. Risk Financial Manage.*, vol. 11, no. 3, p. 48, Aug. 2018.

[46] L. Cocco, R. Tonelli, and M. Marchesi, "An agent-based artificial market model for studying the bitcoin trading," *IEEE Access*, vol. 7, pp. 42908–42920, 2019.

[47] D. Andriosopoulos, M. Doumpos, P. M. Pardalos, and C. Zopounidis, "Computational approaches and data analytics in financial services: A literature review," *J. Oper. Res. Soc.*, vol. 70, no. 10, pp. 1581–1599, Oct. 2019.

[48] D. Xi, T. I. O'Brien, and E. Irannezhad, "Investigating the investment behaviors in cryptocurrency," 2019, *arXiv:1912.03311v1*. [Online]. Available: https://arxiv.org/abs/1912.03311

[49] A. Keller and M. Scholz, "Trading on cryptocurrency markets: Analyzing the behavior of bitcoin investors," in *Proc. 14th Int. Conf. Inf. Syst. (ICIS)*, 2019, p. 11. [Online]. Available: https://aisel.aisnet.org/icis2019/blockchain_fintech/blockchain_fintech/11/

[50] P. M. Monamo, V. Marivate, and B. Twala, "A multifaceted approach to bitcoin fraud detection: Global and local outliers," in *Proc. 15th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2016, pp. 188–194.

[51] H. Sun Yin and R. Vatrapu, "A first estimation of the proportion of cybercriminal entities in the bitcoin ecosystem using supervised machine learning," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2017, pp. 3690–3699.

[52] M. Bartoletti, B. Pes, and S. Serusi, "Data mining for detecting bitcoin Ponzi schemes," in *Proc. Crypto Valley Conf. Blockchain Technol. (CVCBT)*, Jun. 2018, pp. 75–84.

[53] E. Jung, M. Le Tilly, A. Gehani, and Y. Ge, "Data mining-based ethereum fraud detection," in *Proc. IEEE Int. Conf. Blockchain (Blockchain)*, Jul. 2019, pp. 266–273.

[54] K. Toyoda, P. Takis Mathiopoulos, and T. Ohtsuki, "A novel methodology for HYIP operators' bitcoin addresses identification," *IEEE Access*, vol. 7, pp. 74835–74848, 2019.

[55] G. E. Melo-Acosta, F. Duitama-Munoz, and J. D. Arias-Londono, "Fraud detection in big data using supervised and semi-supervised learning techniques," in *Proc. IEEE Colombian Conf. Commun. Comput. (COLCOM)*, Aug. 2017, pp. 1–6.

[56] S. Meiklejohn, M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G. M. Voelker, and S. Savage, "A fistful of bitcoins: Characterizing payments among men with no names," in *Proc. Conf. Internet Meas. Conf. (IMC)*, New York, NY, USA, 2013, pp. 127–140.

[57] A. Biryukov, D. Khovratovich, and I. Pustogarov, "Deanonymisation of clients in bitcoin P2P network," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur. (CCS)*, New York, NY, USA, 2014, pp. 15–29.

[58] F. Sabry, W. Labda, A. Erbad, H. Al Jawaheri, and Q. Malluhi, "Anonymity and privacy in bitcoin escrow trades," in *Proc. 18th ACM Workshop Privacy Electron. Soc. (WPES)*, London, U.K., Nov. 2019.

[59] H. A. Jawaheri, M. A. Sabah, Y. Boshmaf, and A. Erbad, "Deanonymizing TOR hidden service users through bitcoin transactions analysis," *Comput. Secur.*, vol. 89, Feb. 2020, Art. no. 101684.

[60] D. Ermilov, M. Panov, and Y. Yanovich, "Automatic bitcoin address clustering," in *Proc. 16th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2017, pp. 461–466.

[61] S. Ranshous, C. Joslyn, S. Kreyling, K. Nowak, F. N. Samatova, L. C. West, and S. Winters, "Exchange Pattern Mining in the Bitcoin Transaction Directed Hypergraph," in *Proc. Financial Cryptogr. Workshops*, 2017, pp. 248–263.

[62] M. Jourdan, S. Blandin, L. Wynter, and P. Deshpande, "Characterizing entities in the bitcoin blockchain," in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, Nov. 2018, pp. 55–62.

[63] H. H. Sun Yin, K. Langenheldt, M. Harlev, R. R. Mukkamala, and R. Vatrapu, "Regulating cryptocurrencies: A supervised machine learning approach to de-anonymizing the bitcoin blockchain," *J. Manage. Inf. Syst.*, vol. 36, no. 1, pp. 37–73, Jan. 2019.

[64] Y.-J. Lin, P.-W. Wu, C.-H. Hsu, I.-P. Tu, and S.-W. Liao, "An evaluation of bitcoin address classification based on transaction history summarization," in *Proc. IEEE Int. Conf. Blockchain Cryptocurrency (ICBC)*, May 2019, pp. 302–310.

[65] F. Zola, M. Eguimendia, J. L. Bruse, and R. Orduna Urrutia, "Cascading machine learning to attack bitcoin anonymity," in *Proc. IEEE Int. Conf. Blockchain (Blockchain)*, Jul. 2019, pp. 154–167.

[66] L. Péter Juhász, J. Stéger, D. Kondor, and G. Vattay, "A Bayesian approach to identify bitcoin users," *PLoS ONE*, vol. 13, no. 12, pp. 1–21, Dec. 2018.

[67] J. Liang, L. Li, W. Chen, and D. Zeng, "Targeted addresses identification for bitcoin with network representation learning," in *Proc. IEEE Int. Conf. Intell. Secur. Informat. (ISI)*, Jul. 2019, pp. 158–160.

[68] A. Baldominos and Y. Saez, "Coin.AI: A proof-of-useful-work scheme for blockchain-based distributed deep learning," *Entropy*, vol. 21, no. 8, p. 723, Jul. 2019.

[69] H. J. Singh and A. S. Hafid, "Prediction of transaction confirmation time in ethereum blockchain using machine learning," in *Blockchain and Applications*, J. Prieto, A. K. Das, S. Ferretti, A. Pinto, and J. M. Corchado, Eds. Cham, Switzerland: Springer, 2020, pp. 126–133.

[70] C. Feng and J. Niu, "Selfish mining in ethereum," in *Proc. IEEE 39th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Jul. 2019, pp. 1306–1316.

[71] M. Carlsten, H. Kalodner, S. M. Weinberg, and A. Narayanan, "On the instability of bitcoin without the block reward," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur. (CCS)*, New York, NY, USA, Oct. 2016, pp. 154–167.

[72] I. Eyal, "The Miner's dilemma," in *Proc. IEEE Symp. Secur. Privacy*, Washington, DC, USA, May 2015, pp. 89–103.

[73] M. Salimitari, M. Chatterjee, M. Yuksel, and E. Pasiliao, "Profit maximization for bitcoin pool mining: A prospect theoretic approach," in *Proc. IEEE 3rd Int. Conf. Collaboration Internet Comput. (CIC)*, Oct. 2017, pp. 267–274.

[74] R. Zhang, R. Xue, and L. Liu, "Security and privacy on blockchain," *ACM Comput. Surv.*, vol. 52, no. 3, pp. 51:1–51:34, Jul. 2019.

[75] M. Conti, E. Kumar, C. Lal, and S. Ruj, "A survey on security and privacy issues of bitcoin," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 4, pp. 3416–3452, 4th Quart., 2017.

[76] B. Johnson, A. Laszka, J. Grossklags, M. Vasek, and T. Moore, "Game-theoretic analysis of DDoS attacks against bitcoin mining pools," in *Financial Cryptography and Data Security*, R. Böhme, M. Brenner, T. Moore, and M. Smith, Eds. Berlin, Germany: Springer, 2014, pp. 72–86.

[77] U.-J. Baek, S.-H. Ji, J. T. Park, M.-S. Lee, J.-S. Park, and M.-S. Kim, "DDoS attack detection on bitcoin ecosystem using deep-learning," in *Proc. 20th Asia–Pacific Netw. Oper. Manage. Symp. (APNOMS)*, Sep. 2019, pp. 1–4.

[78] M. Vasek, M. Thornton, and T. Moore, "Replication data for: Empirical analysis of denial-of-service attacks in the bitcoin ecosystem," *Harvard Dataverse, V2*, 2014, doi: 10.7910/DVN/25541.

[79] S. Elliott, "Nash equilibrium of multiple, non-uniform bitcoin block withholding attackers," in *Proc. 2nd Int. Conf. Data Intell. Secur. (ICDIS)*, Jun. 2019, pp. 144–151.

[80] W. Li, M. Cao, Y. Wang, C. Tang, and F. Lin, "Mining pool game model and Nash equilibrium analysis for PoW-based blockchain networks," *IEEE Access*, vol. 8, pp. 101049–101060, 2020.

[81] G. M. Caporale, W.-Y. Kang, F. Spagnolo, and N. Spagnolo, "Non-linearities, cyber attacks and cryptocurrencies," *Finance Res. Lett.*, vol. 32, Jan. 2020, Art. no. 101297.

[82] M. Rahouti, K. Xiong, and N. Ghani, "Bitcoin concepts, threats, and machine-learning security solutions," *IEEE Access*, vol. 6, pp. 67189–67205, 2018.

FARIDA SABRY (Member, IEEE) received the M.Sc. and Ph.D. degrees in computer engineering from the Faculty of Engineering, Cairo University, Egypt. She is currently a Lecturer with the Computer Engineering Department, Faculty of Engineering, Cairo University. She is also working as a part-time Postdoctoral Researcher with the Department of Computer Science and Engineering, Qatar University. Her main research interest includes machine intelligence and its applications.

WADHA LABDA (Associate Member, IEEE) received the M.Sc. degree in information systems and the Ph.D. degree from The University of Manchester, U.K., in 2018 and 2012, respectively. She is currently working as an Assistant Professor with the Department of Computer Science and Engineering, College of Engineering, Qatar University, where she is also the Head of the Technology Innovation and Engineering Education (TIEE) Unit, College of Engineering. Her research interests include privacy of data, knowledge engineering, and blockchain and its applications.

AIMAN ERBAD (Senior Member, IEEE) received the M.C.S. degree in embedded systems and robotics from the University of Essex, U.K., and the Ph.D. degree in computer science from The University of British Columbia, Canada. He is currently an Associate Professor with the College of Science and Engineering, Hamad Bin Khalifa University (HBKU). His research interests include cloud computing, edge computing, the IoT, private and secure networks, and multimedia systems. He received the Platinum award from H. H. Emir Sheikh Tamim bin Hamad Al Thani at the Education Excellence Day 2013 (Ph.D. category). He also received the 2020 Best Research Paper Award from *Computer Communications*, the IWCMC 2019 Best Paper Award, and the IEEE CCWC 2017 Best Paper Award. He is an Editor of *KSII Transactions on Internet and Information Systems* and was a Guest Editor of *IEEE Network*.

QUTAIBAH MALLUHI (Member, IEEE) received the B.S. degree from the King Fahd University of Petroleum & Minerals, and the M.S. and Ph.D. degrees from the University of Louisiana at Lafayette. He was the Head of the Department, from 2006 to 2012, and the Director of the KINDI Center for Computing Research, Qatar University (QU), from 2012 to 2016. He was the Co-Founder and CTO of Data Reliability Inc. He served as a Faculty for Jackson State University. He is currently a Professor with the Department of Computer Science and Engineering, Qatar University (QU).