Journal of Engineering Science and Technology Review 17 (3) (2024) 125 - 143

Review Article

JOURNAL OF
**Engineering Science
and Technology Review**

www.jestr.org

Blockchain and Artificial Intelligence for Big Data Analytics in Networking: Leading-edge Frameworks

Patikiri Arachchige Don Shehan Nilmantha Wijesekara*

Department of Electrical and Information Engineering, Faculty of Engineering, University of Ruhuna, Galle 80000, Sri Lanka

Received 31 March 2023; Accepted 7 June 2024

Abstract

Big Data (BD) Analytics (BDA) in networking involves acquisition, sharing, pre-processing, storage, analysis, interpretation, and decision-making. Blockchain (BC) technology incorporates a progression of bonded blocks that fundamentally upholds credibility, protects unquestionability, and protects the partial-anonymity of its transactions on account of distributed consensus methods and cryptographic protocols. So as to fulfill the deficiency of a review paper catering to individual and combined use of BC and Artificial Intelligence (BCandAI) for BDA in the networking domain, in this work, we recognize 6 sections in the leading edge BCandAI BDA notion and rigorously analyze each stratagem concerning blockchain attributes, blockchain/AI techniques, network attributes, and the like. We piled up an opening sample of 89 publication citations by culling articles for screening requirements tracked down from cyber libraries, availing a comprehensive and protracted systemology. Established upon this exploration, we highlight that Artificial Intelligence (AI) can be involved in BDA by analyzing BD, while blockchain can facilitate secure transmission and storage of BD due to its inherent security features of unchangeability, non-deniability, etc., preventing data poisoning attacks, and facilitating hybrid on- and off-chain storage due to the challenges of high volume by availing techniques just like offloading and partial storage. Moreover, we highlight that there are BCandAI integrated approaches where blockchain-anchored secure BD storage is availed for secure federated learning or blockchain and cloud computing are availed for BD fusion for analysis, availing AI to generate accurate insights from BD. Rigorous analysis discloses that from all studies, 17.5% use BC alone, 20% avail of the combined BCandAI concept, 62.5% use AI alone, 70% address one or more BDA stages, 10% implement PoW consensus, 12.5% avail of deep learning, and 17.5% choose generic or IoT networks. Finally, we express the potentialities and problems of the proposition of BCandAI-anchored BDA concepts and then offer counsel to overpower them.

Keywords: Artificial Intelligence; Analytics; Blockchain; Big Data; Mixed BD Recording; Federated Learning

1. Introduction

Big Data (BD) is characterized by high volume (quantity), high variety (heterogeneity), high velocity (rate of generation), high veracity (variation in quality), and high value (significance) and is typically stored in a distributed manner [1]. They can be modeled either in graph form, consisting of a collection of nodes connected with edges with relationships typically containing millions of nodes, or in tensor form, having multiple dimensions that can store all classes of BD, such as structural, semi-structural, and un-structural data [2]. Amidst a Big Data Analytics (BDA) framework, it necessarily involves 6 steps: BD acquisition, sharing, pre-processing, storage, analysis, and interpretation with decision making [3]. Even though the BD analysis is a single step in the whole process of BD analytics, it is the most crucial step, while efficient functioning of acquisition, sharing, pre-processing, and storage are required for obtaining the correct analysis from the BD that can be aided in the decision-making process [4]. Specifically, BD analysis can be two-fold: descriptive analysis, where an overview of data is obtained using techniques like exploratory analysis, association rule mining, clustering, etc., and predictive analysis, where future outcomes or trends are predicted using historical data using techniques like classification, regression, inferential statistics, etc. [5]. Moreover, networking techniques like queuing theory, coflow scheduling, systematic modeling, etc. can be aided in the process of BD analytics to obtain optimum outputs and decisions with a low expenditure of network resources, leading to lower delays and optimal resource utilization [6].

A Blockchain (BC) unavoidably incorporates a progression of blocks bonded in a structured or non-linear mode grounded in the layout of the digital ledger infrastructure [7]. Clearly, transactions/blocks are interconnected, availing a defined block/transaction preserving the hash code of numerous precursor transactions/blocks and ensuring their unchangeability [8]. Likewise, they integrate a general consent workflow; for instance, proof-anchored general consent or vote-anchored general consent for authenticating the blocks among contemporaries in advance of a transaction/block is joined to the digital ledger infrastructure [9]. Blockchains leverage hash encryption to protect credibility and online signatures to protect transaction unquestionability [10]. Likewise, they are capable of embedding vigorous cryptographic protocols, for instance, zero-knowledge succinct non-interactive augment of knowledge and cryptography beyond quantum threats for protecting from quantum incursions and fortifying the qualities of privacy assurance in blockchain [11]. Yet, unaltered blockchain, as it is, steers clear of cryptographic protocols. For instance, the asymmetrical key cryptographic

*E-mail address: nilmantha@eie.ruh.ac.lk

doi:10.25103/jestr.173.16

method for protecting privacy assurance comes up short of privacy assurance due to the fact that blockchain interactions/transactions are partially anonymous, suggesting that interactions/transactions are determined by an encryption-anchored alias as an alternative to the valid addresses of members [12]. Specifically, the grade of privacy assurance may be altered in response to the digital ledger category: restricted, coalition, and unrestricted. Unrestricted blockchain is the typical totally dispersed blockchain, whereas restricted and coalition blockchains retain a grade of dominant governance, presenting enhanced solitude and data entitlement supervision than unrestricted [13].

In reliance on this evaluation, we identify that blockchain and Artificial Intelligence (AI) for BDA can be 6 folded. First, blockchain can be availed in the transmission and storage steps of BDA thanks to its characteristics of immutability, unquestionability, high trustworthiness, etc., preventing possible data poisoning attacks, along with robust authentication and access control like those availed in the framework SBBDA-IoD [14]. Moreover, research in [15] has advocated using a hybrid on-chain and off-chain approach to cater to the challenge of data redundancies and the high volume associated with BD, where blockchain is availed to prevent tampering with BD stored off-chain. Next, due to the additional computational burden of blockchain transactions due to the large volume of BD, researchers have advocated for efficiently storing BD in the blockchain, either by offloading blockchain transactions to the network edge with the aid of hollow blocks and express transactions or by storing only important pieces of BD [16]. Furthermore, AI can be extensively operated for the analysis step of BD that can produce predictive and descriptive analysis for diverse network functions such as intrusion detection, anomaly detection, fault detection, network monitoring, traffic prediction, etc. [17]. Fifthly, we find that the Blockchain and AI (BCandAI) fused approach in BDA is where blockchain-anchored storage of BD is utilized for secure federated learning to generate insights from the stored data [18]. Finally, cloud computing and blockchain have been utilized for secure and privacy-protecting data fusion for BD in an effort to reduce network congestion, along with Machine Learning (ML) for BD analysis [19].

The review paper [20] attempts to discuss the unification of AI, blockchain, and BD. However, work [20] does not emphasize BD analytics at all and rather discusses applications of AI, blockchain, and BD for various fields, for instance medicine, academia, agriculture, etc. Similarly, the review papers [21] and [22] investigate the applications of AI, blockchain, and BD applications in healthcare, but do not inspect the combined or individual use of AI and/or blockchain for BDA. Even though the review paper [23] reviews regarding the operation of blockchain for BD services and applications, it does not emphasize BD analytics, and further integrated use of BCandAI for BD analytics is not discoursed at all. To the limit of our cognition, during this writing, there is an absence of a survey paper reviewing the individual and the combined use of BCandAI, specifically emphasizing BDA. Therefore, this piece of work will fulfill an unattended space in the existing literature, providing a useful reference for future researchers on the integrated and individual use of BCandAI for BDA.

Figure 1 reveals the topic layout of this literary investigation.

2. Methodology

This exploration evaluates the active research work on BCandAI-anchored BDA out in published form over the decades, leveraging a comprehensive and protracted systemology [24]. Likewise, it dissects numerous viewpoints on big data, BDA, AI, and blockchain frameworks. Thus, all fresh academic investigations and cyberspace sheets published in electronic form on BD, BDA, blockchain, AI, and blockchain- and AI-anchored BDA consist of the full population set in the context of this examination. However, full population set references are unfathomable for scrutiny in this examination.

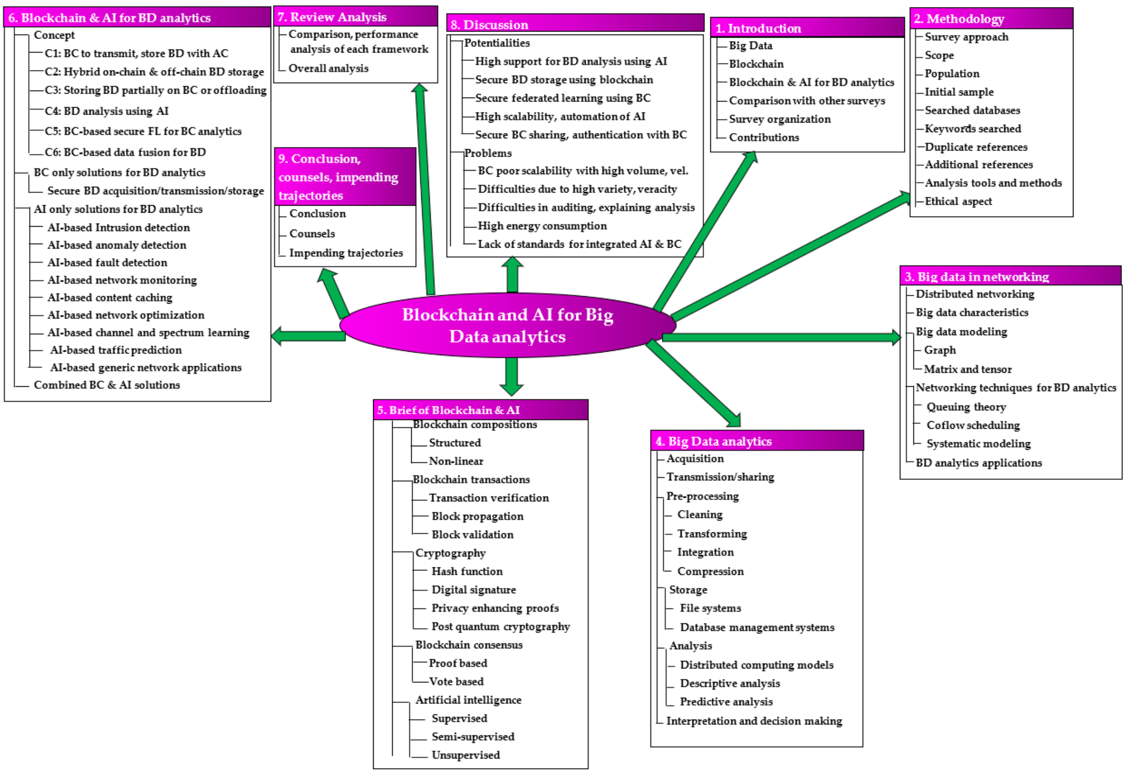


Fig. 1. Topic layout of literary investigation on BC and AI for BDA in networking.

Thus, leveraging applicable search expressions and screening requirements, we compiled 91 references from fresh academic investigations and cyberspace sheets.

We tracked down IEEE Xplore IT data resource, ScienceDirect scientific data collection vault, ACM cyber library, Google Scholar research material search engine, Wiley cyber library, and the MDPI document locator. Our preferred search expressions were "BD" OR "Blockchain" OR "BDA" OR "Blockchain-anchored BDA" OR "AI-anchored BDA" OR "Blockchain and AI-anchored BDA" OR "AI anchored BDA for intrusion diagnosis" OR "AI anchored BDA for anomaly diagnosis" OR "AI anchored BDA for fault diagnosis" OR "AI anchored BDA for network observation" OR "AI anchored BDA for content caching" OR "AI anchored BDA for network optimization" OR "AI anchored BDA for channel and spectrum learning" OR "AI anchored BDA for traffic prediction".

Many facets of culling the articles developed the screening requirements. The first two screening requirements are that publication necessitates English language and that it must be profoundly connected to the search expression. Thirdly, to facilitate rising the correctness of conducted exploration, periodical publications were set as primary concern by comparison with convention presentations and early version documents. On the other hand, we didn't show preference for academic papers in an explicit paper publisher within the screening requirements; in lieu, we thought of all scientific paper publishers alike. The last screening requirement proclaims that an explicit publication demands electronic distribution in the interval of years from 1970.

The opening sample was shrunken to 89 publication citations, next to identification of 2 publication citations was the same. Likewise, we stated clarifications and explanations regarding the spectrum of topics proposed in this exploration using 27 publications. To differentiate this exploration from former explorations, we inserted in a later phase 4 greater exploration articles into the group of literary work, conveying the aggregated figure of publication citations to 120.

To analyze in place blockchain- and AI-anchored BDA anchored on many facets, for instance, blockchain characteristics, AI characteristics, BD characteristics, network facets, and competence, we leveraged the structured table data for exploration's non-quantitative study. Likewise, we composed visual depictions leveraging the MS Excel application to dispassionately review exploration data attached to AI-anchored, blockchain-anchored, and BD-anchored facets.

Ethics lack relevance for the reason that this exploration links to data transmission networks.

3. BD in Networking

3.1 Distributed networking

BD are typically collected and stored in a distributed manner in networking in an effort to provide fast responses for local services deviating from the centralized mode of operation. In particular, Google's file organization structure is scattered, having thousands of clusters to store BD in a distributed manner [25]. Moreover, BD is processed and analyzed typically in a distributed manner using operations such as MapReduce, which functions parallelly in a distributed manner [26]. Thus, BD applications are most suitable to be availed in distributed networks.

3.2 BD characteristics

BD are characterized by Big-5V features that are briefly explained below [27].

3.2.1 Volume

Volume conveys a massive amount of data that is created, collected, and archived, usually in proportions of terabytes to petabytes. Due to the sheer volume, traditional approaches to data storage fail, and they require storage systems like Hadoop and NoSQL.

3.2.2 Variety

Variety specifies the high degree of heterogeneity that exists in BD as the data can come from various formats (video, audio, text, images, websites, etc.), structures (structured/semi-structured/unstructured), and sources having different data models.

3.2.3 Velocity

Velocity implies the high rate at which BD is born and subsequently operated due to the presence of real-time applications and streaming data sources such as stock markets, sensor data from networks, social media feeds, etc.

3.2.4 Veracity

Big veracity refers to the high degree of variation that exists in the quality of BD, as they come from different sources and can contain both high-quality and low-quality data. Thus, robust data cleansing and validation techniques must be implemented in BD systems.

3.2.5 Value

Big value refers to the high significance and knowledge that can be extracted using BD to perform meaningful tasks, optimize processes, etc. by making intelligent decisions to bring massive profits to individuals and organizations.

3.3 BD modeling

3.3.1 Graph

A graph is constituted by a group of points (nodes) associated with edges, where points denote real-world objects and edges represent relationships among the objects. Figure 2 reveals different realizations that BD can be modeled.

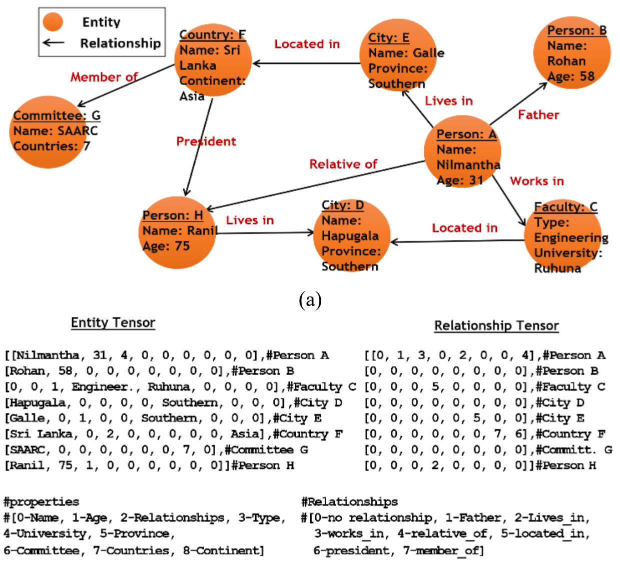


Fig. 2. Big data modeling. (a) Graph. (b) Tensor.

In BD scenarios, graphs can be composed of countless points and edges, such as in the case of social networks, biological networks, communication networks, etc. [28]. Techniques of graph spectra (obtaining eigen values and eigen vectors of the graph's connection matrix to study its spectral properties) and graph summarization (capturing essential characteristics of large graphs using techniques just like node sampling, subgraph extraction, graph compression, etc.) can be availed to make these large graphs processable [29].

3.3.2 Matrix and Tensor
Matrices can be availed to depict BD in a structured tabular form, storing structured data such as sensor readings, user item interactions, etc. to be used for subsequent BDA. Typically, sparse matrices having a large number of zeros and efficient storage are used for representing BD [30]. Compared with graph representation, matrix representation brings advantages in efficient computation, enables structured data analysis, and provides collaborative filtering. Moreover, the tensor representation with a large number of dimensions can effectively be availed to represent all forms of BD, such as structured (using a 2D matrix), semi-structured, and unstructured data [31].

3.4 Networking techniques for BDA

3.4.1 Queuing theory
Queuing theory deals with the study of queues, which is a core concept in networking. In multiple stages of BDA, including data collection, analysis, and output generation; queuing theory can help in optimizing the performance and resource utilization. It can be availed to minimize delays and maximize network throughput by optimizing the arrival rate, service rate, queue length, service time, etc. In [6], queuing theory is availed for request time processing in the in-memory system of BD, which avails queuing representations. Moreover, queuing analysis can be availed for habituative censoring in BD processing, where it is used for modeling a terminal element with plentiful sensor elements where the non-examined data can be depicted in the form of a queue [32].

3.4.2 Coflow scheduling
Coflow (collection of flows) scheduling is a networking technique where communication patterns among multiple tasks in a distributed computing network environment are optimized to minimize communication delays among the nodes. It further involves grouping, prioritizing, and scheduling similar flows together in parallel clusters. VirtCO is a scheme that jointly schedules coflows and does virtual machine placement in an effort to optimize the performance of communication-intensive BD applications of data centers [33]. For BDA in data center networks, a coflow compression minimizing coflow completion time problem is solved using a heuristic technique availing multi-stage pipelining to elevate time-planning and recurrent Neural Networks (NN) to envisage compression rate [34].

3.4.3 Systematic modeling
Systematic modeling in the context of networking refers to creating a mathematical or analytical model to understand the behavior of the network [35]. In particular, it can be availed to predict how changes in the network parameters, just like bandwidth, latency, etc., will affect the data processing. In [36], for a mobile cellular network, the base stations' activities and behaviors are captured by using a BD analytical model to examine call detail records in an effort to predict call behaviors and analyze detail patterns. The Pareto principle has been advocated as an operational mechanism to model BD apps in networking. Pareto distribution has been availed to model extreme events in BD when analyzing BD in a longitudinal and cross-sectional approach using a non-parametric Bayesian formulation [37].

3.5 BDA applications in networking

3.5.1 Network monitoring
Network monitoring within circumstances of BDA involves collecting and analyzing data about network traffic, performance, network behavior, user events, sensor data, policies, configuration data, etc. in an effort to make network decisions to optimize network performance with the aid of network administrators or using an automatic decision-making system. Thus, network monitoring is a high-level activity that encapsulates all monitoring tasks related to BDA in networking, such as intrusion detection, anomaly detection, fault detection, and traffic prediction. Big-DAMA is such a generic network BD traffic monitoring and analysis framework for anomaly detection, intrusion detection, etc. that has both online and offline analysis approaches for real-time and historical BD in a scalable manner [38].

3.5.2 Intrusion detection
A vast amount of data from the network, such as user logs, performance metrics, network traffic, etc., can be minutely evaluated availing BDA to identify trends and relationships and detect malicious or unauthorized activities in the network to trigger alerts for mitigating them. In [39], spark-chi Support Vector Machines (SVMs) are used for BDA in an effort to detect intrusions by using a chi-squared selector for feature selection on the Apache Spark BD platform.

3.5.3 Anomaly detection
Anomaly detection, being an event detection technique, attempts to detect behaviors that deviate from normal behavior by analyzing BD related to network events by using statistical methods, ML techniques, data mining, etc. Thus, anomaly detection enables not only the detection of security anomalies for fraud detection, but also can be availed for the detection of equipment anomalies. In particular, in a cellular network, BDA has been availed to analyze call detail records to detect abnormal activities of user movements and identify the location of the device where the anomaly has occurred [40].

3.5.4 Fault detection
By analyzing BD such as sensor data, operational logs, etc., faults that exist in the network machinery can be detected. Thus, anomaly detection can be availed for network device fault detection. In particular, in [41], sensor anomaly detection is realized to analyze BD using ML, where sensor faults are detected as anomalies. However, for fault detection, other techniques just like signature anchored detection, rule-anchored detection, root cause analysis, time-series analysis, etc. can be availed. Moreover, data mining can be availed to collect BD in order to find fault nodes in a communication network using outlier detection in a route from source to destination [42].

3.5.5 Channel and spectrum learning
Channel and spectrum learning using BDA is useful in wireless communication networks to optimize channel

selection and allocate spectrum for efficient and reliable data transmission. BDA can be availed to examine usage trends, channel quality, interference metrics, spectrum-related data, etc. to dynamically learn about the channels and spectrum in an effort to allocate radio resources anchored on real-time demands. For instance, in underwater acoustic networks, data can be analyzed to- -predict communication channel quality [43]. In particular, in [44], the spectrum is monitored by analyzing BD, availing AI tactics and statistics to learn about the spectrum and forecast its occupancy and usage in an effort to make timely decisions in allotting radio spectrum dynamically in cloud-anchored wireless access networks.

Table 1. A condensation of prevailing literature on BD in networking.

Big data characteristic	**Specific characteristic**	**Procedure**	**Performance**
Distributed networking	Storage	Google file system [25]	Robust and scalable in handling BD
	Processing	MapReduce [26]	Simplified data analysis
BD modeling	Graph	BD graph data modeling [28]	Efficient and scalable
		Graph spectra and graph summarization techniques [29]	No performance analysis
	Matrix	Sparce matrices [30]	Efficient BD storage
	Tensor	Tensor representation and dimension reduction [31]	93% general accuracy
Networking techniques	Queuing theory	Request time processing using queuing models [6]	MG1 model requires least memory size
		Queuing theory for adaptive censoring for BD processing [32]	Successfully deal overcome data congestion
	Coflow scheduling	VirtCO-schedules coflows, place virtual machine [33]	36.5% reduction in coflow scheduling time
		Coflow compression using stage pipelining and RNN [34]	Speed up job completion time, reduces traffic
	Systematic modeling	BD analytical model to examine call detail records [36]	Effective network activity prediction using BD
		Pareto distribution with non-parametric Bayesian formula [37]	For climate events, predict extreme event risk
BD analytics applications	Network monitoring	Traffic monitoring and analysis (Big-DAMA) [38]	Realtime, historical scalable BD analysis
	Intrusion detection	Spark-chi SVMs [39]	High performance with reduced training time
	Anomaly detection	Analyze call detail records to detect movement anomaly [40]	Detect location of the anomaly
	Fault detection	Machine learning to detect sensor faults [41]	Able to extract meanings from complex BD
		Data mining for outlier detection in a route [42]	Feasible outlier detection in a multivariate data
	Channel & spectrum learning	Machine learning and statistics to learn on spectrum [44]	There is a trade-off in accuracy vs. latency
	Traffic prediction	BD regularity and randomness - time series analysis [45]	Regularity-predictable, randomness-unpredictable
	Network optimization	Resource allocation for life threatening incident [4]	Low delay critical data transmission
		Fuzzy logic for optimum radio resource allocation [46]	Efficient resource allocation
	Caching content	Popularity-anchored proactive content caching [47]	Full request satisfaction, offload 98% backhaul
		Location specific frequently accessed content caching [48]	Reduce network latency

3.5.6 Traffic prediction
Traffic prediction involves forecasting network traffic anchored on historical traffic data. Traffic prediction is useful for capacity planning to make the network ready to accept future high demands in network traffic by allocating the required resources at the correct times. Traffic prediction often involves spatial and temporal traffic analysis. In particular, in [45], mobile network BD traffic is evaluated by using time series analysis to divide BD traffic into randomness and regularity sections and avail temporal data envisaging to foretell traffic trends availing regularity component.

3.5.7 Network optimization
Network optimization deals with optimizing network parameters anchored on the predictive evaluation of BD (output of the network monitoring task). By analyzing user behavior, network traffic, and performance metrics, bottlenecks in the network can be identified, and resources such as bandwidth, spectrum, etc., can be optimally allocated, devices can be reconfigured, routing decisions can be optimized, and loads can be balanced to reduce the performance bottlenecks. Thus, network optimization belongs to the interpretation and decision-making step of BDA (refer to Section 4.5). In patient-anchored mobile networks, BD extracted from outpatient medical records and body sensor measurements are evaluated to envisage the probability of a fatal incident, and accordingly, an optimum 4G physical resource block is allocated in the network to transfer crucial data to medical suppliers with a lower time latency [4]. Some have used the MapReduce technique to lower the volume of BD and use orthogonal binary singular value decomposition of binary matrices representing BD in an effort to recognize patterns and group using fuzzy logic for optimum radio resource allocation of mobile 5G users [46].

3.5.8 Caching content
BDA can be availed to optimize BD transmission by using geographically targeted data caching and predictive caching. Thus, content caching is a type of network optimization. Proactive caching involves analyzing the user content for popularity, preferences, and other factors and caching more popular content in an effort to make them readily available to the users by analyzing user patterns. In [47], proactive content

caching at the base stations of a 5G network by collecting BD related to users' mobile traffic and estimating content popularity to optimize the network by achieving gains in backhaul offloading and request satisfaction has been advocated. Moreover, location specific content caching attempts to reduce network latency by caching frequently accessed content in geographical locations closer to the users [48].

Table 1 reveals a condensation of the prevailing literature on BD in networking.

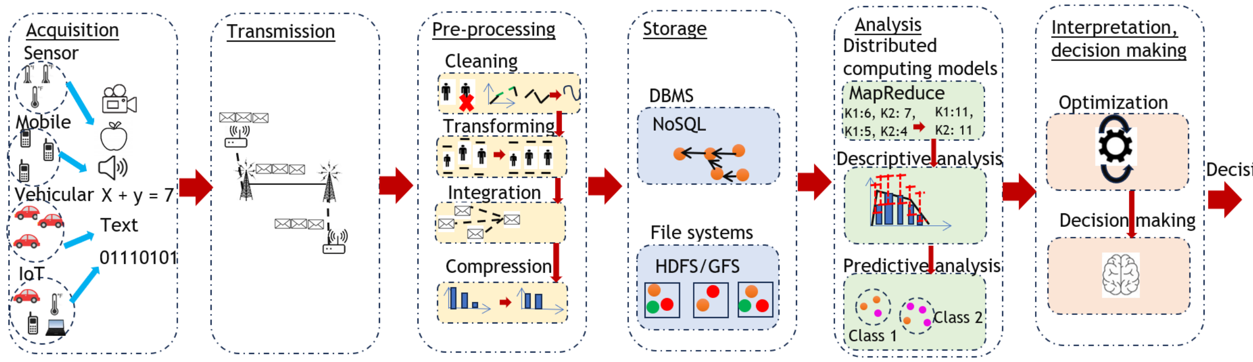


Fig. 3. Big data analytics routine.

4. A Brief review of BDA

BDA participates in the deep examination of massive, complicated data collections to take out knowledge using analysis, and this routine is illustratively revealed in Figure 3.

4.1 Acquisition

The inceptive stride of BDA is data acquisition. It refers to raw data collection from diverse origins, as in sensing elements, mobile appliances, operations, etc., using dedicated data collection approaches for each type of data collection [49]. In the BD scenario, it can be arduous to collect valuable and relevant data owing to the elevated veracity and variety of BD. In [3], for an unmanned aerial vehicular network, a cloud control system is modeled, and the stable condition of the control network is decided by identifying a core-association among the BD by obtaining the speed of sensing elements' data and the robustness of the network. A well-known data acquisition protocol is the Advance Messaging Queuing Protocol (AQMP), which is an open protocol that operates on four layers: transport, messaging, transaction, and security [50].

4.2 Transmission/Sharing

Once BD are acquired from the BD generation origins, they are required to be transmitted to servers or devices on which BD are stored, and subsequent analysis is carried out. Because of the exceptionally elevated amount and production rate of BD, it can be arduous to transmit data using conventional techniques. Thus, efficient transmission techniques, just like real-time data streaming, along with a high-speed data transfer service, can be availed. Apache Kafka is a BD streaming pipeline that is configurable and has features such as offset and partition that can be availed for program scaling [51]. FSA-MPTCP is a BD transmission scheme that provides optimum procedures for diverse data streams while providing an inter-tier methodology for a multiple path transmission control protocol to autonomously render BD transferring procedures [52].

4.3 Pre-processing

Owing to the elevated heterogeneity and veracity of BD, raw data cannot be directly utilized for analysis purposes. Raw BD should necessarily be pre-processed to be ready for analysis or storage. BD pre-processing typically involves enhancing data quality and standardizing data, which typically involves 4 steps:

4.3.1 Cleaning

Data cleaning is the first step of data pre-processing, which involves removing duplicate records, interpolating missing values, mitigating data noise, eliminating inconsistencies, and correcting inaccuracies. In [53], in the industrial center cloud, a BD cleaning method with the aid of mobile edge nodes using angle-anchored outlier detection and SVMs has helped to reduce energy and bandwidth consumption.

4.3.2 Transforming

Transformation involves converting the cleaned data into a regular arrangement, as in the normalization of data to a common scale. A BD quality ensuring framework emphasizes the requirement of data transformation following cleaning to ameliorate the condition of BD [54].

4.3.3 Integration

Integration essentially means data aggregation by combining data from numerous origins to create a consolidated dataset. In [55], a secrecy-conserving BD fusion scheme is availed in a network of unwired sensor elements where sensor elements are divided into clusters and inter- and intra-cluster data aggregation is carried out in an energy-efficient manner in the data pre-processing stage.

4.3.4 Compression

Data compression attempts to reduce the volume of data while preserving essential characteristics of the data at the same time. This involves reducing the number of attributes, balancing skewed data, and compressing data using techniques just like principal component analysis [56].

4.4 Storage

Pre-processed data can be stored either on a long-term or short-term basis using storage infrastructure. Data management software in BD that includes file systems and database management systems is briefly discussed in the following subsections.

4.4.1 File systems

Within the scope of BD, traditional file systems are not suitable to storing BD due to their limitations in scalability. Hadoop Distributed File System (HDFS) is one such example of a scalable and fault-tolerant BD file system that breaks files into smaller blocks and distributes them across nodes [57]. Another example is the Google File System (GFS), which breaks large BD into equal-sized, smaller blocks and distributes them [25].

4.4.2 Database management systems (DBMS)
Traditional relational databases are not suitable for BD storage owing to their boosted speed and diversity features. Thus, typically, NoSQL databases such as Apache Hbase (DBMS for HDFS), Cassandra, MongoDB, etc. that are capable of handling semi-structured and unstructured data having high scalability and consistency are availed [58].

4.5 Analysis
Analysis is the core of the BDA routine that extracts meaningful insights from the BD.

4.5.1 Distributed computing models
Batch processing involves processing an extensive quantity of data in batches by breaking data into smaller chunks and processing them parallelly across multiple nodes in a cluster. Distributed computing models often make use of batch processing. MapReduce is the most famous general-purpose distributed computing model that maps (sorts key-value pairs) input data and then reduces data by aggregating the output of the map function [26]. Apache Spark and Apache Flink are alternatives to MapReduce, supporting batch and stream processing having iterative algorithms and queries [59].

4.5.2 Descriptive Analysis
Descriptive analytics involves extracting the properties or overview of data, such as summarizing the data to gain an understanding of its characteristics, patterns, relationships, etc. This involves an initial assessment of the data using data mining. Several descriptive analysis techniques exist.

- Exploratory analysis involves visualizing data using graphical means such as histograms and scatter plots in an effort to identify trends, outliers, and relationships. In particular, the tile-anchored visual data analysis system developed to visualize the BD Twitter dataset uses exploratory data analysis to provide interactive visualization [60].
- Association rule mining detects patterns of co-occurrence among data items. It aims to extract the dependence between variables. Examples are the apriori algorithm and the frequent pattern growth algorithm. In [61], a Swarm Intelligence Optimization (SIO)-anchored growth procedure has been availed for discovering association rules from BD.
- Clustering groups similar data items together anchored on specific characteristics to ensure that items in the same category have a high similarity. Unsupervised ML methods, just like K-means and density-anchored clustering, are typical examples. Hybrid approaches for clustering by combining clustering techniques have resulted in better clustering performance than individual clustering techniques [62].
- Sequential pattern matching identifies patterns that occur in a sequence. In particular, this enables analyzing the sequence of user actions, tracking the order of a process, etc. An example of such an algorithm is a generalized sequential pattern. In [63], Knuth Morris Pratt anchored sequential pattern matching is implemented in the Hadoop distributed file system, which also involves partitioning and clustering of text documents where pattern mining tasks are implemented as map operations.
- Descriptive statistics render insight into the statistical properties of the dataset. In particular, R-tool has been advocated as a statistical data analysis framework for BD that is open source and provides packages for statistical analysis in an effort to make decisions anchored on the analysis [64]. Thus, R supports BD analysis using descriptive statistics.

4.5.3 Predictive Analysis
The aim of predictive analysis is to avail antecedent records to envisage eventual consequences or tendencies [65]. Thus, predictive analysis involves forecasting. There are several techniques for forecasting using predictive analysis.

- Classification is essentially a guided learning approach that assigns an assemblage of items to target categories (discrete outputs). Typical classification algorithms are supervised ML methods just like SVMs, deep NNs, etc., having different intentions such as intrusion recognition, network appliance classification, traffic classification, etc. [66].
- Regression attempts to estimate an operation to identify the association amidst objective attributes. It is used for predicting continuous figures, as in cases of linear and logistic. In particular, in [67], an apriori algorithm mined household BD in a mobile network is classified using descriptive analytics, and then non-linear logistic regression is used for identifying the poor households.
- Inferential statistics make inferences about a population (a larger dataset) using a sample of data (an observed dataset). Again, R-tool can be availed for the generation of inferential statistics for BD [64].
- Stochastic modeling techniques, just like Markov models, Bayesian networks, etc., model uncertainty and estimate parameters using probabilistic techniques using a sample of data. In particular, in [68], BD from a transportation network is analyzed using Bayesian networks to predict the bus arrival time.

Table 2. A condensation of prevailing literature on BDA.

BD analytics step	BD analytic component/aspect	Specific procedure	Performance
Acquisition	Acquisition rate	Identify relationship in BD acquisition rate with stability [3]	Raising of bandwidth curtails scheme's reaction time
	Data acquisition protocol	Advance Messaging Queuing Protocol (AQMP) [50]	High flexibility, security, ubiquity, interoperability
Transmission	Streaming pipeline	Apache Kafka having partitioning and offset [51]	Improves performance, accuracy of BD storing, processing
	Transmission scheme	Optimum strategies for multipath transmission control [52]	Better performance for diverse flows
Pre-processing	Cleaning	Angle-anchored outlier detection, SVM [53]	Reduced energy, bandwidth consumption
	Transforming	Improve quality of BD by transforming [54]	Save cost, help for accurate data analysis

	Integration	Inter- and intra-cluster data aggregation [55]	Scalable solution, protect privacy
	Compression	Reducing attributes, balancing skewed data, compression [56]	Improves data quality
Storage	File systems	HDFS [57]	Scalable and fault tolerant file system
		Google File System having distributed small blocks [25]	Robust and scalable in handling BD
	DBMS	NoSQL databases to handle semi, unstructured data [58]	High scalability, consistency
Analysis	Distributed computing models	MapReduce to map input data, aggregate output [26]	Simplified data analysis
		Spark, Flink - batch and stream processing [59]	Spark has better runtimes than Flink
	Descriptive analysis	Tile anchored visual data analysis [60]	Enables exploratory data analysis of massive data
		SIO for discovering association rules [61]	Effective rule mining
		Hybrid clustering combining clustering techniques [62]	High accuracy, recall, precision, F1-score
		Knuth Morris Pratt anchored sequential pattern matching [63]	Effective MapReduce model on time requirements
		R-tool for descriptive statistics analysis [64]	Capable of analyzing various data types
	Predictive analysis	Supervised ML for classification [66]	SVM has a good trade off in accuracy, computational cost
		Non-linear logistic regression [67]	92% recognition accuracy, only 18s mining time
		R-tool for inferential statistics analysis [64]	Capable of analyzing various data types
		Bayesian network as a stochastic modeling technique [68]	Effective predictive analysis for a use case
Interpretation and decision making	Optimization techniques	Network optimization using BDA [4]	High SINR, low margin of error
	Decision making module	Makes optimized decisions anchored on BD analysis [69]	30% throughput increment, 25% reduction in response time

4.6 Interpretation and decision making
The outcome of the predictive analysis is interpreted, and action recommendations are provided to achieve the desired goal. It involves making decisions anchored on predictions. Optimization and decision-making are the main components of this step of BDA.

- Optimization techniques involve linear programming, non-linear programming, genetic algorithms, swarm optimization, etc. to minimize or maximize a set of outcomes under a set of constraints to find the optimum solution for achieving tasks such as resource allocation, scheduling, variable selection, etc. [4].
- A decision-making module such as multiple-criteria decision-making assesses the predictive outcomes to provide a set of alternative solutions and selects the best solution among them. In particular, in self-optimizing networks, the BD-driven decision-making module takes optimization decisions such as adjusting network parameters and allocating resources, anchored on the insight from the deep examination of BD [69].

Table 2 reveals a condensation of the prevailing literature on BDA.

5. A Brief review on Blockchain and AI

5.1 A Condensation of Blockchain System
A progression of bonded blocks or interactions/transactions incorporates the digital ledger identified as blockchain.

5.1.1 Composition
Every one of block in the realm of a structured blockchain, which incorporates a payload division and header division, is bonded to its foregoing block (other than the foundation block) by harnessing the foregoing block's hash code, and the interactions/transactions in the realm of a payload division are arranged within a Merkle tree paradigm [8].

A non-linear blockchain incorporates an aggregate of bonded interactions/transactions, where one interaction/transaction could verify diverse additional interactions/transactions that were produced prior to its creation. These interactions/transactions are scant in payload divisions and header divisions; as a result, Merkle trees are not there [9].

5.1.2 Transactions/Interactions
An end-user can kick-off a blockchain transaction/interaction, which is consequently disseminated to all contemporaries in the realm of the network and safeguarded by harnessing the owner's key. A consensus mechanism will kick-off once each end-user harnesses the public asymmetric key to authenticate the transaction/interaction. Block producers commonly interact using consensus/general consent by joining the transaction/interaction in the realm of a block, which is consequently disseminated to the digital ledger network and assisted by each end-user in the digital ledger network with consecutive block authentication [70]. The transaction verification and generic transaction routine in a blockchain are illustratively revealed in Figure 4.

5.1.3 Cryptography
To protect the credibility of interactions/transactions in blockchain, hash encryption is harnessed to bestow fixed proportion hash codes with diminished clashes [71].

By harnessing an online signature, an asymmetrical key cryptographic method holding a private-public key combination is harnessed to authenticate interactions/transactions. In an effort to fortify the concealment of records, it's equally likely to be harnessed to conceal blockchain interactions/transactions [72].

Zero-knowledge-SNARKs are harnessed to authenticate interactions'/transactions' accuracy in a clandestine fashion, providing the user-specific details of interactions/transactions, fortifying concealment, and obstructing the dissemination of classified content [107].

Cryptography beyond quantum threats harnesses powerful cryptographic strategies that fortify against incursions from quantum computational devices, for instance, Kyber, Ed25519, and such [11].

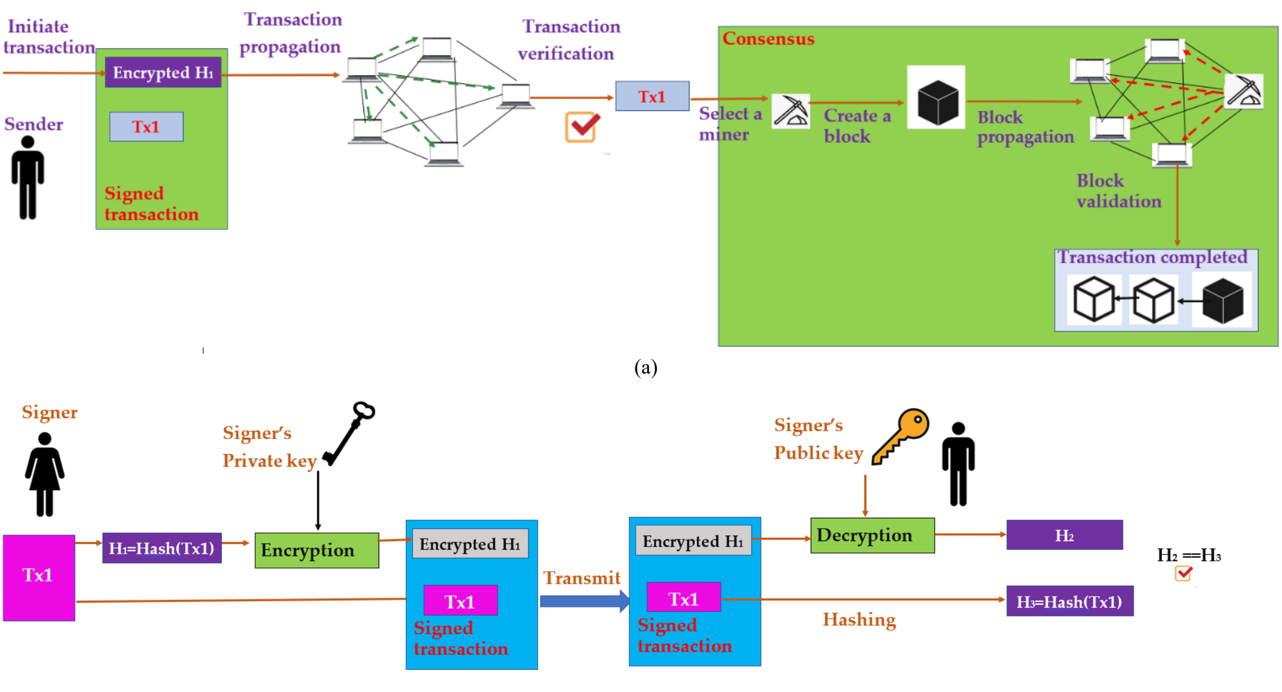


Fig. 4. Transactions of blockchain. (a) Transaction verification using digital signature. (b) Generic transaction process.

5.1.4 Consensus/General consent

Blockchain consensus harnesses widespread general consent to produce and authenticate unprecedented blocks, protecting the credibility of the digital ledger infrastructure.

In vote-anchored general consent, facts are transferred and procured in the realm of the contemporaries as they corporate to authenticate blocks. The universally adored vote-anchored general consent strategy harnesses byzantine fault-resilience consent, throughout which a supervisor joins interactions/transactions in the realm of a block, disseminates it, and end-users re-disseminate it to authenticate the block procured with the parent is equal [12]. When each end-user gets equal facsimiles of an unprecedented block by a greater than 66% majority of the network's end-users, the block is joined to the digital ledger.

Proof-anchored general consent requires end-users to bestow compelling demonstration on the grounds that they must be remunerated for joining an unprecedented block to the digital ledger. The most admired proof-anchored general consent strategy is denoted as proof-of-work, mandating an end-user to contribute effort by handling a daunting obstacle in an effort to protect its credibility [70].

5.2 AI

AI techniques involve a broad scope containing ML methods, fuzzy logic, meta-heuristics, and so on. ML is the most widely used sub-field in AI that involves supervised, semi-supervised, and unsupervised learning [73].

5.2.1 Supervised learning

In supervised learning, there exist labeled inputs and outputs for the learning model to comprehend a relationship that maps the input parameters to the output parameters (for model fitting/training). While training, underlying patterns are learned by the model. Typical examples of supervised ML involve classification and regression problems.

5.2.2 Unsupervised learning

Unsupervised learning involves learning entirely from an unlabeled data set, where the model typically learns the relationships among the input data to pick a bunch of inter-related objects from the fed-in data collection without any guidance.

5.2.3 Semi-supervised learning

In this experience-acquiring method, a mixed mode where both annotated and non-annotated data are availed for fitting a ML model.

6. Blockchain and AI for BDA

6.1 Notion

Built upon this review of literature, the blockchain- and AI-anchored BDA notion can be sectioned into the underneath 6 sections.

- C1 -- Exercising blockchain to store and transmit all BD securely, preventing data poisoning attacks with or without authentication and access control;
- C2 -- Mixed on- and off-blockchain BD recording;
- C3 -- Efficient BD storage in blockchain by offloading to the network edge or storing a selected subset of BD;
- C4 -- Analysis of BD using AI;
- C5 -- Blockchain-anchored secure federated learning for BDA;
- C6 -- Blockchain-anchored data fusion for BD;

The notion of BCandAI for BDA is revealed in Figure 5.

6.2 Blockchain only solutions for BDA

6.2.1 Secure BD acquisition/transmission/storage

For BDA in medical networks, blockchain has been availed for secure electronic medical record BD sharing and management by availing a cryptographic hash generator in the HDFS, using asymmetric cryptography for authentication of sensitive data, and using grey wolf optimization for validating the requests [74]. "MapChain" is a scalable BDA framework for IoT that avails a blockchain to provide verifiable access to healthcare data with the aid of zero-knowledge proofs to prevent unauthorized access, along with a BDA tracking system that is anchored on algorithms [75]. "SBBDA-IoD" is a blockchain-driven secure authentication key administration scheme drone in beyond 5G networks, along with a big-data analytics approach using data mining, predictive analytics, etc. that has shown high security and efficiency [14]. A BD sharing and transaction scheme using blockchain has been exercised for multiple users as a distributed secure storage, along with a cryptographic algorithm to prevent data modification during user storage [76]. A BD management system that collects, transmits, and stores BD has been feasible by integrating blockchain for non-tampering and traceability by availing of a data caching structure consisting of an on-blockchain and an off-blockchain to address data redundancy and storage problems associated with BD [15]. So as to facilitate BD sharing in edge networks, blockchain has been availed in a resource-efficient approach by using proof-of-collaboration consensus, which is computationally efficient, a blockchain transaction offloading mechanism, and express transactions and hollow blocks [16]. Driven by the non-scalability of storing all BD in blockchains, a scalable solution for storing only sensitive BD in the blockchain while using 3 algorithms: allocating data to users, data searching, and confirming communication has been feasible for the Smart City (SC) IoT system [77].

6.3 AI only solutions for BDA

6.3.1 AI-anchored Intrusion detection

So as to efficiently detect intrusions by BD analysis in network environments, a mixed Deep Learning (DL) algorithm comprised of a convolutional NN to draw out meaningful insights from the BD and a Long-Short Term Memory (LSTM) to sustain dependencies in the long term has been more effective than traditional ML approaches [78]. Similarly, in [79], LSTM and DL are jointly availed for a network intrusion detection technique along with BDA techniques to improve detection speed and accuracy, known as BDL-IDS. Alternatively, Louati et al. have used a multi-agent system availing Deep Reinforcement Learning (DRL) on the Hadoop platform for parallel big-data analysis for the agent decision-maker to detect network intrusions in a distributed approach [80]. Using network BD, a generic ML framework has evaluated the effectiveness of several ML algorithms, as in Random Forest (RF), Decision Tree (DT), etc., to deeply evaluate the BD to uncover intrusions in a distributed manner using a new multi-objective feature selection approach in an effort to expand the lifespan (accuracy) of the ML tactic in the long run [81].

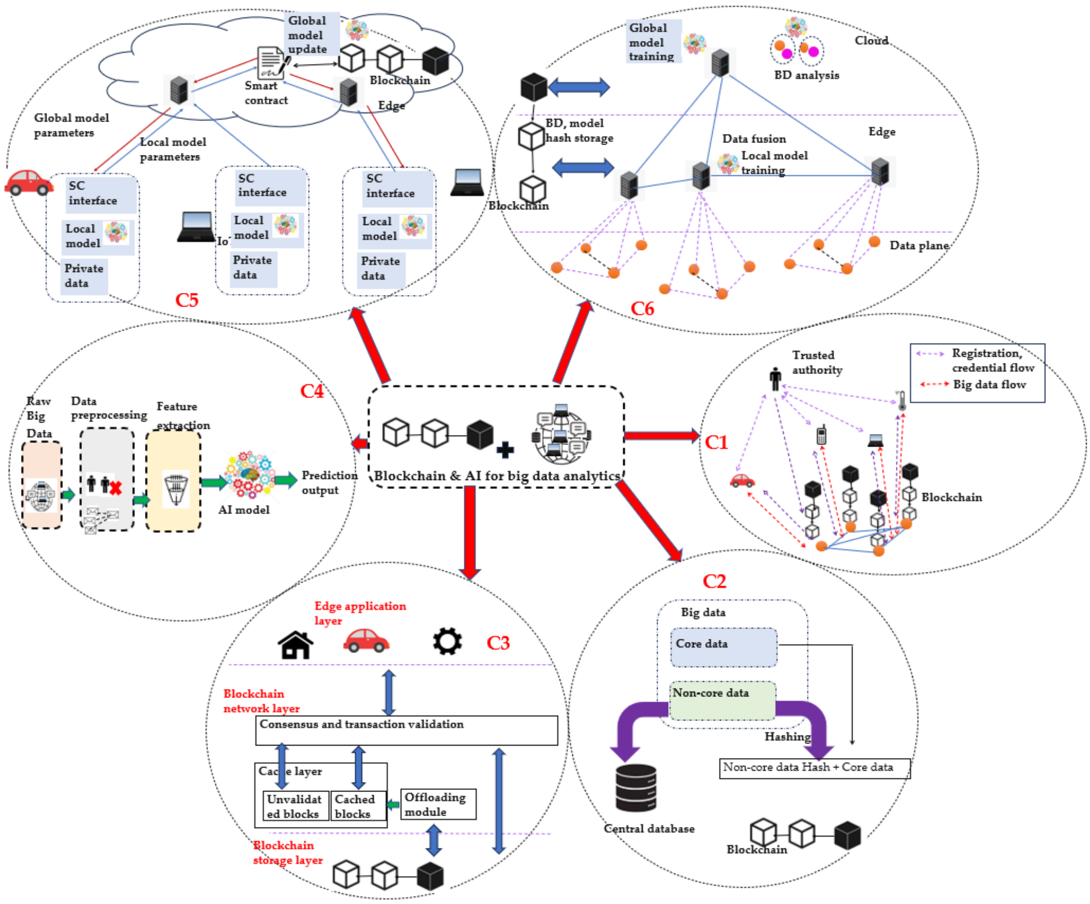


Fig. 5. The notion of BCandAI for BDA in networking.

6.3.2 AI-anchored Anomaly detection

In fourth-generation LTE networks, anomalous network conduct has been uncovered by using semi-supervised ML along with a statistical approach to detect sleeping cells (unusual low activity regions) and unusual high activity regions, which require more network resources, by analyzing BD generated from the cellular network [82]. Moreover, a BDA framework known as “Big-DAMA”, which is competent of deep examination and archiving BD, is able to process as batches and streams, where anomaly detection is realized using a set of ML models in an effort to benchmark each specific model [83]. Furthermore, the K-means unsupervised ML technique has been availed for BDA along with principal component analysis to detect network anomalies in Apache Spark clusters, which has resulted in a high accuracy of detection [84]. Similarly, BD related to user activity in mobile networks has been stored to analyze the anomalous behavior using the unsupervised K-means clustering ML technique along with hierarchical clustering that can detect anomalies such as unusual traffic at a specific location by analyzing the BD [85].

6.3.3 AI-anchored Fault detection

For efficient monitoring of the Internet of Things (IoT) in the manufacturing domain, using large unstructured BD from IoT sensors (Kafka, Storm, and MongoDB) for messaging, real-time processing, and storage, respectively, along with a mixed prediction model availing an outlier identification algorithm and a RF ML classifier for fault identification, has been effective [86]. For automatic fault diagnosis in software-defined networks, BDA, by harnessing Bayesian Networks (BN), has been availed as a self-diagnostic service along with a self-healing approach to either fix or bypass the failure [87].

6.3.4 AI-anchored Network monitoring

A combined ML approach by combining 5 ML classifiers, as in SVMs, DT, etc., and availing weighted, soft, and plurality voting for analyzing BD from a mobile IoT network for security monitoring has shown better classification performance than individual ML approaches [88]. Work in [89] renders an outline for preservation and then specifies how BD in mesh networks can be availed for network observation, non-normal behavior uncovering, and underlying fault evaluation availing ML and BN. In an IoT-anchored smart home automation and management system, BD are analyzed and monitored to learn the behavior of users and patterns of energy usage using the J48 ML technique for classification in an effort to provide energy-saving recommendations anchored on the analysis using rule languages [90].

6.3.5 AI-anchored content caching

“Learn to cache” is a generic ML-anchored BDA system for BDA at the network edge to escalate operating efficiency and lessen the requests for wireless spectrum where BD are estimated for content popularity and a caching approach is allocated accordingly [91]. Similarly, an anticipatory information caching method for 5G avails ML for analysis of mobile BD to predict item well-likeness, and items are cached at the base-station considering popularity to improve user satisfaction [92]. In software-defined and virtualized networks, a networking, caching, and computing resource allocation framework by analyzing BD using DRL (deep Q learning) has resulted in converged performance in an effort to improve applications related to smart cities [93].

6.3.6 AI-anchored Network optimization

A multimedia IoT network quality of experience optimization by availing data fusion of BD to map from uncontrolled user data to controlled network data using a NN for quality of experience assessment of the BD, where the fused data is optimized to control bandwidth allocation and path selection in an effort to improve user satisfaction [94].

6.3.7 AI-anchored Channel and spectrum learning

Channel learning, specifically a channel modeled using a finite-state Markov channel model, is learned using BD collected from wireless networks using DRL to obtain optimal interference alignment [95]. Moreover, ensemble learning anchored regression has been availed to analyze BD from a 5G network for intra-cell interference evaluation and optimize the network accordingly to improve user quality of experience and reduce interferences [96]. A framework known as “RFLearn” extracts spectrum knowledge from a wireless network by analyzing the network's I/Q BD samples using DL in an effort to make spectrum-anchored decisions such as changing communication parameters just like radio frequency, coding rate, modulation technique, etc. [97]. In wireless communication networks, DL has been availed for BDA to generate knowledge regarding mobile user demands using BD, which represents user spectrum and other demands, in an effort to intensify the effectiveness of channels [98].

6.3.8 AI-anchored Traffic prediction

Deep learning and DRL are jointly utilized along with unsupervised online incremental ML to diagnose content drifts, separately identify traffic events, forecast traffic flows, and optimize control decisions availing network-acquired heterogeneous BD [99]. Moreover, a convolutional NN is availed for spatial property recognition, while a LSTM recurrent NN is availed for temporal property recognition in traffic flow prediction by analyzing BD generated from SC networks in an effort to control the congestion of the network [100]. Similarly, spatio-temporal modeling using a hybrid DL technique consisting of an autoencoder-anchored DL algorithm to predict space-related properties and LSTM to predict time-related properties, along with parallel training for traffic load prediction in a cellular network, has shown high prediction accuracies [17].

6.3.9 AI-anchored Generic Network applications

A BDA system for the IoT-SC using DL and a parallel strategy of convolutional NNs where digital twins are availed to construct the SC has yielded high energy efficiency and prediction accuracy [101].

6.4 Combined BCandAI solutions for BDA

Blockchain has been advocated as a scheme for storing validated and quality BD for subsequent analytics using AI, where a use case related to a business network application has been advocated in which blockchain-stored BD is used for developing insights using federated AI [18]. A framework integrating cloud computing and blockchain is availed for less threatened and solitude conservative data fusion for BD in an effort to reduce network congestion, while a hierarchical fuzzy hashing technique has been availed for BDA in an effort to detect and locate ML models' anomalies [19]. So as to provide secure BDA services, blockchain is amalgamated with Federated Learning (FL) to ensure secure FL in IoT networks to protect the ML models' integrity while using fuzzy hashing to detect FL models' anomalies as a data analytics step [102]. In cognitive IoT networks, blockchain has been availed as a measure to prevent BD's data poisoning attacks, such that BDA availing AI/ML methods has shown good classification performance when BD is archived in blockchain compared to when data is archived in cloud storage under data poisoning attacks [103]. Similarly, another research for beyond 5G network-driven edge computing uses FL for privacy-preserving ML in BDA by integrating blockchain with Wasserstein generative adversarial networks to generate random noise-anchored differential privacy in an effort to protect model parameters [104]. “BlockDeepNet” is a framework for collaborative DL using blockchain for BD analysis, ensuring confidentiality and integrity, which allows local training at the equipment stage and combination of the locally fitted models at the edge-server stage, availing blockchain interactions [105]. Likewise, “BlockIoTIntelligence” is an IoT framework for BD analysis using AI by availing blockchain for making data immutable for decentralized IoT applications having four tiers: cloud, fog, edge, and device intelligence, where BCandAI are availed in each tier, achieving scalability and security [106]. In IoT networks in cyber-physical systems to achieve BD analysis with low latency and high accuracy while preventing the single spot of failure and adversarial attacks, “DeepBlockIoTNet” has been advocated as a blockchain-anchored framework that carries out DL at the edge tier in a reliable mode thanks to the blockchain [108].

Table 3. Evaluation of Blockchain- and AI-anchored BDA frameworks.

BCandAI involvement	Application	BD analytics stage	Procedure	Notion	Blockchain composition	Blockchain consensus	Blockchain kind	AI/BC techniques	Network kind	Performance
Pure blockchain anchored	Medical	Sharing	BC-CHG [74]	C1	Structured	Generic	Generic	Asymmetric cryptography	Medical	High efficiency and lower time
	Healthcare	All	MapChain [75]	C1	Structured	PoW	Public	Zero knowledge proofs	IoT	Solve privacy issues
	Drones	All	SBBDA-IoD [14]	C1	Structured	PBFT	Generic	Authentication key administration	5G	High security and efficiency
	Data sharing	sharing	Tamper-proof [76]	C1	Structured	Generic	Private	Cryptographic algorithm	Generic	Lower failure rate
	Generic	Acquisition, sharing	BDM [15]	C2	Structured	RBFT	Permissioned	On and off-chain	Generic	Low response time
	Generic	sharing	Open-edge [16]	C3	Structured	Po Collaboration	Generic	Offloading, express transactions	Edge	Low computational cost, wasted energy
	Smart city	Sharing	PP-BD [77]	C3	Structured	BFT	Permissioned	3 algorithms	IoT	Good BFT performance, preserve privacy
AI only	Intrusion detection	Analysis	Hybrid DL [78]	C4	-----	-----	-----	CNN+LSTM	Generic	More than 95% accuracy
		Analysis	BDL-IDS [79]	C4	-----	-----	-----	DL+LSTM	Generic	High detection rate, low false alarms
		Analysis	DIID [80]	C4	-----	-----	-----	Deep Reinforcement Learning (DRL)	Generic	Accuracy:75%, Precision:79%, F1-score-72%
		Analysis	MOA [81]	C4	-----	-----	-----	RF, DT, etc.	Generic	Longer model lifespan, 10Gbps throughput
	Anomaly detection	Analysis	AD-WN [82]	C4	-----	-----	-----	Semi-supervised ML	4G-LTE	Timely and reliable detection
		Storage, Analysis	Big-DAMA [83]	C4	-----	-----	-----	Supervised ML	Generic	10 times speed in computations
		Analysis	Netflow [84]	C4	-----	-----	-----	K-Means + PCA	Public network	96% anomaly detection accuracy
		Storage, analysis	User-privacy [85]	C4	-----	-----	-----	K-means, hierarchical clustering	Wireless	Low prediction error in anomaly-free data
	Fault detection	All	Hybrid-prediction [86]	C4	-----	-----	-----	Outlier detection, RF	IoT-automative	Efficient monitoring, high accuracy
		Analysis, interpretation	Bayesian [87]	C4	-----	-----	-----	Bayesian networks	SDN	high accuracy, precision, recall, F1-score
	Network monitoring	Analysis	Combined ML for MIoT-BD [88]	C4	-----	-----	-----	SVM, KNN, GNB, DT, and plurality voting	Mobile IoT	Combined technique has better classification performance
		Storage, analysis	Mesh-BDA [89]	C4	-----	-----	-----	ML, Bayesian network	Mesh	Practically viable architecture
		Analysis, interpretation	HEMS-IoT [90]	C4	-----	-----	-----	J48 ML classifier	IoT-SHA	Ensure safety, comfort, low energy consumption
	Caching content	All	Learn to cache [91]	C4	-----	-----	-----	Deep learning	Edge	Low CPU computations, high user satisfaction
		All	ML-backhaul [92]	C4	-----	-----	-----	Machine learning	5G	High user satisfaction, backhaul offloading
		Analysis, interpretation	DRL-SDVN [93]	C4	-----	-----	-----	Deep reinforcement learning	SDVN	Solution converges, utility decreases with CPU cycles
	Network optimization	Analysis, interpretation	QoEDF [94]	C4	-----	-----	-----	Neural network	IoT	Improvements in QoE levels
	Channel and spectrum sharing	Analysis, interpretation	FSMC [95]	C4	-----	-----	-----	Deep reinforcement learning	Wireless	Converges, high sum rate
		Analysis, interpretation	MNO [96]	C4	-----	-----	-----	Ensemble learning anchored regression	5G	Improved user QoS, throughput
		Analysis	RFLearn [97]	C4	-----	-----	-----	Deep learning	Wireless	Decrease latency and power consumption by 15 times
		Analysis, interpretation	WC-BDA [98]	C4	-----	-----	-----	Deep learning	Wireless	Implemented for underwater sensor networks
	Traffic prediction	Analysis, interpretation	Smart traffic [99]	C4	-----	-----	-----	DL+DRL, unsupervised incremental ML	ITS	Demonstrated on smart sensor network traffic data
		Analysis, interpretation	Urban-BDF [100]	C4	-----	-----	-----	CNN+LSTM	Smart city	RMSE-49, 92.3% accuracy
		Analysis	Spatio-temporal [17]	C4	-----	-----	-----	Autoencoder-anchored DL+LSTM	Cellular	Improved prediction accuracy
	Generic	Analysis	Digital twin [101]	C4	-----	-----	-----	DL+CNN	IoT-SC	High energy efficiency, prediction accuracy
Combined BCandAI	Business	Storage, analysis	BC-BDA [18]	C5	Structured	PBFT	Permissioned	Federated AI	Business	Improved customer satisfaction
	Data fusion	All	EC-DF [19]	C6	Structured	Generic	Generic	Hierarchical fuzzy hashing	IoT	Low overhead and processing times
	Secure BD analytics	All	FML-BC [102]	C5	Structured	PoW	Private	Federated learning, fuzzy hashing	IoT	Prevent model poisoning attacks, low energy consumption
	Secure BD analytics	All	CIoT [103]	C1	Structured	PBFT	Public	ML classification, BC storage	Cognitive IoT	Prevent data poisoning attack to get high classification performance

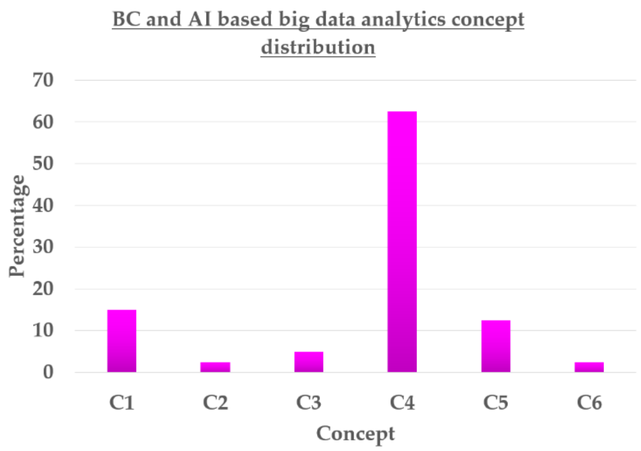
	Secure FL for BDA	All	B5G-BC [104]	C5	Structured	Custom	Generic	Wasserstein generative adversarial network	B5G-EC	High convergence, accuracy, good data utility
	Privacy preserving BDA	All	BlockDeepNet [105]	C5	Structured	PoW	Private	Collaborative DL	IoT	High accuracy, acceptable latency, overhead
	Multi-tier applications	All	BlockIoTIntelligence [106]	C1, C4	Structured	Generic	Generic	BC, AI in each tier	IoT	Good accuracy, security, privacy
	BD analytics in CPS	All	DeepBlockIoTNet [108]	C5	Structured	PoW	Private	Deep learning at edge	IoT-CP	low latency and high accuracy

7. Review Evaluation

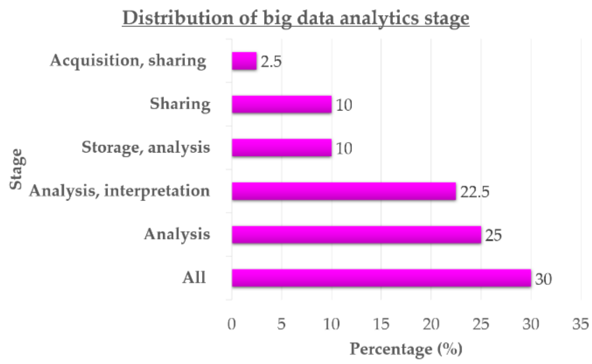
7.1 Evaluation of singular schemes
Table 3 reveals the BCandAI-anchored BDA frameworks' evaluation with respect to BDA concept, application, blockchain-related parameters, techniques, network type, performance, and time of proposal.

7.2 Overall Evaluation
Figure 6 illustratively reveals the dispersal of various attributes in evaluated BCandAI anchored BDA frameworks.

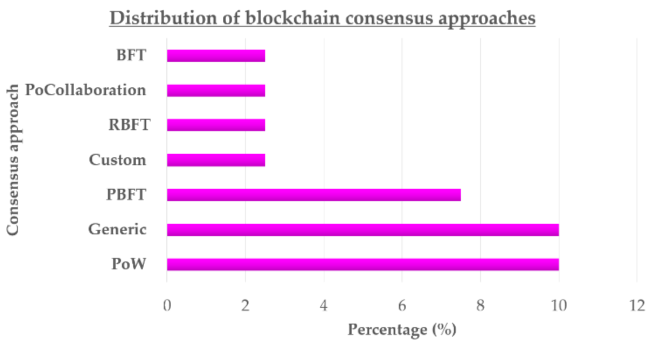
As revealed in Figure 6a, the greatest (62.5%) of the frameworks in the reviewed work implement concept C4 (BD analysis using AI), while the remaining 37.5% are distributed for BCandAI for BDA (C5-12.5%) or blockchain itself alone for analytics (C1-15%, C2-2.5%, C3-5%, and C6-2.5%). Next, when deliberating on the stage of the BD analytics dispersion among frameworks, as revealed in Figure 6b, it is transparent that 30% of frameworks are involved in the whole BDA routine, while other (70%) frameworks address one or more specific stages of BD analytics. Specifically, 25% address BC analysis, 22.5% are involved in both analysis and interpretation, 10% are involved in both storage and analysis, another 10% are involved in sharing, and 2.5% are involved in BD acquisition and sharing. Furthermore, out of the literature reviewed, the highest (10%) of frameworks implement PoW general consent, with the same percentage by generic general consent, 7.5% by PBFT, and remaining equal percentages by RBPT, custom, PoCollaboration, and BFT, as revealed in Figure 6c. Moreover, when deliberating on the specific technique for BD analytics, as revealed in Figure 6d, DL is the topmost predominantly used BD analysis technique (12.5%), succeeded by DRL (7.5%), CNN+LSTM (5%), and DL+LSTM (5%). Blockchain-anchored specific techniques, just like the use of asymmetric cryptography, zero-knowledge proofs, and the like, and BCandAI combined techniques, just like federated AI, BC storage with AI classification, and the like, have relatively lower implementation. Additionally, as revealed in Figure 6e, most blockchain or BD analytics frameworks have been either formulated for generic networks or IoT, as the highest percentage of 17.5% exists for such networks, succeeded by wireless, 5G, and edge networks, while remaining networks have the least prevalence. Finally, as revealed in Figure 6f, the literature associated with BC and AI for BDA commenced to be produced around 2016 and has been prevailing since then, achieving a peak in 2018 and prevailing still with high and low variations in between the years with no plain variation on the trend.



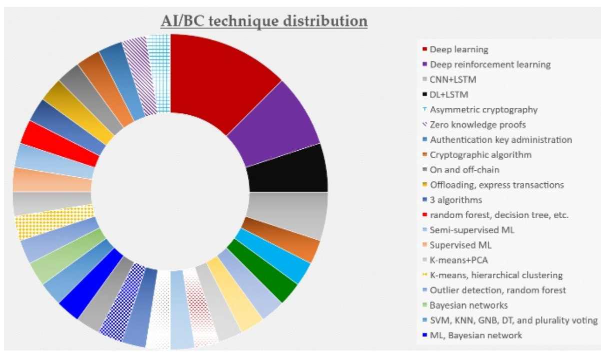
(a)



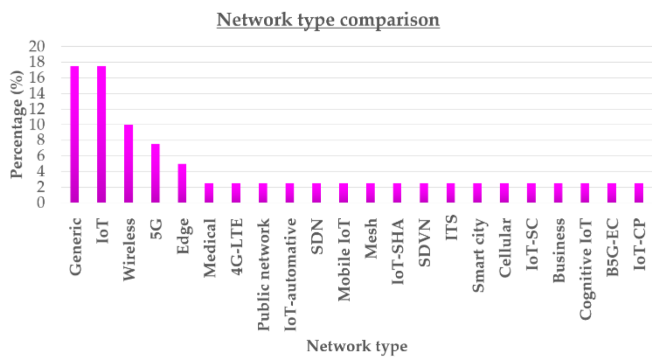
(b)



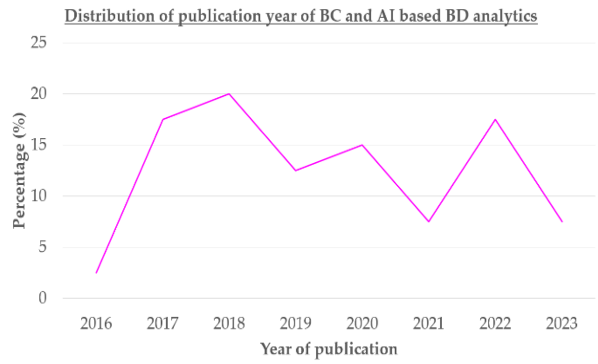
(c)



(d)



(e)



(f)

Fig. 6. Overall intricate review (a) Notion (b) BDA stage (c) BC general consent (d) BC/AI technique (e) Type of network (f) Posted year

8. Discussion

8.1 Potentialities

8.1.1 High support for BD analysis availing AI

AI methods are extensively availed in BD analysis to achieve different network objectives such as intrusion detection, anomaly detection, fault detection, network monitoring, content caching, network optimization, traffic prediction, etc. Furthermore, data fusion can be utilized to effectively reduce the quantity of data, making BDA more efficient [109]. AI-anchored prediction tasks in BD involve both descriptive analysis to extract an overview of data by using techniques just like clustering and predictive analysis techniques just like classification and regression to predict future outcomes anchored on historical data. AI techniques just like ML and fuzzy logic can be integrated with other analysis techniques just like descriptive statistics, inferential statistics, association rule mining, sequential pattern matching, etc. to improve the global analysis task [110].

8.1.2 Secure BD archiving availing blockchain

BD archiving is a phase in BDA. Blockchains can be availed to store well pre-processed BD that is cleaned, transformed, integrated, and compressed, as raw BD is not suitable to be stored in blockchains. Thus, validated and quality BD are stored in the blockchain to be availed for subsequent analysis. Moreover, blockchain can also be utilized for secure BD fusion. Investigations reveal that BDA, by availing data archived in the blockchain, is superior in performance to that stored in the cloud, fundamentally as a result of the data stored in the blockchain is free from data poisoning attacks. Furthermore, a network can be divided into several tiers, such as cloud, edge, fog, device, etc., where BCandAI can be integrated into each tier to achieve BD analysis in each tier. As BD is large in volume, off-chain storage can be assisted to efficiently archive BD, where the hash code of the data stored in off-chain storage can be stored in blockchain.

8.1.3 Secure FL availing blockchain

FL is a dispersed mode of ML that permits models to be trained across decentralized devices without exchanging local data. Blockchain-anchored BDA has paved the path for obtaining inferences using federated AI by analyzing BD locally. So as to achieve secure FL, blockchain protects the integrity of the ML models used for FL for BD analysis. Federated learning in BDA allows privacy preserving ML model training for analyzing BD, as private data are not shared in the network. Furthermore, local ML models can be aggregated at the edge servers using blockchain transactions.

8.1.4 High growth potential and automation of AI techniques

Traditional BDA techniques, just like statistical methods and data mining, can strive to analyze BD that have a large volume of high-velocity data in a large network. AI practices can be trained and tested on large datasets and, thus, have high growth potential. For instance, it has a high growth potential in analyzing network data and predicting real-time parameters to aid in network optimizations such as routing [111]. Moreover, they can be trained online as more new data becomes available in an automated fashion without the involvement of a third party, reducing manual intervention and human errors, in contrast to traditional techniques that require human supervision [112]. Furthermore, AI techniques are capable of producing inferences in actual-time availing BD to aid in making network decisions.

8.1.5 Secure BD sharing and authentication availing blockchain

Availing blockchain, cryptographic practices can be availed for authentication, and HDFS can be incorporated to share BD. Moreover, Zero-knowledge-SNARKs can be availed on blockchain to render admission control for BD. Furthermore, blockchain can be availed for secure authentication key administration in BDA frameworks. The proof-of-collaboration consensus technique has been advocated to be availed in collaborative BD sharing frameworks utilizing blockchain to share BD in an energy-efficient manner having express transactions.

8.2 Problems

8.2.1 Poor versatility of blockchain as a result of high data volume and velocity

BD inherits features of high data volume and velocity, meaning that large quantities of data that are generated at high speeds are required to be analyzed. Thus, BDA systems are obliged to be capable of handling an enormous bulk of BD per unit time. This can be challenging, especially when the network scale gets larger, further elevating the burden of data quantity per unit time and making data collection to a centralized point very difficult [113]. Therefore, blockchain frameworks will strive to provide their duties of secure storage/transmission/sharing of BD. This is very true, especially for conventional blockchain networks that can incur extra delays in BD processing as a result of time-consuming distributed consensus approaches that are intensified with the growth of network scale, velocity of data, and volume of data [114].

8.2.2 Problems as a result of high variety and veracity

BD can originate from heterogeneous origins, as in video, audio, text, images, etc., having large differences in the quality of the data. In knowledge-defined networks, data needs to be collected to a centralized location for analysis and knowledge generation, which can be challenging from a big data perspective when integrated with blockchain [115]. Due to this high variety and veracity, it is challenging for the blockchains to archive or transmit BD and AI to analyze the data. Thus, it will be difficult for the AI techniques to analyze the BD by integrating different varieties of data. Instead, multiple AI models may be required to predict different varieties of data, as different varieties of data exist to represent purposes [116]. Therefore, even though the effect of veracity can be reduced by proper cleaning of data, the effect of variety cannot be nullified. Moreover, blockchains often require standardized data structures for archiving data. Since raw BD can contain fully, partially, or zero unstructured data, it is arduous to store them on blockchain without proper pre-processing techniques just like cleaning and transforming.

8.2.3 Troubles in auditing and explaining analysis

When blockchain is used for secure storing and transmitting BD, and AI is used for analysis, it can be very arduous to evaluate and audit such BDA frameworks. First, even though blockchain transactions are traceable, auditors can find it difficult to understand the history of transactions and verify data accuracy. Moreover, blockchain networks can change over time having protocol updates and changes in consensus mechanisms, making it difficult to audit transactions over time. On the other hand, the inner workings of AI models are also difficult to interpret. For example, in the prediction of link lifetime using machine learning, it is difficult to interpret the cause of the prediction [117]. Moreover, it can be hard to predict whether data transformation and integration lead to better performance than data analysis using individual predictions for each data stream.

8.2.4 Elevated energy wastage

The whole routine of BD analysis is a high-energy demanding task, from data acquisition up to interpretation and decision-making. When BCandAI are integrated into the process, the energy consumption is even higher than using conventional BDA. Even though blockchain improves the purity, transparency, and trust of BD in the BDA routine, it can cause additional network energy wastage because of the peer-to-peer distributed consensus-anchored approach followed by them for transaction generation and validation. However, whether integration of AI boosts or reduces energy consumption is debatable, as it can depend on various factors such as the complexity of the AI model, hardware resources, parallel processing capability, etc.

8.2.5 Lack of systematizations for integrated BCandAI use

In BCandAI-integrated BDA systems, to the climax of our perception, there are no systematizations that recommend combinations of a particular blockchain framework and an AI model. In fact, the integration is still in the research stage, and these two techniques exist to serve different purposes in the BDA framework. However, as reviewed in the literature, there is sufficient evidence to prove that two techniques in combination have provided very promising results to provide secure, privacy-preserving, attack-preventing, and high accuracy BDA despite lack of systematizations.

8.2.6 High implementation challenges

First, there exist regulatory and compliance issues, like in the cases of liability issues for AI decisions and automated actions of smart contracts [118]. Next, maintaining data quality in a system of integrated BCandAI is difficult due to the high heterogeneity, velocity, and quantity that exist in big data [119]. Furthermore, the implementation cost is much higher in BCandAI combined BDA approaches than using AI alone for BDA, which incurs costs in terms of additional initial infrastructure and operational costs. Next, due to the integration of blockchain, it can introduce additional overheads in communication and additional latencies. Finally, the integration can negatively affect towards the overall system's scalability since both AI and blockchain are known to have scalability issues, making implementation challenging [120].

9. Conclusion, Counsels, and Impending trajectories

In this evaluation, we first evaluated the notion of BD in networking and then evaluated a brief of the BDA process. Post to a condensation of BCandAI concepts, we evaluated the application of BCandAI for BDA in networking. Specifically, built upon this review of literature, we sectioned the notion into 6 sections in which BCandAI can be availed for BDA: blockchain for archiving and transmitting BD securely, mixed on-blockchain and off-blockchain BD recording, BD analysis availing AI, secure FL availing blockchain for BDA, and blockchain anchored BD fusion. Moreover, we intricately reviewed the evaluated studies with regard to blockchain concerning attributes, AI concerning attributes, and BDA concerning attributes to evaluate shifts and breaks related to them. Finally, we discoursed the potentialities and problems of utilizing AI and blockchain for BDA.

This review article offers a profitable piece of knowledge on the concept of integrated and individual use of BCandAI for BDA. Using this evaluation, researchers can promptly evaluate shifts and breaks in the BCandAI application for BDA. This will grant a foundation for them to explore more in this realm stemming from the intricated review, findings, and counsels.

Underneath counsels can be advocated to overpower the problems discoursed.

- Conventional blockchain can strive to scale under the characteristics of high data volume and velocity of BD. So as to face these challenges, a non-linear blockchain with parallel arithmetic capability can be availed to replace the conventional blockchain. Moreover, if traditional blockchain is used, sharding can improve throughput and decrease delay. Furthermore, a hybrid approach with on-chain and off-chain storage can be availed under the demands of high quantities and rates of BD.
- The high veracity and heterogeneity of BD are challenging for both BCandAI. These effects of BD can be reduced by proper pre-processing techniques involving cleaning, transforming, integration, and compression. Data pre-processing is typically done off-chain; however, it is possible to pre-process data using smart contracts and integrate them into the blockchain. However, as the effect of variety cannot be nullified, AI may involve transfer learning or a hierarchy of AI models to fuse the results of outputs

from different varieties of data to predict the desired output of BDA.

- To overcome difficulties in auditing BDA, various methods can be availed. First, for blockchains, periodic audits are obliged to be carried out, as it is difficult to track the history of transactions over time as a result of the evolution of protocols and principles of consensus over time. When it comes to predictions and analysis using AI, it is advisable to use explainable AI models to evaluate the causes of prediction outputs and to understand how these models make decisions.
- So as to reduce the high energy consumption in traditional blockchain, one can avail energy-efficient consensus approaches, just like green practical byzantine fault tolerance. Moreover, an energy-efficient graph-anchored blockchain can be availed to reduce energy consumption for computations. Furthermore, dedicated hardware designed for blockchains can be availed as infrastructure for blockchains to improve the energy efficiency of secure BD storage using blockchain.
- So as to tackle the difficulty of the lack of standardizations for integrated use of BCandAI, they can be selected by considering the application-case of BDA and selecting the best combination by referring to existing research literature. Moreover, industry involvement and collaborations can be promoted to develop standards for use cases of BCandAI integrated BDA.

In the financial and banking sector, BCandAI-driven BDA can contribute to effective fraud detection and risk management, helping to improve the overall system's transparency. In healthcare, it helps in providing personalized treatments by analyzing a large quantity of medical data. From the viewpoint of supply chain and logistics, BDA using BCandAI can indeed contribute towards better demand prediction, route optimization, and inventory management. Blockchains can help in energy trading in decentralized energy grids, while AI can make that system more sustainable by analyzing big data to reduce energy wastage.

Blockchains can reinforce the credibility, secrecy, and reliability of the BDA, while AI can be availed to accurately produce outcomes by analyzing the BD descriptively or predictively. Later research may encompass developing a unified AI model for predicting diverse BD. Moreover, in blockchain terms, the influence of quantum processing attacks can be studied for the security and subsequent analysis of BD.

References

[1] G. Kapil, A. Agrawal, and R.A. Khan, "A study of big data characteristics," in *Proc. 2016 Inter. Conf. Commun. Electron. Syst. (ICCES)*, Coimbatore, India: IEEE, Oct. 2016, pp. 1-4. doi: 10.1109/CESYS.2016.7889917.

[2] A. Chebotko, A. Kashlev, and S. Lu, "A big data modeling methodology for Apache Cassandra," in *Proc. 2015 IEEE Inter. Congr. Big Data*, New York City, USA: IEEE, Jun. 2015, pp. 238-245. doi: 10.1109/BigDataCongress.2015.41.

[3] F. Luo, C. Jiang, S. Yu, J. Wang, Y. Li, and Y. Ren, "Stability of cloud-based UAV systems supporting big data acquisition and processing," *IEEE Trans. Cloud Comput.*, vol. 7, no. 3, pp. 866-877, Jul. 2019. doi: 10.1109/TCC.2017.2696529.

[4] M.S. Hadi, A.Q. Lawey, T.E. El-Gorashi, and J.M. Elmirghani, "Patient-centric cellular networks optimization using big data analytics," *IEEE Access*, vol. 7, pp. 49279-49296, Apr. 2019. doi: 10.1109/ACCESS.2019.2910224.

[5] A. Shi-Nash and D. R. Hardoon, "Data analytics and predictive analytics in the era of big data," in *Internet of Things and Data Analytics Handbook*, 1st ed., H. Geng, Ed., Wiley, 2017, pp. 329–345. doi: 10.1002/9781119173601.ch19.

[6] R. Srivastava, "Exploration of in-memory computing for big data analytics using queuing theory," in *Proc. 2nd Inter. Conf. High Performance Compilation, Comput. Commun.*, Hong Kong, China: ACM, Mar. 2018, pp. 11-16. doi: 10.1145/3195612.3195621.

[7] P.A.D.S.N. Wijesekara, and S. Gunawardena, "A Review of Blockchain Technology in Knowledge-Defined Networking, Its Application, Benefits, and Challenges," *Network*, vol. 3, no. 3, pp. 343-421, Aug. 2023. doi: 10.3390/network3030017.

[8] M. Wang, M. Duan, and J. Zhu, "Research on the security criteria of hash functions in the blockchain," in *Proc. 2nd ACM Works. Blockch, Cryptocurr. Contr.,* New York, USA: ACM, May 2018, pp. 47-55. doi: 10.1145/3205230.3205238.

[9] G. Sun, M. Dai, J. Sun, and H. Yu, "Voting-based decentralized consensus design for improving the efficiency and security of consortium blockchain," *IEEE Internet Things J.*, vol. *8*, no. 8, pp. 6257-6272, Oct. 2020. doi: 10.1109/JIOT.2020.3029781.

[10] L. Chen, W.K. Lee, C.C. Chang, K.K.R. Choo, and N. Zhang, "Blockchain based searchable encryption for electronic health record sharing," *Future Gener. Comput. Syst.*, vol. *95*, pp.420-429, Jun. 2019. doi: 10.1016/j.future.2019.01.018.

[11] P.A.D.S.N. Wijesekara, "A Review of Blockchain-Rooted Energy Administration in Networking," *Indonesian J. Comput. Sci.*, vol. 13, no. 2, pp. 1607-1642, Apr. 2024. doi: 10.33022/ijcs.v13i2.3818.

[12] I. Gutiérrez-Agüero, S. Anguita, X. Larrucea, A. Gomez-Goiri, and B. Urquizu, "Burnable pseudo-identity: A non-binding anonymous identity method for ethereum," *IEEE Access*, vol. *9*, pp.108912-108923, Jul. 2021. doi: 10.1109/ACCESS.2021.3101302.

[13] R. Yang, R. Wakefield, S. Lyu, S. Jayasuriya, F. Han, X. Yi, X. Yang, G. Amarasinghe, and S. Chen, "Public and private blockchain in construction business process and information integration," *Automat. Construction*, vol. *118*, Oct. 2020, Art. no. 103276. doi: 10.1016/j.autcon.2020.103276.

[14] A.K. Mishra, M. Wazid, D.P. Singh, A.K. Das, J. Singh, and A.V. Vasilakos, "Secure Blockchain-Enabled Authentication Key Management Framework with Big Data Analytics for Drones in Networks Beyond 5G Applications," *Drones*, vol. 7, no. 8, Art. no. 508, Aug. 2023, doi: 10.3390/drones7080508.

[15] J. Chen, Z. Lv, and H. Song, "Design of personnel big data management system based on blockchain," *Future Gener. Comput. Syst.*, vol. 101, pp. 1122-1129, Dec. 2019. doi: 10.1016/j.future.2019.07.037.

[16] C. Xu, K. Wang, P. Li, S. Guo, J. Luo, B. Ye, and M. Guo, "Making big data open in edges: A resource-efficient blockchain-based approach," *IEEE Trans. Parallel Distrib. Syst.*, vol. 30, no. 4, pp. 870-882, Apr. 2019. doi: 10.1109/TPDS.2018.2871449.

[17] J. Wang, J. Tang, Z. Xu, Y. Wang, G. Xue, X. Zhang, and D. Yang, "Spatiotemporal modeling and prediction in cellular networks: A big data enabled deep learning approach," in *Proc. IEEE Conf. Comput. Commun. (IEEE INFOCOM 2017)*, Atlanta, USA: IEEE, May 2017, pp. 1-9. doi: 10.1109/INFOCOM.2017.8057090.

[18] H.T. Vo, M. Mohania, D. Verma, and L. Mehedy, "Blockchain-powered big data analytics platform," in *Proc. Big Data Anal.: 6th Int. Conf. (BDA 2018)*, Warangal, India: Springer, Dec. 2018, pp. 15-32. doi: 10.1007/978-3-030-04780-1_2.

[19] J. Dong, C. Song, T. Zhang, Y. Li, and H. Zheng, "Integration of edge computing and blockchain for provision of data fusion and secure big data analysis for Internet of Things," *Wireless Commun. Mobile Comput.*, vol. 2022, pp. 1-9, May 2022. doi: 10.1155/2022/9233267.

[20] K. Rabah, "Convergence of AI, IoT, big data and blockchain: a review," *Lake Inst. J.*, vol. 1, no. 1, pp. 1-18, Mar. 2018.

[21] V.K. Chattu, "A review of artificial intelligence, big data, and blockchain technology applications in medicine and global health," *Big Data Cognit. Comput.*, vol. 5, no. 3, Art. no. 41, Sep. 2021, Art. no. 41, doi: 10.3390/bdcc5030041.

[22] R.G.J. Pablo, D.P. Roberto, S.U. Victor, G.R. Isabel, C. Paul, and O.R. Elizabeth, "Big data in the healthcare system: A synergy with artificial intelligence and blockchain technology," *J. Integr. Bioinformatics*, vol. 19, no. 1, Art. no. 20200035, Mar. 2022. doi: 10.1515/jib-2020-0035.

[23] N. Deepa, Q.V. Pham, D.C. Nguyen, S. Bhattacharya, B. Prabadevi, T.R. Gadekallu, P.K.R. Maddikunta, F. Fang, and P.N. Pathirana, "A survey on blockchain for big data: Approaches, opportunities, and future directions," *Future Gener. Comput. Syst.*, vol. 131, pp.209-226, Jun. 2022. doi: 10.1016/j.future.2022.01.017.

[24] P.A.D.S.N. Wijesekara, "A study in University of Ruhuna for investigating prevalence, risk factors and remedies for psychiatric illnesses among students," *Sci. Rep.*, vol. 12, no. 1, Art. no. 12763, Jul. 2022, doi: 10.1038/s41598-022-16838-4.

[25] J. Yang, "From Google file system to omega: a decade of advancement in big data management at Google," in *Proc. 2015 IEEE 1st Int. Conf. Big Data Comput. Service Applicat.*, Redwood City, USA: IEEE, Mar. 2015, pp. 249-255. doi: 10.1109/BigDataService.2015.47.

[26] S. Maitrey, and C.K. Jha, "MapReduce: simplified data analysis of big data," *Proc. Comput. Sci.,* vol. 57, pp. 563-571, Jan. 2015. doi: 10.1016/j.procs.2015.07.392.

[27] U.G. Gupta, and A. Gupta, "Vision: a missing key dimension in the 5V Big Data framework," *J. Int. Bus. Res. Market.*, vol. 1, no. 3, pp. 50-56, Mar. 2016. doi: 10.18775/jibrm.1849-8558.2016.13.3005.

[28] A. Kiourtis, P. Karamolegkos, A. Karabetian, K. Voulgaris, Y. Poulakis, A. Mavrogiorgou, and D. Kyriazis, "An autoscaling platform supporting graph data modelling big data analytics," *Stud. Health Tech. Inform.*, vol. 295, pp. 376-379, Jun. 2022. doi: 10.3233/SHTI220743.

[29] A. Khan, S.S. Bhowmick, and F. Bonchi, "Summarizing static and dynamic big graphs," in *Proc. VLDB Endowment*, Munich, Germany: VLDB Endowment Inc., vol. 10, no. 12, Aug. 2017, pp.1981-1984. doi: 10.14778/3137765.3137825.

[30] K. Atanassov, and V. Bureva, "Index matrix representation of Big data structures," in *Proc. Bulgarian Acad. Sci.*, vol. 75, no. 5, May 2022, pp. 719-725. doi: 10.7546/CRABS.2022.05.12.

[31] L. Kuang, F. Hao, L.T. Yang, M. Lin, C. Luo, and G. Min, "A tensor-based approach for big data representation and dimensionality reduction," *IEEE Trans. Emerging Topics Comput.*, vol. 2, no. 3, pp.280-291, Sep. 2014. doi: 10.1109/TETC.2014.2330516.

[32] H. Zhu, H. Qian, X. Luo, and Y. Yang, "Adaptive queuing censoring for big data processing," *IEEE Signal Process. Lett.*, vol. 25, no. 5, pp. 610-614, Mar. 2018. doi: 10.1109/LSP.2018.2815006.

[33] D. Shen, J. Luo, F. Dong, and J. Zhang, "VirtCo: joint coflow scheduling and virtual machine placement in cloud data centers," *Tsinghua Sci. Technol.*, vol. 24, no. 5, pp. 630-644, Apr. 2019, doi: 10.26599/TST.2018.9010098.

[34] Q. Zhou, K. Wang, P. Li, D. Zeng, S. Guo, B. Ye, and M. Guo, "Fast coflow scheduling via traffic compression and stage pipelining in datacenter networks," *IEEE Trans. Comput.*, vol. 68, no. 12, pp. 1755-1771, Dec. 2019. doi: 10.1109/TC.2019.2931716.

[35] P.A.D.S.N. Wijesekara, and Y.K. Wang, "A Mathematical Epidemiological Model (SEQIJRDS) to Recommend Public Health Interventions Related to COVID-19 in Sri Lanka," *COVID*, vol. 2, no. 6, pp. 793-826, Jun. 2022. doi: 10.3390/covid2060059.

[36] D. Jiang, L. Huo, and H. Song, "Rethinking behaviors and activities of base stations in mobile cellular networks based on big data analysis," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 1, pp. 80-90, Jan. 2020. doi: 10.1109/TNSE.2018.2861388.

[37] S. Shenoy, and D. Gorinevsky, "Predictive analytics for extreme events in big data," in *Proc. 2015 IEEE 1st Int. Conf. Big Data Comput. Service Applicat.*, Redwood City, USA: IEEE, Mar. 2015, pp. 184-193. doi: 10.1109/BigDataService.2015.66.

[38] P. Casas, A. D'Alconzo, T. Zseby, and M. Mellia, "Big-DAMA: big data analytics for network traffic monitoring and analysis," in *Proc. 2016 Workshop Fostering Latin-American Res. Data Commun. Netw.*, Florianopolis, Brazil: ACM, Aug. 2016, pp. 1-3. doi: 10.1145/2940116.2940117.

[39] S.M. Othman, F.M. Ba-Alwi, N.T. Alsohybe, and A.Y. Al-Hashida, "Intrusion detection model using machine learning algorithm on Big Data environment," *J. Big Data*, vol. 5, no. 1, pp. 1-12, Dec. 2018. doi: 10.1186/s40537-018-0145-4.

[40] I.A. Karatepe, and E. Zeydan, "Anomaly detection in cellular network data using big data analytics," in *Proc. European Wireless 2014; 20th European Wireless Conf.*, Barcelona, Spain: IEEE, May 2014, pp. 1-5. Available: https://ieeexplore.ieee.org/abstract/document/6843095.

[41] A.C. Onal, O.B. Sezer, M. Ozbayoglu, and E. Dogdu, "Weather data analysis and sensor fault detection using an extended IoT framework with semantics, big data, and machine learning," in *Proc. 2017 IEEE Int. Conf. Big Data (Big Data)*, Boston, USA: IEEE, pp. 2037-2046, Dec. 2017. doi: 10.1109/BigData.2017.8258150.

[42] P. Gottumukkala, and G.S. Rao, "Fault Detection in Mobile Communication Networks Using Data Mining techniques with big data analytics," *Int. J. Cybern. Inf.*, vol. 5, no. 4, pp. 81-89, Aug. 2016.

[43] P.A.D.S.N. Wijesekara, W.M.A.K. Sangeeth, H.S.C. Perera, and N.D. Jayasundere, "Underwater Acoustic Digital Communication Channel for an UROV," in *Proc. 5th Annu. Research Symp. (ARS2018)*, Hapugala, Sri Lanka: University of Ruhuna, Jan. 2018, p. E17.

[44] P. Baltiiski, I. Iliev, B. Kehaiov, V. Poulkov, and T. Cooklev, "Long-term spectrum monitoring with big data analysis and machine learning for cloud-based radio access networks," *Wireless Pers. Commun.*, vol. 87, pp. 815-835, Apr. 2016. doi: 10.1007/s11277-015-2631-8.

[45] F. Xu, Y. Lin, J. Huang, D. Wu, H. Shi, J. Song, and Y. Li, "Big data driven mobile traffic understanding and forecasting: A time series approach," *IEEE Trans. Services Comput.*, vol. 9, no. 5, pp. 796-805, Sep. 2016. doi: 10.1109/TSC.2016.2599878.

[46] P. Kiran, M.G. Jibukumar, and C.V. Premkumar, "Resource allocation optimization in LTE-A/5G networks using big data analytics," in *Proc. 2016 Int. Conf. Inform. Netw. (ICOIN),* Kota Kinabalu, Malaysia: IEEE, Jan. 2016, pp. 254-259. doi: 10.1109/ICOIN.2016.7427072.

[47] E. Baştuğ, M. Bennis, E. Zeydan, M.A. Kader, I.A. Karatepe, A.S. Er, and M. Debbah, "Big data meets telcos: A proactive caching perspective," *J. Commun. Netw.*, vol. 17, no. 6, pp. 549-557, Dec. 2015. doi: 10.1109/JCN.2015.000102.

[48] M.S. Hadi, A.Q. Lawey, T.E. El-Gorashi, and J.M. Elmirghani, "Big data analytics for wireless and wired network design: A survey," *Comput. Netw.*, vol. 132, pp. 180-199, Feb. 2018. doi: 10.1016/j.comnet.2018.01.016.

[49] P.A.D.S.N. Wijesekara, K.L.K. Sudheera, G.G.N. Sandamali, and P.H.J. Chong, "Data Gathering Optimization in Hybrid Software Defined Vehicular Networks," in *Proc. 20th Academic Sessions*, Matara, Sri Lanka: University of Ruhuna, Jun. 2023, p. 59.

[50] K. Lyko, M. Nitzschke, and A.C. Ngonga Ngomo, "Big data acquisition," *New horizons for a data-driven economy: A roadmap for usage and exploitation of big data in Europe*, Switzerland: Springer open, 2016, ch. 4, pp. 39-61. doi: 10.1007/978-3-319-21569-3_4.

[51] B. Leang, S. Ean, G.A. Ryu, and K.H. Yoo, "Improvement of Kafka streaming using partition and multi-threading in big data environment," *Sensors*, vol. 19, no. 1, Art. no. 134, Jan. 2019, doi: 10.3390/s19010134.

[52] Y. Xing, J. Han, K. Xue, J. Liu, M. Pan, and P. Hong, "MPTCP meets big data: Customizing transmission strategy for various data flows," *IEEE Netw.*, vol. 34, no. 4, pp.35-41, Jul. 2020. doi: 10.1109/MNET.011.1900438.

[53] T. Wang, H. Ke, X. Zheng, K. Wang, A.K. Sangaiah, and A. Liu, "Big data cleaning based on mobile edge computing in industrial sensor-cloud," *IEEE Trans. Ind. Informat.*, vol. 16, no. 2, pp. 1321-1329, Feb. 2020. doi: 10.1109/TII.2019.2938861.

[54] I. Taleb, R. Dssouli, and M.A. Serhani, "Big data pre-processing: A quality framework," in *Proc. 2015 IEEE Int. Congr. Big Data*, New York City, USA: IEEE, Jun. 2015, pp. 191-198. doi: 10.1109/BigDataCongress.2015.35.

[55] D. Wu, B. Yang, and R. Wang, "Scalable privacy-preserving big data aggregation mechanism," *Digit. Commun. Netw.*, vol. 2, no. 3, pp.122-129, Aug. 2016. doi: 10.1016/j.dcan.2016.07.001.

[56] Z. Guan, T. Ji, X. Qian, Y. Ma, and X. Hong, "A survey on big data pre-processing," in *Proc. 2017 5th Int. Conf. Appl. Comput. Inf. Technol./4th Int. Conf. Comput. Sci./Intell. Appl. Inf./2nd Int. Conf. Big Data, Cloud Comput., Data Sci. (ACIT-CSII-BCD)*, Hamamatsu, Japan: IEEE, Jul. 2017, pp. 241-247. doi: 10.1109/ACIT-CSII-BCD.2017.49.

[57] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The hadoop distributed file system," in *Proc. 2010 IEEE 26th Symp. Mass Storage Sys. Technol. (MSST)*, Incline Village, NV: IEEE, May 2010, pp. 1-10. doi: 10.1109/MSST.2010.5496972.

[58] A. Oussous, F.Z. Benjelloun, A.A. Lahcen, and S. Belfkih, "NoSQL databases for big data," *Int. J. Big Data Intell.*, vol. 4, no. 3, pp. 171-185, Jul. 2017. doi: 10.1504/IJBDI.2017.085537.

[59] D. García-Gil, S. Ramírez-Gallego, S. García, and F. Herrera, "A comparison on scalability for batch big data processing on Apache Spark and Apache Flink," *Big Data Anal.*, vol. 2, no. 1, pp.1-11, Dec. 2017. doi: 10.1186/s41044-016-0020-2.

[60] D. Cheng, P. Schretlen, N. Kronenfeld, N. Bozowsky, and W. Wright, "Tile based visual analytics for twitter big data exploratory analysis," in *Proc. 2013 IEEE Int. Conf. Big Data*, Silicon Valley, USA: IEEE, Oct. 2013, pp. 2-4. doi: 10.1109/BigData.2013.6691787.

[61] T. Su, H. Xu, and X. Zhou, "Particle swarm optimization-based association rule mining in big data environment," *IEEE Access*, vol. 7, pp. 161008-161016, Nov. 2019. doi: 10.1109/ACCESS.2019.2951195.

[62] S. Kumar, and M. Singh, "A novel clustering technique for efficient clustering of big data in Hadoop Ecosystem," *Big Data Mining Anal.*, vol. 2, no. 4, pp. 240-247, Aug. 2019. doi: 10.26599/BDMA.2018.9020037.

[63] A.V. Ramya, and E. Sivasankar, "Distributed pattern matching and document analysis in big data using Hadoop MapReduce model," in *Proc. 2014 Int. Conf. Parallel Distrib. Grid Comput.*, Solan, India: IEEE, Dec. 2014, pp. 312-317. doi: 10.1109/PDGC.2014.7030762.

[64] A. Malviya, A. Udhani, and S. Soni, "R-tool: Data analytic framework for big data," in *Proc. 2016 Symp. Colossal Data Anal. Netw. (CDAN)*, Indore, India, Mar. 2016, pp. 1-5. doi: 10.1109/CDAN.2016.7570960.

[65] H.M.D.P.M. Herath, W.A.S.A. Weraniyagoda, R.T.M. Rajapaksha, P.A.D.S.N. Wijesekara, K.L.K. Sudheera, and P.H.J. Chong, "Automatic Assessment of Aphasic Speech Sensed by Audio Sensors for Classification into Aphasia Severity Levels to Recommend Speech Therapies," *Sensors*, vol. 22, no. 18, Art. no. 6966, Sep. 2022, doi: 10.3390/s22186966.

[66] D. Arora, K.F. Li, and A. Loffler, "Big data analytics for classification of network enabled devices," in *Proc. 2016 30th Int. Conf. Adv. Inf. Netw. Appl. Workshops (WAINA)*, Crans-Montana, Switzerland: IEEE, Mar. 2016, pp. 708-713. doi: 10.1109/WAINA.2016.131.

[67] W. Zhao, and N. Herencsar, "Logistic regression analysis of targeted poverty alleviation with big data in mobile network," *Mobile Netw. Appl.*, vol. 27, no. 6, pp. 2553-2564, Dec. 2022. doi: 10.1007/s11036-022-02068-5.

[68] A. Cuzzocrea, C.K. Leung, M. Hajian, and M.D. Jackson, "Effectively and Efficiently Supporting Predictive Big Data Analytics over Open Big Data in the Transportation Sector: A Bayesian Network Framework," in *Proc. 2022 IEEE DASC/PiCom/CBDCom/CyberSciTech*, Calabria, Italy: IEEE, Sep. 2022, pp. 1-8. doi: 10.1109/DASC/PiCom/CBDCom/Cy55231.2022.9927788.

[69] A. Mirzaei Somarin, M. Barari, and H. Zarrabi, "Big data based self-optimization networking in next generation mobile networks," *Wireless Pers. Commun.*, vol. 101, pp. 1499-1518, Aug. 2018. doi: 10.1007/s11277-018-5774-6.

[70] G. Pîrlea, and I. Sergey, "Mechanising blockchain consensus," in *Proc. 7th ACM SIGPLAN Int. Conf. Certified Programs Proofs*, Los Angeles, USA: ACM, Jan. 2018, pp. 78-90. doi: 10.1145/3167086.

[71] P.A.D.S.N. Wijesekara, "Ethical Knowledge Sharing Leveraging Blockchain: An Overview," *Sci. Eng. Technol.*, vol. 4, no. 1, pp. 112-136, Apr. 2024. doi: 10.54327/set2024/v4.i1.126.

[72] P.A.D.S.N. Wijesekara, "A Literature Review on Access Control in Networking Employing Blockchain," *Indonesian J. Comput. Sci.*, vol. 13, no. 1, pp. 734-768, Feb. 2024. doi: 10.33022/ijcs.v13i1.3764.

[73] P.A.D.S.N. Wijesekara, "Deep 3D Dynamic Object Detection towards Successful and Safe Navigation for Full Autonomous Driving," *Open Transp. J.*, vol. 16, no. 1, Art. no. e187444782208191, Oct. 2022, doi: 10.2174/18744478-v16-e2208191.

[74] V.S. Marichamy, and V. Natarajan, "Blockchain based Securing Medical Records in Big Data Analytics," *Data Knowl. Eng.*, vol. 144, Art. no.102122, Mar. 2023, doi: 10.1016/j.datak.2022.102122.

[75] U. Demirbaga, and G.S. Aujla, "MapChain: A blockchain-based verifiable healthcare service management in IoT-based big data ecosystem," *IEEE Trans. Netw. Service Manag.*, vol. 19, no. 4, pp. 3896-3907, Dec. 2022. doi: 10.1109/TNSM.2022.3204851.

[76] J. Yang, J. Wen, B. Jiang, and H. Wang, "Blockchain-based sharing and tamper-proof framework of big data networking," *IEEE Netw.*, vol. 34, no. 4, pp. 62-67, Jul. 2020. doi: 10.1109/MNET.011.1900374.

[77] I.L.H. Alsammak, M.F. Alomari, I.S. Nasir, and W.H. Itwee, "A model for blockchain-based privacy-preserving for big data users on the internet of thing," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 26, no. 2, pp. 974-988, May 2022. doi: 10.11591/ijeecs.v26.i2.pp974-988.

[78] M.M. Hassan, A. Gumaei, A. Alsanad, M. Alrubaian, and G. Fortino, "A hybrid deep learning model for efficient intrusion detection in big data environment," *Inf. Sci.*, vol. 513, pp. 386-396, Mar. 2020. doi: 10.1016/j.ins.2019.10.069.

[79] M. Mahdavisharif, S. Jamali, and R. Fotohi, "Big data-aware intrusion detection system in communication networks: a deep learning approach," *J. Grid Comput.*, vol. 19, pp.1-28, Dec. 2021. doi: 10.1007/s10723-021-09581-z.

[80] F. Louati, F.B. Ktata, and I.A.B. Amor, "A Distributed Intelligent Intrusion Detection System based on Parallel Machine Learning and Big Data Analysis," in *Proc. 11th Int. Conf. Sensor Netw. (SENSORNETS)*, Vienna, Austria: SCITEPRESS, Feb. 2022, pp. 152-157. doi: 10.5220/0010886300003118.

[81] E. Viegas, A.O. Santin, and V. Abreu Jr, "Machine learning intrusion detection in big data era: A multi-objective approach for longer model lifespans," *IEEE Tran. Netw. Sci. Eng.*, vol. 8, no. 1, pp. 366-376, Nov. 2020. doi: 10.1109/TNSE.2020.3038618.

[82] B. Hussain, Q. Du, and P. Ren, "Semi-supervised learning based big data-driven anomaly detection in mobile wireless networks," *China Commun.*, vol. 15, no. 4, pp. 41-57, Apr. 2018. doi: 10.1109/CC.2018.8357700.

[83] P. Casas, F. Soro, J. Vanerio, G. Settanni, and A. D'Alconzo, "Network security and anomaly detection with Big-DAMA, a big data analytics framework," in *Proc. 2017 IEEE 6th Int. Conf. Cloud Netw. (CloudNet)*, Prague, Czech Republic: IEEE, Sep. 2017, pp. 1-7. doi: 10.1109/CloudNet.2017.8071525.

[84] D.S. Terzi, R. Terzi, and S. Sagiroglu, "Big data analytics for network anomaly detection from netflow data," in *Proc. 2017 Int. Conf. Comput. Sci. Eng. (UBMK)*, Antalya, Turkey: IEEE, Oct. 2017, pp. 592-597. doi: 10.1109/UBMK.2017.8093473.

[85] M.S. Parwez, D.B. Rawat, and M. Garuba, "Big data analytics for user-activity analysis and user-anomaly detection in mobile wireless network," *IEEE Trans. Ind. Informat.*, vol. 13, no. 4, pp. 2058-2065, Aug. 2017. doi: 10.1109/TII.2017.2650206.

[86] M. Syafrudin, G. Alfian, N.L. Fitriyani, and J. Rhee, "Performance analysis of IoT-based sensor, big data processing, and machine learning model for real-time monitoring system in automotive manufacturing," *Sensors*, vol. 18, no. 9, Art. no. 2946, Sep. 2018, doi: 10.3390/s18092946.

[87] F. Benayas, A. Carrera, and C.A. Iglesias, "Towards an autonomic Bayesian fault diagnosis service for SDN environments based on a big data infrastructure," in *Proc. 2018 5th Int. Conf. Softw. Defined Syst. (SDS)*, Barcelona, Spain: IEEE, Apr. 2018, pp. 7-13. doi: 10.1109/SDS.2018.8370415.

[88] I. Kotenko, I. Saenko, and A. Branitskiy, "Framework for mobile Internet of Things security monitoring based on big data processing and machine learning," *IEEE Access*, vol. 6, pp. 72714-72723, Nov. 2018. doi: 10.1109/ACCESS.2018.2881998.

[89] L. Maccari, and A. Passerini, "A Big Data and machine learning approach for network monitoring and security," *Security and Privacy*, vol. 2, no. 1, Art. no. e53, Jan. 2019, doi: 10.1002/spy2.53.

[90] I. Machorro-Cano, G. Alor-Hernández, M.A. Paredes-Valverde, L. Rodríguez-Mazahua, J.L. Sánchez-Cervantes, and J.O. Olmedo-Aguirre, "HEMS-IoT: A big data and machine learning-based smart home system for energy saving," *Energies*, vol. 13, no. 5, Art. no. 1097, Mar. 2020,. doi: 10.3390/en13051097.

[91] Z. Chang, L. Lei, Z. Zhou, S. Mao, and T. Ristaniemi, "Learn to cache: Machine learning for network edge caching in the big data era," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 28-35, Jun. 2018. doi: 10.1109/MWC.2018.1700317.

[92] E. Zeydan, E. Bastug, M. Bennis, M.A. Kader, I.A. Karatepe, A.S. Er, and M. Debbah, "Big data caching for networking: Moving from cloud to edge," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 36-42, Sep. 2016. doi: 10.1109/MCOM.2016.7565185.

[93] Y. He, F.R. Yu, N. Zhao, V.C. Leung, and H. Yin, "Software-defined networks with mobile edge computing and caching for smart cities: A big data deep reinforcement learning approach," *IEEE Commun. Mag.*, vol. 55, no. 12, pp. 31-37, Dec. 2017. doi: 10.1109/MCOM.2017.1700246.

[94] X. Huang, K. Xie, S. Leng, T. Yuan, and M. Ma, "Improving Quality of Experience in multimedia Internet of Things leveraging machine

learning on big data," *Future Gener. Comput. Syst.*, vol. 86, pp. 1413-1423, Sep. 2018. doi: 10.1016/j.future.2018.02.046.

[95] Y. He, C. Liang, F.R. Yu, N. Zhao, and H. Yin, "Optimization of cache-enabled opportunistic interference alignment wireless networks: A big data deep reinforcement learning approach," in *Proc. 2017 IEEE Int. Conf. Commun. (ICC)*, Paris, France: IEEE, May 2017, pp. 1-6. doi: 10.1109/ICC.2017.7996332.

[96] P. Chiu, J. Reunanen, R. Luostari, and H. Holma, "Big data analytics for 4.9 G and 5G mobile network optimization," in *Proc. 2017 IEEE 85th Veh. Technol. Conf. (VTC Spring)*, Sydney, Australia: IEEE, Jun. 2017, pp. 1-4. doi: 10.1109/VTCSpring.2017.8108601.

[97] F. Restuccia, and T. Melodia, "Big data goes small: Real-time spectrum-driven embedded wireless networking through deep learning in the RF loop," in *Proc. IEEE Conf. Comput. Commun. (IEEE INFOCOM 2019)*, Paris, France: IEEE, Apr. 2019, pp. 2152-2160. doi: 10.1109/INFOCOM.2019.8737459

[98] P. Rohini, S. Tripathi, C.M. Preeti, A. Renuka, J.L.A. Gonzales, and D. Gangodkar, "A study on the adoption of Wireless Communication in Big Data Analytics Using Neural Networks and Deep Learning," in *Proc. 2022 2nd Int. Conf. Adv. Comput. Innovative Technol. Eng. (ICACITE)*, Greater Noida, India: IEEE, Apr. 2022, pp. 1071-1076. doi: 10.1109/ICACITE53722.2022.9823439.

[99] D. Nallaperuma, R. Nawaratne, T. Bandaragoda, A. Adikari, S. Nguyen, T. Kempitiya, D. De Silva, D. Alahakoon, and D. Pothuhera, "Online incremental machine learning platform for big data-driven smart traffic management," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 12, pp. 4679-4690, Dec. 2019. doi: 10.1109/TITS.2019.2924883.

[100] S. Khan, S. Nazir, I. García-Magariño, and A. Hussain, "Deep learning-based urban big data fusion in smart cities: Towards traffic monitoring and flow-preserving fusion," *Comput. Electr. Eng.*, vol. 89, Art. no.106906, Jan. 2021, doi: 10.1016/j.compeleceng.2020.106906.

[101] X. Li, H. Liu, W. Wang, Y. Zheng, H. Lv, and Z. Lv, "Big data analysis of the internet of things in the digital twins of smart city based on deep learning," *Future Gener. Comput. Syst.*, vol. 128, pp. 167-177, Mar. 2022. doi: 10.1016/j.future.2021.10.006.

[102] D. Unal, M. Hammoudeh, M.A. Khan, A. Abuarqoub, G. Epiphaniou, and R. Hamila, "Integration of federated machine learning and blockchain for the provision of secure big data analytics for Internet of Things," *Comput. Security*, vol. 109, Art. no. 102393, Oct. 2021, doi: 10.1016/j.cose.2021.102393.

[103] A. Mitra, B. Bera, A.K. Das, S.S. Jamal, and I. You, "Impact on blockchain-based AI/ML-enabled big data analytics for Cognitive Internet of Things environment," *Comput. Commun.*, vol. 197, pp. 173-185, Jan. 2023. doi: 10.1016/j.comcom.2022.10.010.

[104] Y. Wan, Y. Qu, L. Gao, and Y. Xiang, "Privacy-preserving blockchain-enabled federated learning for B5G-Driven edge computing," *Comput. Netw.*, vol. 204, Art. no.108671, Feb. 2022, doi: 10.1016/j.comnet.2021.108671.

[105] S. Rathore, Y. Pan, and J.H. Park, "BlockDeepNet: A blockchain-based secure deep learning for IoT network," *Sustainability*, vol. 11, no. 14, Art. no. 3974, Jul. 2019, doi: 10.3390/su11143974.

[106] S.K. Singh, S. Rathore, and J.H. Park, "Blockiotintelligence: A blockchain-enabled intelligent IoT architecture with artificial intelligence," *Future Gener. Comput. Syst.*, vol. 110, pp. 721-743, Sep. 2020. doi: 10.1016/j.future.2019.09.002.

[107] P.A.D.S.N. Wijesekara, "A Review on Deploying Blockchain Technology for Network Mobility Management," *Int. Trans. Electr. Eng. Comput. Sci.*, vol. 3, no. 1, pp. 1-33, Mar. 2024. doi: 10.62760/iteecs.3.1.2024.83.

[108] S. Rathore, and J.H. Park, "A blockchain-based deep learning approach for cyber security in next generation industrial cyber-physical systems," *IEEE Trans. Ind. Informat.*, vol. 17, no. 8, pp. 5522-5532, Aug. 2021. doi: 10.1109/TII.2020.3040968.

[109] C. Seneviratne, P.A.D.S.N. Wijesekara, and H. Leung, "Performance analysis of distributed estimation for data fusion using a statistical approach in smart grid noisy wireless sensor networks," *Sensors*, vol. 20, no. 2, Art. no. 567, Jan. 2020, doi: 10.3390/s20020567.

[110] S. Ahn, S.V. Couture, A. Cuzzocrea, K. Dam, G.M. Grasso, C.K. Leung, K.L. McCormick, and B.H. Wodi, "A fuzzy logic based machine learning tool for supporting big data business analytics in complex artificial intelligence environments," in *Proc. 2019 IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE),* New Orleans, USA: IEEE, Jun. 2019, pp. 1-6. doi: 10.1109/FUZZ-IEEE.2019.8858791.

[111] P.A.D.S.N. Wijesekara, and S. Gunawardena, "A Machine Learning-Aided Network Contention-Aware Link Lifetime- and Delay-Based Hybrid Routing Framework for Software-Defined Vehicular Networks," *Telecom*, vol. 4, no. 3, pp. 393-458, Jul. 2023. doi: 10.3390/telecom4030023.

[112] A. Nazir, "Seamless automation and integration of machine learning capabilities for big data analytics," *Int. J. Distrib. Parallel Syst*, vol. *8*, no. 3, pp. 1-18, May 2017. doi: 10.5121/ijdps.2017.8301.

[113] P.A.D.S.N. Wijesekara, K.L.K. Sudheera, G.G.N. Sandamali, and P.H.J. Chong, "An Optimization Framework for Data Collection in Software Defined Vehicular Networks," *Sensors*, vol. 23, no. 3, Art. no. 1600, Feb. 2023, doi: 10.3390/s23031600.

[114] S. Ma, S. Wang, and W.T. Tsai, "Delay analysis of consensus communication for blockchain-based applications using network calculus," *IEEE Wireless Commun. Lett.*, vol. *11*, no. 9, pp. 1825-1829, Jun. 2022. doi: 10.1109/LWC.2022.3183197.

[115] P.A.D.S.N. Wijesekara, and S. Gunawardena, "A Comprehensive Survey on Knowledge-Defined Networking," *Telecom*, vol. 4, no. 3, pp. 477-596, Aug. 2023. doi: 10.3390/telecom4030025.

[116] A.M. Rahmani, E. Azhir, S. Ali, M. Mohammadi, O.H. Ahmed, M.Y. Ghafour, S.H. Ahmed, and M. Hosseinzadeh, "Artificial intelligence approaches and mechanisms for big data analytics: a systematic study," *PeerJ Comput. Sci.*, vol. *7*, Art. no. e488, Apr. 2021, doi: 10.7717/peerj-cs.488.

[117] P.A.D.S.N. Wijesekara, K.L.K. Sudheera, G.G.N. Sandamali, and P.H.J. Chong, "Machine Learning Based Link Stability Prediction for Routing in Software Defined Vehicular Networks," in *Proc. 20th Academic Sessions*, Matara, Sri Lanka: University of Ruhuna, Jun. 2023, Art no. 60. Available: http://ir.lib.ruh.ac.lk/handle/iruor/13317.

[118] K. Karisma, and P. Moslemzadeh Tehrani, "Blockchain: Legal and Regulatory Issues," in *Sustainable Oil and Gas Using Blockchain,* Cham: Springer International Publishing, vol. 98, 2023, pp. 75-118. doi: 10.1007/978-3-031-30697-6_4.

[119] R. Kashyap, "Big Data Analytics challenges and solutions," in *Big Data Anal. Intell. Healthcare Manag.,* Academic press: Elsevier, 2019, ch. 2, pp. 19-41. doi: 10.1016/B978-0-12-818146-1.00002-7.

[120] 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, Jan. 2019. doi: 10.1109/ACCESS.2018.2890507.