Multimedia Tools and Applications

Big Image Data 15Vs Model for Intelligent Data Ocean of Multimedia Things

--Manuscript Draft--

Manuscript Number:	MTAP-D-23-07072
Full Title:	Big Image Data 15Vs Model for Intelligent Data Ocean of Multimedia Things
Article Type:	Track 1: General Multimedia Topics
Keywords:	Big Image Data 15Vs Model; Dark data; Real-time Cyber-physical System; Bot-oriented Processing; Internet of Multimedia Things; Intelligent Data Ocean; Digital Twin
Corresponding Author:	Aminollah Mahabadi Shahed University Tehran, IRAN, ISLAMIC REPUBLIC OF
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Shahed University
Corresponding Author's Secondary Institution:	
First Author:	Aminollah Mahabadi
First Author Secondary Information:	
Order of Authors:	Aminollah Mahabadi
	Faezeh Abdolazimi
Order of Authors Secondary Information:	
Funding Information:	
Abstract:	The classification of large image data using traditional methods has been based on changes in image gray features, extraction of edge and contour feature information, or conversion between image coordinate sets. However, due to the growth of new image size or big image data (BID) in real-time multimedia communication and future online big data (BD) applications, these methods have become increasingly complex and resulting in poor real-time performance. The complex methods suffer from complex algorithms, massive data communication, slow processing speed, unintelligent predictive modeling, weak data classification, limited accuracy, uncleaned usage data, combined destructive artificial and natural noise, exchanging data value over time, and exhausting updating data for accurate operation of media storage. In the era of the Internet of Multimedia Things (IoMT), most modern devices can accurately capture vast amounts of observational data in the form of valuable images, text, and acoustic recordings. This reliable data must be kept valid for local data processing at different times and for globally extracting knowledge from information at various data levels to address forthcoming challenges related to big image data. The challenges of secure communication require accurate, reliable, fast, and effective information gathering for local decision-making and global knowledge mining of BID to support cyber-physical systems for speedy data virtualization of smart city devices. Complex and uncontrollable problems persist on the outskirts of grid, fog, and cloud networks regarding data cleansing and privacy protection for intelligent data ocean management. This study developed a new 15Vs model empirically to examine distributed big image data processing based on a modern BID connectivity approach. The model accurately extracts textured features from visible data images and overcomes a unique set of key challenges. The model introduces a strategic analyzer, uses intelligent local agents, and recruits clever global bots to provide a suitable platform for generous support of bot-oriented BD processing. This instantly forms a hierarchical data level for better data acquisition and cleansing, safe privacy protection, suitable information diffusion, and speedy knowledge extraction. Our study presents a novel definition of BID to address the ultimate challenges of data management in a high-risk environment for further big image data modeling research needs.

Big Image Data 15Vs Model for Intelligent Data Ocean of Multimedia Things

Aminollah Mahabadi[a,b,*], Faezeh Abdolazimi[b]

[a]School of Computer Science, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran
[b]Computer Engineering Department, Shahed University, Tehran, Iran

Abstract

The classification of large image data using traditional methods has been based on changes in image gray features, extraction of edge and contour feature information, or conversion between image coordinate sets. However, due to the growth of new image size or *big image data* (BID) in real-time multimedia communication and future online *big data* (BD) applications, these methods have become increasingly complex and resulting in poor real-time performance. The complex methods suffer from complex *algorithms*, massive data *communication*, slow *processing* speed, unintelligent *predictive* modeling, weak data *classification*, limited *accuracy*, uncleaned usage *data*, combined destructive *artificial* and *natural* noise, exchanging data value over *time*, and exhausting *updating* data for accurate operation of media storage. In the era of the Internet of Multimedia Things (IoMT), most modern devices can accurately capture vast amounts of observational data in the form of valuable *images*, *text*, and *acoustic* recordings. This reliable data must be kept valid for local data processing at different *times* and for globally extracting knowledge from information at various data *levels* to address forthcoming challenges related to big image data. The challenges of secure communication require *accurate*, *reliable*, *fast*, and *effective* information gathering for local decision-making and global knowledge mining of BID to support cyber-physical systems for speedy data *virtualization* of smart city devices. Complex and uncontrollable problems persist on the outskirts of *grid*, *fog*, and *cloud* networks regarding data *cleansing* and *privacy protection* for intelligent data ocean management. This study developed a new *15Vs model* empirically to examine distributed big image data processing based on a modern BID connectivity approach. The model accurately extracts textured features from visible data images and overcomes a unique set of key challenges. The model introduces a *strategic* analyzer, uses intelligent local *agents*, and recruits clever global *bots* to provide a suitable platform for generous support of bot-oriented BD processing. This instantly forms a hierarchical data level for better data acquisition and *cleansing*, safe privacy *protection*, suitable *information* diffusion, and speedy *knowledge* extraction. Our study presents a *novel definition* of BID to address the ultimate challenges of data management in a *high-risk* environment for further big image data modeling research needs.

Keywords: Big Image Data 15Vs Model, Dark data, Real-time Cyber-physical System, Bot-oriented Processing, Internet of Multimedia Things, Intelligent Data Ocean, Digital Twin.

1. Introduction

Advanced information technology has facilitated the efficient handling of a vast volume of historical data at high velocity, typically coupled with the intellectual ability to continuously gather and carefully store the data stream, which will predictably lead to increasingly complex challenges across all foreseeable future domains [1]. This broad area encompasses topics for data such as *modeling*, *capturing*, *sensing*, *storing*, *validating*, *cleaning*, *polishing*, *phishing*, *computing*, *processing*, *analysis*, *representation*, *visualization*, *communication*, *diffusion*, *retrieval*, *protection*, *security*, and *decision-making* to become more competitive and grow without limits, such as an *intelligent data ocean* (IDO). Undoubtedly, owing to the fundamental quality of big data, which has recently been generated and subjected to cleaning and cleansing processing, there is a need for contemporary computing trends to establish practical terminology and definitions such as notable *volume*, local *variety*, effective *velocity*, and intrinsic *value*. These steps for a big data *strategy* include *redefining*, *modeling*, *classification*, *clearing*, *duplicating*, *debugging*, *missing data*, *outlier filtering*, *validation*, and *protection* of historical data for exact *fit*, *denoising*, *accessibility*, *correctness*, *consistency*, *usability*, and *privacy protection* for *smart data* and innovative *data intelligence* in a *multi-level* model [2].

*Corresponding Author. Tel.: +98(21)51212078; fax: +98(21)51212020.
Email address: am@ipm.ir; mahabadi@shahed.ac.ir (Aminollah Mahabadi)

Table 1: List of used acronyms, symbols, metrics, models and their descriptions, definitions, or dimensions [authors].

Acronym	Description	Acronym	Description
2D	Two-dimensional	HKGB	Health Knowledge Graph Builder
3D	Three-dimensional	ICPS	Intelligent Cyber-Physical System
AI	Artificial Intelligence	ICT	Information and Communications Technology
BD	Big Data	IDA	Intelligent Data Analysis
BDA	Big Data Analytics	IDC	Intelligent Data Core
BDC	Big Data Cleaning	IDM	Information Diffusion Model
BDCS	Big Data Cleansing	IDO	Intelligent Data Ocean
BDO	Big Data Ocean	IoT	Internet of Things
BAD	Big Acoustic Data	IIoT	Internet of Industrial Things
BI	Business Intelligence	IoE	Internet of Everything
BID	Big Image Data	IoMT	Internet of Multimedia Things
BIDM	Big Image Data Model	IoMET	Internet of Medical Things
BIDL	Big Image Data Lake	IoSCT	Internet of Smart Cities Things
BIDS	Big Image Data Strategy	IRC	Internet Relay Chat
BTD	Big Text Data	IS	Information System
BoP	Bot-oriented Processing	IT	Information Technology
BVs	Blood Vessels	ITS	Intelligent Transportation Systems
CIA	Confidentiality, Integrity and Availability	KBIP	Knowledge-Based Image Processing
CNC	Computer Numerical Control	KG	Knowledge Graph
CNN	Convolutional Neural Networks	KT	Knowledge Transfer
COC	Cells-on-chips	MARV	Marburg Virus
CPS	Cyber-Physical System	MRI	Magnetic Resonance Imaging
CSP	Communication Service Providers	ML	Machine Learning
CT	Computerized Tomography	MWT	Metaworth- or Metaverse-based Ecosystem
DBMS	Database Management System	OCR	Optical Character Recognition
DD	Dark Data	OOC	Organs-on-chips
DDoS	Distributed Denial-of-Service	PAM	Proximity Attention Model
DMS	Decision Making System	PCs	Personal Computers
DRP	Digital Rock Physics	PET	Positron Emission Tomography
DSL	Domain Specific Language	RFID	Radio-frequency Identification
DT	Digital Twin	RTEC	Run-Time Event Calculus
DTA	Digital Twin Aggregate	RES	Real-time Embedded Systems
DTC	Digital Twin City	SC	Smart Cities
DTI	Digital Twin Instance	SIDM	Smart Information Diffusion Model
DTP	Digital Twin Prototype	SAR	Source-to-Artifacts Ratio
DV	Data Virtualization	SDR	Source-to-Distortion Ratio
EAM	Enterprise Architecture Model	SIR	Source-to-Interference Ratio
EMR	Elastic MapReduce — Amazon EMR	SLA	Service level Agrement
ESM	Enterprise Strategic Management	SMS	Short Message Service
EBO	Ebola Virus	SPECT	Single-Photon Emission Computerized Tomography
FDHM	Full-Duration-at-Half-Maximum	SQL	Structured Query Language
GIS	Geographic Information Systems	TD	Tracer Dilution
GPS	Global Positioning System	VAG	Vitality Adjacent Graph
GPU	Graphics Processor Unit	VAR	Video Assistant Referee
HTML	Hyper Text Markup Language	VMI	Vendor Market Intelligence
HTTP	Hypertext Transfer Protocol	WSI	Whole Slide Image
HTTPS	Hypertext Transfer Protocol Secure	XML	eXtensible Markup Language
Symbol	**Description**	**Symbol**	**Description**
Bot	A computer program	RoBot	An internet Bot
Bot-herder	A single attack party	Zombie	A bad Bot
BotMaster	A person who controls BotNets	5G	5^{th} generation mobile network
BotNet	A group of Internet-connected computer devices	x-ray	A medical imaging type
Metric	**Definition**	**Metric**	**Definition**
GB	Gigabytes or 10^9 bytes	fps	frame per second
TB	Terabytes or 10^{12} bytes	nm	$nano$meter
ZB	Zettabytes or 10^{21} bytes	ms	$mili$second
YB	Yottabytes or 10^{24} bytes	ns	$nano$second
Model	**Dimensions**	**Model**	**Dimensions**
V3C	Variety, volume, value, and complexity.	15Vs	Variety, volume, value, velocity, veracity, volatility, variability, vindication, validity, valence, vulnerability, visualization, virtualization, vicinity, and vitality.
4Vs	Variety, volume, velocity, and veracity.		
5Vs	Variety, volume, value, velocity, and veracity.		
6Vs	Variety, volume, value, velocity, veracity, and variability.		

Figure 1: An enterprise architecture model of big data analytic for strategic management [authors].

Identifying and repairing dirty data is one of the key challenges in data analytics for reliable decision-making, which requires an appropriate big data model for various forms of *text*, *images*, and *acoustics* owing to the functional or structural dependencies at different levels of *multimedia* technology [4]. Big image data *streams* have become ubiquitous because a considerable number of *online* multimedia applications naturally generate massive amounts of various types of data at an incredible velocity in 2D and 3D forms. Multimedia applications combine different data types in *text*, *speech*, *sound*, *music*, *image*, and *video* formats [5]. This work has to be directly managed by new devices in big data streams because of the built-in dynamic characteristics of different types of data, with an incredible *speed* of presented mining tools, applied technologies, designed methods, heterogeneous hardware, and hybrid techniques from starting data construction to ending useful information production for reasonable speed of knowledge extraction in decision-making at various data stream network levels [6, 7].

Note that *dark data* (DD) represents a significant portion of the enormous and complex big data world that has been collected but not utilized or turned into value [8]. Our focus in this study is on "*big data*," and we possess little work to do with the concept of "*dark data*," collected from various computer systems and currently unsuitable for decision-making. Also, a list of all acronyms, symbols, metrics, and models used in this study and their descriptions, definitions and dimensions is presented in Table 1. Finally, in this section, we provide a description of big data in Section 1.1, discuss different data levels and data intelligence in Section 1.2, present the big image data 15v's conceptual model in Section 1.3, propose motivation and innovation in Section 1.4, and outline the paper organization in Section 1.5.

1.1. Big Data Oceans

In recent times, several interactive concepts have been carefully developed, some of which have the ability to adapt to and interact appropriately with their surrounding environment and its various elements, such as wind, heat, sound, light, and human presence [9]. Interactive kinetic installations have the potential to revolutionize the architecture by providing *adaptable elements* that improve the user experience. These dynamic features can adapt to changing conditions, providing privacy, sun protection, and increased comfort. They create an engaging dialogue between the built environment and occupants through responsive motion, allowing customization and optimization of the space. Moreover, they consistently acquires knowledge and improves their cognitive abilities by optimizing performance and energy efficiency through the appropriate utilization of real-time data and user behavior, employing artificial intelligence (AI) and machine learning (ML) [10]. Therefore, this real-time embedded system (RES) technology enhances the user experience and promotes sustainability and energy efficiency in smart buildings in the future [11].

Most big data experts agree that the amount of data generated by an intelligent data core (IDC) will grow exponentially in the future. Useable data processing approaches set the direct challenge to huge search spaces for accurate, correct, fast, and impressive mining to *privacy* protection for naturally supporting big data, and instantly become modern complex and difficult to overcome the following problems. Fig.1) [12] shows the construction of an enterprise architecture model (EAM) for enterprise strategic management (ESM). *Application* software is a type of computer program that implements a specific personal, educational, or business routine to help users perform various tasks related to productivity, creativity, or communication. *Technology* refers to the physical and cyber elements that comprise computers, devices, tools, and everything else involved that is physically or cyber-tangible. *Data science* is a subset of AI that relates to the application of data models, scientific analysis, and statistical techniques to extract insights and significance from data. *Knowledge* transfer (KT) refers to a process in which employers share their skills, information, experience, or ideas with other departments or individuals in a business.

It is possible to reduce computational *complexity* at a reasonable cost using software techniques in multicore hardware, grid structures, fog environments, and cloud computing to precisely control large search spaces, correct

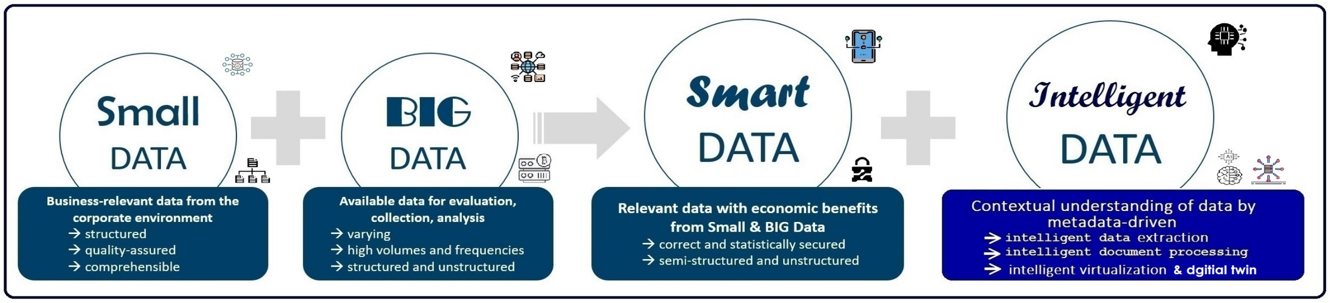

Figure 2: Various data type for intelligent information extraction and automatic document processing [authors].

resource consumption, network load balancing, timely real-time reactions, and direct generation of consequence transactions. The possible speed and correct computing of big data, careful control of data propagation, and proper monitoring of information diffusion in various *large complex* networks for modern *text*, visible *image*, and *acoustic* data types naturally increase the time complexity and memory usage for big *text* data [74], big *image* data [6], big *acoustic* data [65], and the possible combination of exclusive video and audio for multimedia applications in uncertain and *high-risk* environments [97], different *topology* streaming [15], and various *channel* utilization [16] for automatic *decision-making*. This potential problem routinely requires modern definitions and economic modeling for the subsequent definition of big image data generation in a reliable form suitable for various online interactions of intelligent *multimedia* applications, from camera *imaging* to knowledge extraction of sequential images [17, 5].

Future system functionality can be influenced by the *location*, environmental *conditions*, and other *proximity*systems. Users expect a certain behavior of applications, and they must be *predictable* in real-time monitoring applications for high *computation*, streaming *communication*, and multimedia usage in *virtual reality* visualization of different objects such as video assistant referees (VAR) in football games [20, 21]. An important rule in VAR image processing and vision analysis is 2D object *detection* and 3D *construction* from 2D imaging using different *engines* to maintain the relationships and track them for visualizing real *physical* attributes, detecting a *collision*, rendering *graphics*, clarifying *texts*, playing real *sounds*, and preparing *virtual reality* animations by using *artificial intelligence*, *machine learning*, and *knowledge graphs* for real-time decision-making [22, 23, 24].

1.2. Data Intelligence

Data in our world requires a revolution from small-scale data to large-scale *data intelligence* (See Fig.2). First, classical data escaped from the form of *local* data in classical structured data for small businesses to *global* big data in structured and unstructured data for the information provisioning of large business. Secondly, the process of collecting limited local data by private entities is transitional towards the use of correct and secure *smart* big data in the appropriate format of semi-structured and unstructured data, which can be utilized for real-time decision-making within large business systems. Finally, obtaining relevant, smart, correct, and secured data will require a critical transition to *data intelligence*. This includes classically structured, semi-structured, and unstructured data for finding past events and understanding why they occurred, utilizing AI and ML for better comprehension. Additionally, we must use a knowledge graph (KG) to extract all entities, real concepts, their relationships, maximum attributes, known values, and other information from the data structure. These valuable tools will enable us to summarize and analyze primary information for making real-time decisions on a large scale in IoMT. In contrast to big data, smart data is characterized by a greater degree of contextualization and calibration, which allows for the identification of specific needs and facilitates a higher level of customization. This type, in turn, enables large companies to obtain precise information about their local operations and propose globally customized and ad hoc solutions. Data intelligence represents a digital transformation in the manner in which large businesses know how to analyze and consume various large volumes of big image data documents.

Intelligent *virtualization* represents a real-time process that focuses on creating virtual big data to eliminate physical environmental problems. To automate this process, one solution is to combine optical character recognition (OCR) technology with *text* classification for name entity recognition. The technology employed can scan and extract text from image documents for automatic identification and extract the key data needed from a document, such as a healthcare document, to identify prescriptions, patient transfer forms, physician notes, etc. Data intelligence includes specific tools and methods that enterprise-scale organizations use to better understand the information they collect, store, and use to improve their products and services. The KG has been developed with the aim of enhancing the cognitive abilities and operational efficiency of engines. This tool comprises conceptual entities, their corresponding attributes, and tangible interconnections [51]. This method for efficient KG construction includes knowledge extraction, knowledge fusion, ontology learning, entity learning, knowledge representation, knowledge verification, knowledge reasoning, and knowledge storage.

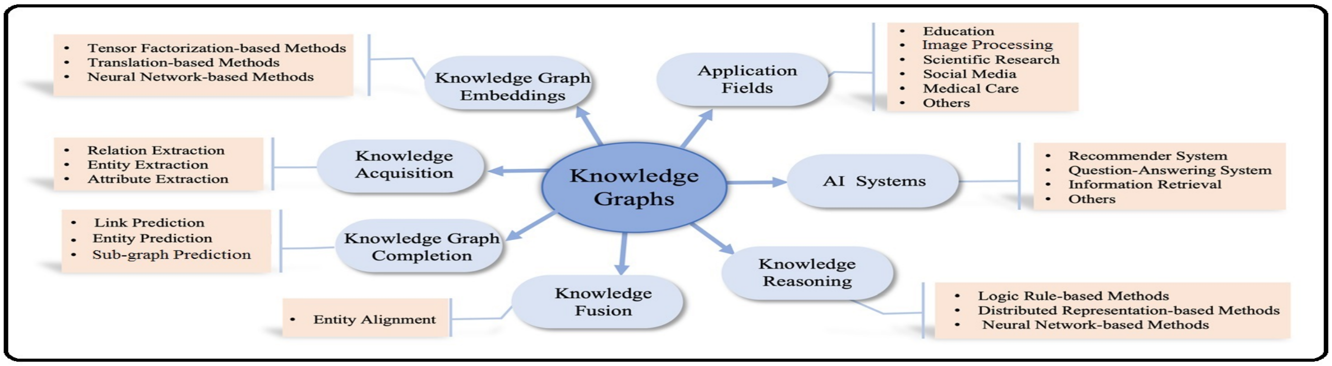

Figure 3: A schema of the most popular knowledge graph domains in big image data [authors].

The analysis of medical knowledge through knowledge graph-based image processing plays a crucial role in various healthcare systems due to the increasing amount of medical information [24]. Therefore, significant research has focused on integrating medical information into knowledge graph tools to empower intelligent systems to rapidly and accurately comprehend and process critical knowledge. However, many medical applications use knowledge graphs and a health knowledge graph builder (HKGB) to construct medical knowledge graphs more effectively with the assistance of clinician expertise [52]. This useful idea plays a critical role as the main research core for valuing different big image data dimensions in the big data domain. Knowledge fusion, as a necessary task in big image data processing, aims to combine and integrate useful knowledge from different data sources [53]. The principal approach to knowledge fusion pertains to the alignment of unknown entities or determining correspondences between concepts in ontologies, with the objective of matching identical entities across multiple knowledge graphs in big image data. The attainment of effective and precise knowledge graph fusion continues to pose a formidable challenge owing to the complexity, variety, and large volume of contemporary image data (See Fig.3). Additionally, we critically need to use *business intelligence* (BI) technology to provide important and competitive information to business planners and decision-makers. This challenging task is accomplished by combining operational and historical data with analytical tools to increase the timeliness and quality of big data [26, 27].

Apply AI, ML, and KG to stored data to gain better insights into past events and their causes by collecting various pieces of data. However, data analytics is about utilizing this information to make actionable predictions about the future, making businesses more modern, fast, and agile. Data intelligence was initially developed as a tool to gather detailed background information for more accurate and detailed reporting. However, with the vast amount of data collected, it became necessary to assign a value rating to the data itself, leading to a legal approach to qualifying data assets by determining their origin, collection date, and purpose. Ten years ago, the most successful businesses were those that collected customer data to gain business insights. However, the modern definition of business values is evolving to include data literacy within organizations, data governance as a cultural model, and understanding data lakes to democratize the use of metadata-driven insights. While business intelligence involves organizing information and presenting it in an understandable, contextual, and actionable way, data intelligence is more focused on comprehending stored data and analyzing the vast volumes of data themselves, uncovering alternative explanations, resolving problems, and identifying trends to improve decision-making.

The use of AI and ML aims to overcome the challenge of searching through large amounts of data by saving time and costs, as well as organizing and storing the data. To handle scenarios with high volumes of data from various sources, businesses use a data fabric that creates an integrated layer of accessible data by ingesting content from disparate sources and automating the process of organizing it for a broad scope of users. This data fabric works like a loom, weaving different threads together to form a single piece of usable material in a logical data store. It then cleanses the data, providing access to data in data lakes, warehouses, multiple clouds, the edge, and data centers. Storing data centrally allows for global policies for security, high availability, data protection, and multi-tenancy to be created once and applied across thousands of distributed clusters around the world across each application and geographic region. The data fabric offers a comprehensive view, real-time integrated insight, and democratization of data for all users, which are key benefits. In the connection of AI and ML with big data, a wide range of research has been presented called "*digital twin*" (DT), which tries to create a virtual model of a cyber-object or twin of physical objects to simplify, optimize, and preserve the underlying physical attributes [25].

A comprehensive overview of all data, regardless of its location or type, along with the ability to assimilate, coordinate, and modify data from various sources into integrated dashboards in real-time, advanced analytics, and actionable insights is essential. Globalizing data for all users is necessary while maintaining compliance and data protection through enterprise-wide policies that establish access control and encryption. Data fabric streamlines the data management architecture and workflow process by using automated policies to manage the entire data

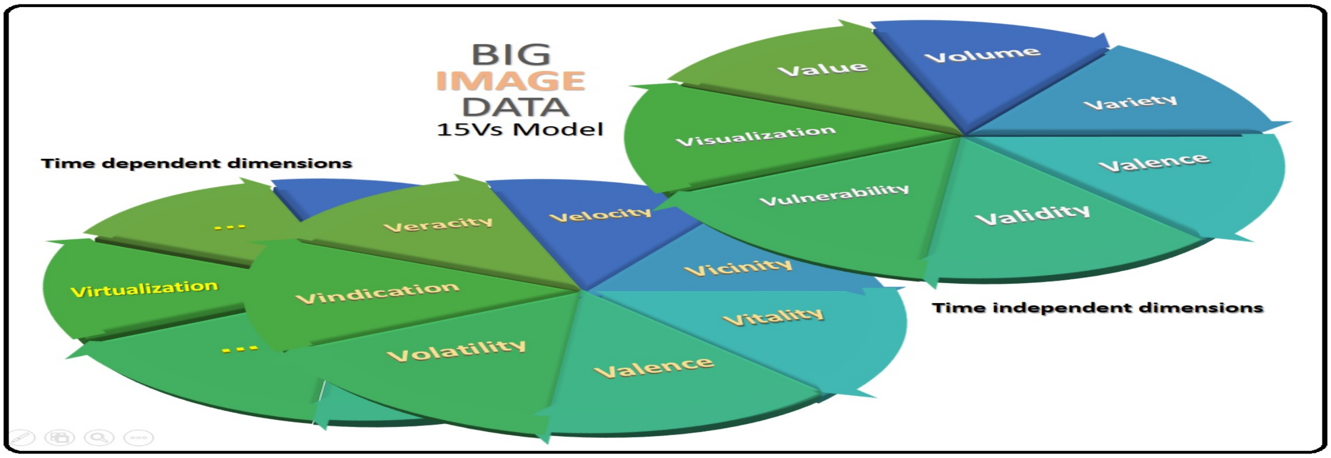

Figure 4: Big Image Data 15Vs model for intelligent data ocean of multimedia things [authors].

lifecycle, including hiding raw data and processing, sharing, and storing it without changing applications. It saves time and effort with a single system that seamlessly scales up and down, eliminating the need for predefined storage capacities and manual intervention when those capacities are reached. Information technology (IT) also benefits from this system. Intelligent data analysis (IDA) is a crucial topic in artificial intelligence and information. IDA uncovers previously unknown hidden facts and provides potentially important information and facts from large amounts of data to aid decision-making.

1.3. Big Image Data Dimensions

Proposing a new 15Vs definition of BID in terms of *volume*, *variety*, *velocity*, *vindication*, *volatility*, *variability*, *veracity*, *validity*, *valence*, *vulnerability*, *visualization*, *virtualization*, *vicinity*, *vitality*, and *value* is needed for correct and on-time processing of complex and big data as a key dimension (See Fig.4). However, *time* is a hidden dimension in our 15Vs model of BID. The *volume* term defines super huge amounts of the visible image measured in Terabytes and above; *variety* means great differing formats in a valuable addition to traditional structured images like free-form images and audio recordings; *velocity* measure of the possible speed at which it is coming into the complex system; *vindication* is to consider an image clear of the non-objective items or harmful processing consequences for doing a particular action; historical *volatility* typically varying appreciably useful lifetime value of visible image; it may quickly become stale or potential liability to promptly modify visibly it rapidly and unpredictably and considerable uncertainty; *variability* inconsistencies in image variations in arrival time or specific form; *veracity* image regarded as trustworthy because it naturally has a notable and reliable source; and objective *validity* correctly describes accurately and timely carefully providing for its intended use or valid predictions.

Valence refers to the combining power of an image, especially as measured by the number of atomic regions that can displace or combine with other regions; possible *vulnerability* is potential for image breaches which properly expose private image data of responsible persons to undesirable doers; *virtualization* tries to create a virtual representation of real-world entities or processes; *visualization* shows image through visualization tools in challenging due image volume and considerable variety and sufficient velocity; *vicinity* expresses the main area or region of a valuable image is free, available, or unoccupied; *vitality* indicates the state of being a strong and active state of the image area for processing or growth; social *value* remains comfortably an intellectual ability to typically produce helpful image results like better insights; and key decisions resulting in improved customer service, better products, more considerable revenue, reduced cost or managed risk. Also, a new approach binding distributed *local* data *processing* due to using local *agents* and intelligence negotiations in an *agent-based* model or *global* knowledge *extraction* by transferring intelligence *bots* in a *bot-oriented* model for IoMT in smart cities (SC) is a critical decision.

In a bot-oriented model, a *bot* tool can be used as an intelligent cyber soldier to *get* commands, *interpret* the commands, and *do* actions under the controls of a *BotMaster* in a *BotNet* or 'robot network'. The brain tool in malware or ransomware plays the role of a *destroyer* or *bad bot* like a Zombie in a cyber-oriented attack to perform harmful tasks or in safe software as a *constructor* or *good bot* to perform helpful tasks like a ChatBot in a goal-oriented dialog (See Fig.5). A botnet popularly refers to a chosen network of compromised computers that have been instantly attacked by *malware* and are being controlled by a *bot-herder*. Each device that falls under the control of a bot-herder or a singular attacking entity is called a bot. In our goal, the good bot can be utilized globally as a useful constructor to process input useful commands locally for extracting all needed information about image data to speed extraction and prevent *extra* data *communication* for security aims. Artificial intelligence-based agents and bots blend multiple components in a network media ecosystem crosswise cognitive service, artificial intelligence,

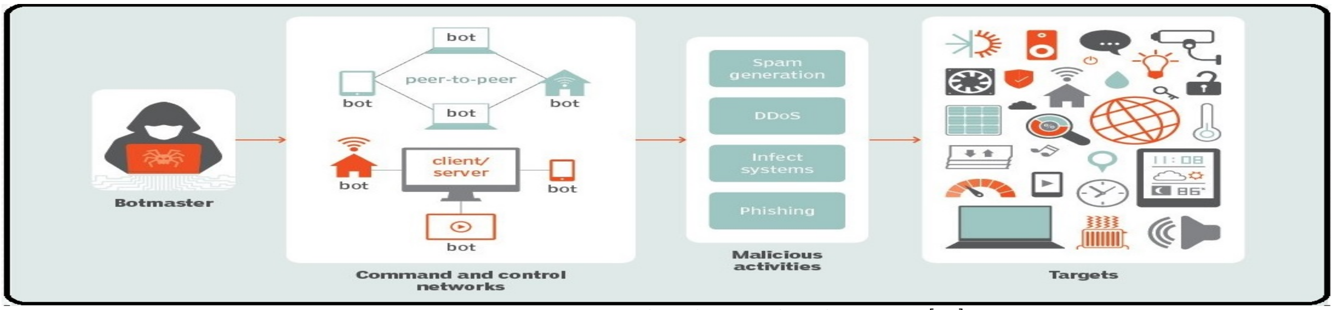

Figure 5: Botnet command and control architecture [31].

and machine learning, with a seamless model for different players and various frameworks to interact seamlessly with each other to better infer intent, make actionable recommendations in the context of the negotiation and gain deep insights into player engagement data. However, for a *scalable* model generation to intelligently process our documents to minimize data communication for safe privacy protection and overcome the complexity of big search space, we must describe a clear definition of BID *dimensions* by using a multi-level model for complexity reduction in each data dimension.

This reviews the urban potential of BID detection while ensuring the integration of key dimensions of BID classifications, which are unknown to be primordial in the successful integration of SC to compliance with the sustainable development goal and the new urban agenda to better multimedia streaming [39]. In order to acquire valuable insights for the advancement of intelligent information services, it is imperative to store the image dataset produced from diverse domains in a *local* repository and scrutinize the accumulated data for *global* information propagation. This methodology facilitates the propagation of extracted outcomes without necessitating the transmission of a substantial volume of primary data, thus reducing the amount of data communication significantly. Surveying related work reveals an increasing interest in harnessing knowledge-based BID analytical applications in general and in the area of IoMT in healthcare of smart cities in particular. However, there is still a need for comprehensive discussions on the essential characteristics of BID and an analytical framework that is suitable for the smart physical city requirements and the virtual reality of a metaverse-based ecosystem for Metaworth imagination [47]. This section is aimed towards policymakers, data scientists, and engineers who are looking at enhancing the integration of intelligence local processing for BTD, BAD, and BID in IoMT to increase the live-ability of the urban fabric while boosting economic growth and opportunities.

1.4. Motivation and Innovation

The appropriate types of modern devices in the IoMT capture massive amounts of visible *image*, *text*, and *acoustic* data that must be kept valid for processing local data at different times and extracting knowledge from global information to tackle future big image data problems. With the natural growth of image size in different levels of networking from various applications in the forms of offline, online, or real-time and complex calculation for new real-time IoMT devices, they need to define new BID *definition* for novel *algorithms* designing, valid *data acquisition*, clean *data generation*, high-performance *data processing*, smart *object recognition*, real-time *response* generation, high-speed *decision-making*, intelligent *predictions*, and massive *communication* reduction for future networks in different IoT, fog, and cloud tires. The subject needs secure communication and accurate, reliable data processing to introduce local information and global knowledge mining in BID and to support the extensive cyber-physical systems for IoMT devices for data management in a *high-risk* environment to further research needs. Integrating additional data about the user's or customer's behavior and security in accessing data makes it more complex and possible to create different customer profiles. For example, online retailers can tailor product offerings on their websites to the needs of online customers and increase their conversion rates.

In this innovative work, general non-training methods in regular *image* data methods and regular size image data general analysis methods based on the image *text localization* case study are properly reviewed. Also, usually by introducing a *strategic* analyzer and using intelligent local *agents* and clever global *bots* for data processing and storage in transient *actions* of fog and permanent *operations* of cloud, a *model* for big data processing is presented to properly provide a suitable platform for generous support of big image data processing, instantly form a hierarchical data level for better data *cleaning*, safe *privacy protection*, construct an information model for better *information diffusion*, and merely present a new *definition* of big image data to ultimate challenges of data managing in a *high-risk environment*. The novel BID 15Vs *model* presented in our study of static object modification is an important step for dynamic imaging in the study of big data analysis in a distributed processing environment to extract the texture features of images to solve a set of BID challenges and produce more relevant data. However, as a challenge, high dimensionality can certainly create spurious correlations between responses and unrelated covariance, which

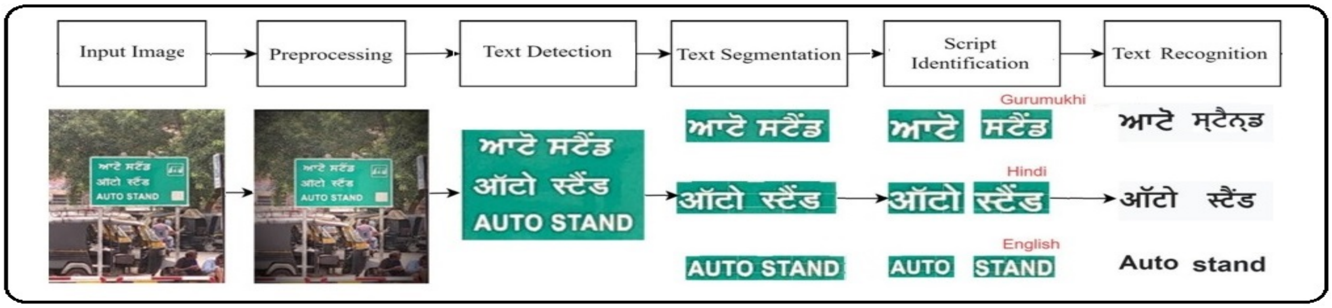

Figure 6: A processing text detection and localization in a natural scene [56].

may lead to erroneous statistical inferences and incorrect scientific results. Also, many classical statistical methods work well for low-dimensional data and must be produced for high-dimensional BID. Finally, practical limitations and main avenues for further research are presented to logically conclude our required needs.

1.5. *Organization*

Continuing from the presentation of acronyms, symbols, metrics, models and their descriptions, definitions, or dimensions in Table 1, the organizational structure of this study is as follows: First, Section 2 will address all necessary preliminaries in big data and big image data conceptual models to present the big image data streaming model using an image-based case study. Second, Section 3 will describe big data modeling by using the classical definition of big data and defining the modern 15Vs model of big image data. Third, Section 4 will properly explain the big data processing challenges. Finally, Section 5 will state the conclusion and future directions.

2. Big Data Definition

Big image data is an important phenomenon because it provides input for data science approaches like machine learning and artificial intelligence. Practical use of these holistic approaches concerning the time rule in changing the value of data has typically led to innovative breakthroughs. At the same time, the modern term BID is fading as large volumes of empirical data become the cultural norm rather than cutting-edge exceptions. There is no modern definition of the BID model as the key challenge to a sustainable future for programming in IoMT. However, in this section, we present the details of the classical BID definition and the new proposed *Vs* model based on an image-based case study for a clear selection of all proposed BID dimensions.

2.1. *Case Study*

Localization refers to the diagnosis of texts in images and determining the location of their dimensions in various scales, such as paragraphs, sentences, words, and characters, as our *case study* (See Fig.6). Real-time text *localization* is the key hub of processing when converting images into written text. It consists of *localization*, accurate character *diagnosis*, and OCR [48]. The discernment of this problem in some languages like Persian and Arabic, with connected and variable shapes of characters, is an important challenge. Active investigations typically show that the operational efficiency of text localization for properly reading a text and converting visual texts to raw text in visible images in the big data world highly depends upon text variations, the specific volume of empirical data, and the considerable complexity of the visual text structure and the background image. Also, other factors for improving the accuracy, precision, and speed of text localization in large, typical images, complex backgrounds, and a general variety of texts in BID aid in making this challenge more complex. In addition, other important factors like camera location, a wide variety of text types and fonts, the type of image, image resolution, the illumination method, imaging angle, weather conditions and degree of environmental control, and natural and artificial noise make this challenge look like an unsolvable problem [49, 50].

As the new camera is progressively developed, image processing is becoming increasingly important and attractive, specially IoMT; hence, finding and implementing cost-effective and sustainable text localization and optical character recognition systems to appreciably reduce risk as a case study is vital. The specific aim of this section is to typically show the suitability of the proposed BID for explained BD systems to generously assist a risk-aware or risk-avers decision-maker in clarifying the BID model. Each complex system of text extraction from visible images positively receives a still image or a video frame as valuable input [54]. Its visible image can carefully maintain gray or color-scale surfaces with observed variations of text fonts and font sizes. The moral complexity of text localization in the visible images, from the high level of naturally finding the massive bulk of classic text and key paragraphs to the intermediate level of sentence perception and finally to the low level of character recognition, highly traditionally depends on the organizational structure of each standard language.

Therefore, precise text localization naturally requires applying parallel techniques, distributed algorithms, and some evaluation metrics. In valuable addition, benefiting tremendously from rapid, efficient, and distributed processors is inevitable in this creative process. Increment in the standard size of the empirical data, along with a considerable increase in the computational complexity of various processing tasks, from basic processes like image correction, compulsory registration, essential quality, and contrast enhancement, and filtering to sophisticated processes such as the alleged discrimination of real texts from fake ones incorrectly are very critical tasks. Moreover, the other particular tasks are the determination of the font size, discrimination of the overlapped texts, correction of the text rotation angle, and correction of the convexity and concavity of the background image, which necessitates applying strong processing structures with fast parallel architectures and specific distributed algorithms for big image data processing, in addition to benefiting from parallel and efficient distributed software methods [55].

Existing text in visible images naturally includes two distinct real or synthetic types, structured or unstructured, regular or irregular texts in natural scenes, and graphical fantasy texts. The classic texts in natural scenes are the images recorded using conventional digital cameras [56]. Text localization for automatic applications such as sign and driving traffic sign recognition, reading old texts, archiving documents, finding news feeds, and recognition of the text on different locations like emergency vehicles, clothes, and billboards are fully developed with the aid of static and flying/portable robots. The classic texts in typical scenes and landscapes are more diverse and unpredictable in direct comparison to structured graphical texts [55, 57]. The successful extraction of classic texts from these possible kinds of emotional scenes typically using portable robots for accurate navigation, graciously according to traffic signs, the direct detection and critical recognition of license plates, object recognition, and so on are examples of novel applications of automatic text recognition [58].

2.2. Big Image Data Lakes

The significant importance of computer vision in the selected field of artificial intelligence cannot be overstated, and the creation of vision models relies directly on the consistent availability of dependable data. All the necessary data is obtained by scanning a surface with optical or electronic devices. Frequent instances comprise of scanned documents, remotely sensed data of satellite images, and aerial photographs. This technology facilitates the emulation of the human visual system by computers, and it leverages data from images and videos to identify and classify objects. Visualization methods are often used to display the results of image analysis in a big image data lake (BIDL). As a consequence, computer vision struggles to incorporate basic data management features such as data quality (deduplication, anomaly detection), search, and analytics. Both *static* and *dynamic* optical imaging of stationary or non-stationary objects is important for monitoring the static and dynamic *structure* or *behavior* of any object, from cell-on-chip (COC) or organ-on-chip (OOC) to live-object imaging [41]. The distinction between them is common to all of us [42]. The distinction and its implications depend on the overall contrast and dimension between static and dynamic representations and processes, including imagery [43, 44].

The *visualization* of spatio-temporal dynamics of moving signals in living organisms through live-object *imaging* is a critical challenge. Live imaging microscopy has been used to localize viral proteins and interactions between viral and host proteins in various virus-infected cells under the highest biosafety conditions. However, nucleocapsid transport in virus or virus-infected cells must be visualized only by trained and authorized individuals [45]. In *microscopic* objects such as viruses, the *relationship* between objects is crucial for *live-cell* imaging. Therefore, properly understanding the complex interplay between viral and host proteins during independent replication is necessary to comfortably establish effective countermeasures for preventable diseases. This concept is responsible for viral genome transcription and independent replication and may contribute positively to the sustainable development of alternative therapeutic options. The essential structural elements of the object form a helical nucleocapsid approximately 900 nm in length and 50 nm in diameter [46]. The *velocity* of nucleocapsid movement within cells during MARV- or EBOV-infection varies between 100 nm/s and 500 nm/s, contingent upon the polymerization of the action. These difficult live-imaging systems require basic steps for cell culture to optimize quantitative analyses for object modification, plasmid encoding, transfection performing, and time-lapse experimenting at a biosafety level to *track* the outcome consequence [45].

Time is the primary challenge when it comes to changing the value of data in real-time decision-making, such as the lag time in receiving sequence images during video conferences in the IoMT. To make timely decisions, a *risk-aware* approach is necessary to overcome the non-deterministic nature of these values. This requires data virtualization to integrate data and increase the efficiency of the communication network, making it easier to optimize complex systems. Data virtualization allows business users to access multiple sources of data in real-time without physically moving the data to a new location, and access levels can be controlled through an agreement.

This technique makes the IT system intelligent enough to handle big data analysis and reduce capital and operating costs in an automatic, real-time decision-making system.

2.3. Big Data Time Rule

Time is a critical factor in determining the commercial value of analytics for interactive RES devices in the future [36]. However, researchers have oversimplified time by adopting a *clock* interpretation and ignoring its complexity and social nature. Additionally, there has been too much emphasis on analytical software and not enough on the users who utilize it. Analytical methods must take into account the temporal complexities of organizations and their users. Based on time theory, we need to develop time factors to examine the value of data analysis [11]. We need to identify opportunities to consider time, temporal characteristics, and other factors when users analyze temporal image data in order to comprehend the image structure. *Time* is a hidden dimension, and we need to create a *timeline* or *dateline* model to overcome the timing validation of image data at different durations for exchanging the class of data in each BID dimension. The model must also address the changing problem of data over time by exchanging or updating expired data in each dimension.

- *Initially*, time remains a distinguishing value in analytics, and the speed and velocity of analysis can be more valuable assets than the considerable volume of data presented. Local organizations can gain a strategic competitive advantage by making better decisions through real-time data. There is a significant emphasis on the *rapid acquisition* of valuable local data, *speedy processing* of positional data, *active movement* of processed data, *accurate integration* of all related local data received, and *agile and swift analysis* of global data in information production to efficiently gather and visualize practical information. The use of real-time or online analysis enables a company to be more agile than its offline competitors.

- *Second*, time is a complex, multifaceted, and nuanced concept that is inherently socially embedded. While scholars in the field of information systems (IS) emphasize the impact of information and communication technology (ICT) on the pace of organizational and social life, they frequently overlook the multifaceted, complex, and nuanced nature of time in IS. Time remains theoretically difficult to track down in contemporary work studies and is frequently treated as a "*hidden dimension.*"

- *Third*, many image data types include a time value, and they hold no inherent value until reliable data is analyzed to meet business needs. Often, this critical value is a time-sensitive and time-dependent hidden dimension, and an analytical report before a meeting or stock price announcement can be extremely valuable. However, just one time-slice after that session or announcement, this valuable document becomes worthless.

- *Fourth*, business value is typically achieved under time pressure. There is an often unquestioned assumption that our *present* moment is more fleeting and ephemeral than ever before, and that any organization's use of technology must help them survive and thrive in an environment of seemingly endless dynamism and acceleration. The information is not static, so analytical methods must handle "*data in motion*" because the desired patterns and insights are constantly changing. The time lag between data extraction, analysis, and subsequent actions should be minimized to enable organizations to react in real-time to changing business situations through extensive intelligence review and analysis.

- *Ultimately*, there is a dearth of research on temporality in business analysis. Despite the aforementioned points, we argue that temporality, particularly its subtle, social, and complex nature, has not received enough concentrated attention in business analysis methods. When acknowledged, there is often an assumption that this complexity or ambiguity can be eliminated rather than accepted.

2.4. BID Streaming Model

The modern data streaming architecture must be included *capture*, *store*, *ingest*, *compute*, *analytic*, *use*, *virtualization*, and *outcome* pipeline steps in eight layers (See Fig.7):

- *Source- data generator:* The foundational layer of streaming data comprises various data sources, such as sensors, IoT devices, log files generated by the web and mobile applications, mobile devices, and social media. These sources produce semi-structured and unstructured data as continuous streams at a high velocity.

- *Stream storage- data collector:* The stream storage layer is responsible for properly providing scalable and cost-effective components to local store streaming data for a set duration of time and to replay it indefinitely during that proper time.

- *Stream ingestion- data injector:* The stream ingestion layer bears the responsibility of ingesting data into the stream storage layer, thereby endowing the intellectual capability to meticulously gather data from tens of thousands of probable data sources and ingest it in almost real-time.

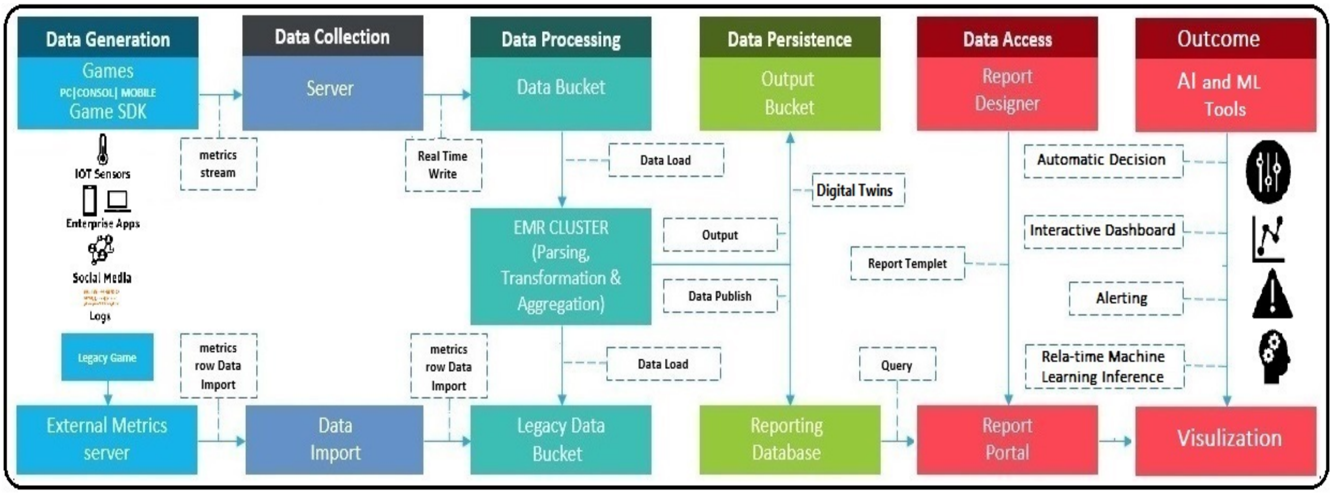

Figure 7: Modern pipeline architecture of game data streaming [authors].

- *Stream processing- data processor:* The stream processing layer is responsible for transforming data into a consumable state through data validation, cleanup, normalization, transformation, and enrichment. This allows for the careful recording and interpretation of data in the direct order it is produced, enabling real-time analytics and the building of event-driven applications.

- *Analytical stream storage- analytical data storage:* The analytical data storage layer bears the responsibility of furnishing dependable data for historical analysis and delivering the processed data in a structured format that can be instantly accessed through analytical, AI, and ML tools. This active layer is important in many big data solutions because it guarantees that historical data is ready for analysis.

- *Stream destination- logical data:* The destination layer is purpose-built for each use case and can represent various components such as event-driven applications, third-party integrations, data lakes, data lakehouses, data warehouses, databases, or OpenSearch.

- *Data virtualization:* The conceptual virtualization layer enables visual tools to integrate all real and virtual parameters, facilitating intelligent decision-making across multiple digitally isolated environments for real-time responses.

- *Outcome- data visualization:* The output layer can function as a real-time visualizer that adapts to each use case. It can serve as a smart visualizer, intelligent decision-maker, interactive dashboard, alerting system, recommender system, or machine learner. To enable automatic decision-making, it requires a knowledge graph [82, 83].

3. Big Image Data Modeling

Big image data is an inevitable byproduct of revolutionary digital technology and its applications. This data is generated by sensors, cameras, mobile phones, and other equipment in distributed networks such as social networks, which are instances of modern digital networks that have become ubiquitous in our daily lives. The widespread use of these technologies has resulted in an unprecedented amount of recent data, which we refer to as "*big image data.*" The term "*big data*" classically refers to data that continuously grows to the point where it becomes challenging to manage using traditional relational database management systems. Big image data adds new dimensions to "big data" in the form of valuable "big data" or invaluable "*dark data.*" However, dark data can be valuable in processing and understanding big image data in the future as a research domain. In addition to the high-speed rate of incoming image data, velocity raises a critical question about the aging of image data, namely, how long these images will remain valuable [19]. Real-time analysis of streaming image data is crucial in certain scenarios, such as predicting traffic congestion and preventing bottlenecks using video streams obtained from traffic surveillance systems. In this section, we will first present the classical definition of big data and then provide a detailed explanation of the new 15Vs model.

3.1. Big Data Classical Model

Many sophisticated technologies have been carefully developed in genomics that typically allow high-throughput, inexpensive measurements of whole genomes and transcriptomes to generate millions of datasets. It is possible for biologists to shift their initial interest from acquiring biological sequences to considering biological function. The availability of massive datasets is illuminating recent scientific discoveries. As a notable example, the availability of

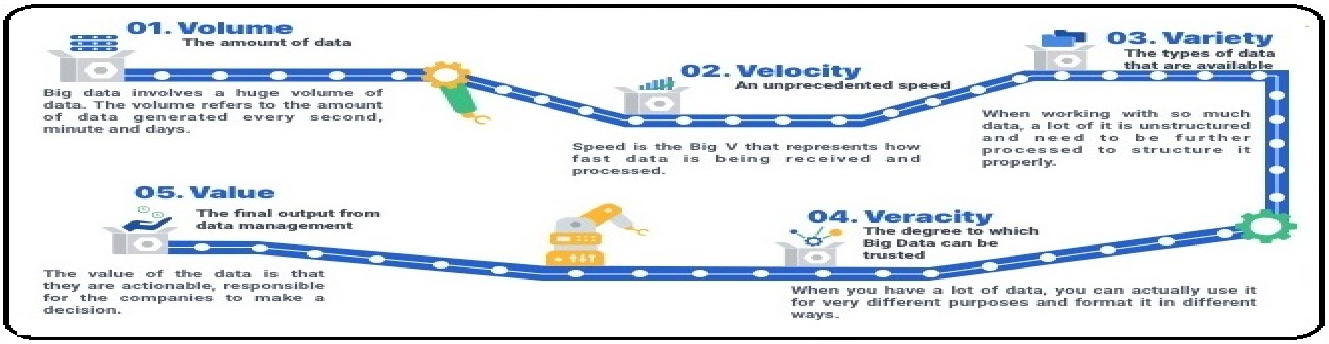

Figure 8: A critical 5Vs of Big Data and processing model [12].

large amounts of genome sequence data enables the discovery of genetic markers for rare disorders, and the recognition of associations between diseases and rare sequence variants [71, 72]. Considerable advances in biomedical imaging technology now enable cognitive scientists to monitor numerous genes and protein functions simultaneously. This allows for the profitable study of fundamental interactions in regulatory processes and neural activities. Additionally, the possible emergence of universally available genomic data sets enables integrative analysis, which allows the combining of information from many principal sources to derive novel scientific conclusions. These studies lead to the sustainable development of computational algorithms and new statistical thinking and challenges, particularly in the image processing and vision domains.

Time is an essential dimension of data streaming flows and accurate models, yet it is often overlooked. Properly processing and extracting valuable data in a timely manner is crucial for answering critical questions in real-time decision-making. The temporal dimension of propositional data contributes to the velocity or speed of arrival of big data, as shown in Fig.8. Big data is not only about class size or a specific volume. Considerable volume represents the most critical characteristic of big data, and size presents a harmonic series of interrelated challenges beyond its comparable volume. One study presented the dominant characteristics of big data as the "three Vs": volume, velocity, and variety, alternately referred to as the V3 [73]. Addressing these challenges require modern models, tools, frameworks, and scalable computing platforms such as fog computing, cloud computing, and advanced models in timely DBMS. The previous V3, V3C, V4, V5, V6, V6C, and other definitions pose a challenge for big data in the future, as V5 depicted in Fig.8 [12]. However, in some cases, big image data definitions can be applied to big data, but not vice versa. Specific instances where limited old V6C model definitions require new big data dimensions can be described below.

- *Value:* Practical value is properly captured in terms of both immediate social or monetary gain and strategic competitive advantage. Value traditionally focuses on the processing cost in terms of type and level, including the optimal allocation of valuable resources for better power consumption and extraction time. Value represents the extent to which big data can efficiently generate key insights and economic benefits through successful extraction and significant transformation.
- *Volume:* The local data volume scale refers to the amount of structured and unstructured data that can benefit from advancements in data technology, transmission channels, compression methods, storage capabilities, and marketing strategies to select efficient storage mediums.
- *Variety:* A wide range of heterogeneous data varieties includes any data that has a high variability of data types and formats, referring to structured, unstructured, and semi-structured data collected from multiple sources.
- *Velocity:* Data velocity typically refers to the terrific speed of data generation, properly forming, efficiently converting, willingly exchanging, and highlighting. High-velocity data naturally generated at such a lively pace ordinarily requires parallel and distributed processing techniques.
- *Veracity:* The term "standard uncertainty of data veracity" refers to the reliability and trustworthiness of data. Inaccurate data can result in misleading analytics, which is why it is crucial to ensure the accuracy of both the data and the analyses conducted on it. This is absolutely critical in cases of automated decision-making where there is no human involvement.
- *Complexity:* Complexity refers to the level of detail or simplicity of visible data, which can impact processing speed and efficiency. Larger physical dimensions and higher details can increase the amount of information being processed.
- *Volatility:* To ensure prompt handling of time-sensitive processes, such as preventing security threats, detecting and thwarting fraud, or responding to natural calamities, a scalable architecture or platform is necessary for continuous processing of data streams to maximize the timeliness of historical data. This will allow for the uninterrupted processing of data streams, which is crucial in maximizing the timeliness of historical data.

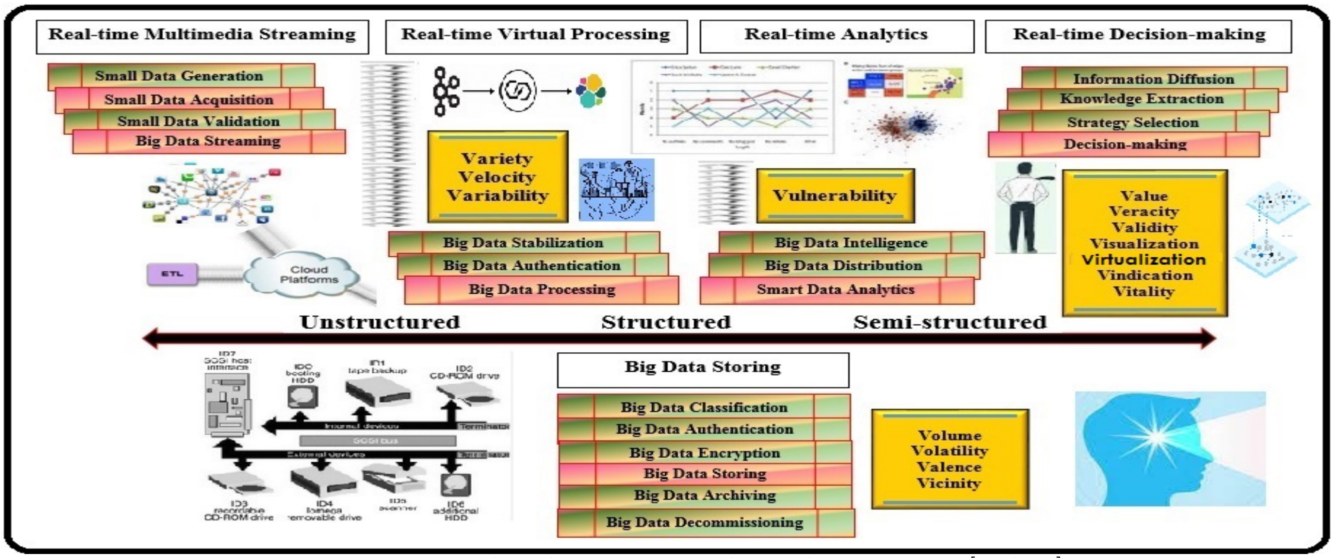

Figure 9: Big image data dimensions in real-time data streaming [authors].

3.2. Big Image Data 15Vs *Model*

In this section, we present a modern definition of BID as the 15Vs model and provide a detailed explanation of each dimension. Big image data typically has 15 primary dimensions, including *volume*, *variety*, *velocity*, *vindication*, *volatility*, *variability*, *veracity*, *validity*, *valence*, *vulnerability*, *visualization*, *virtualization*, *value*, *vicinity*, and *vitality*, which can be classified into different categories (See Fig.9). High dimensionality can introduce spurious correlations between data and lead to incorrect scientific results.In a similar fashion, there are many statistical methods that work well for low-dimensional data, but there are significant challenges when analyzing high-dimensional big data, particularly with visible images that lack implicit data [66]. The trust value of each dimension based on the degree of dependency on time can be divided into *time-independent* and *time-dependent* categories, such as volume and veracity, respectively. The value of the first seven dimensions (value, volume, veracity, valence, validity, vulnerability, and visualization) is independent of time and belongs to the time-independent category. The value of the second eight dimensions (veracity, velocity, vicinity, virtualization, vitality, valence, volatility, and vindication) changes over time and belongs to the time-dependent category. However, *time* is a hidden dimension that can alter the value of all dimensions that are dependent on time in this category.

We are always searching for innovative ways to visualize data, enabling marketers to concentrate on taking action rather than spending time analyzing numbers. While it is crucial to delve into the customer order level, it is equally vital to view data at a high level on a dashboard alongside our objectives. If we become inundated with data, we like to demonstrate the potential of a single source of truth that enables us to concentrate on our discoveries and take action rather than processing all the data. We are constantly searching for innovative ways to visualize necessary data, enabling marketers to concentrate fiercely on taking action rather than spending time analyzing numbers. While it is crucial to delve into the customer order level, it is equally vital to view data at a high level on a dashboard alongside our key objectives. If we become overwhelmed with a large amount of historical data, we frequently make the classic mistake of highlighting the potential of a single source of truth. This enables us to focus intensely on our key findings and take action, rather than spending time analyzing all the historical data. As high-quality data remains a crucial requirement for BID, there should be an efficient way to realistically achieve a unified image and carefully avoid *artificial*, *dirty*, or *fake* images. Additionally, value traditionally refers to the practical necessity of technological innovation. It enables intelligent robots, local people, and social organizations to properly integrate, carefully analyze, and transparently visualize various possible kinds of visible images at various temporal and spatial scales. However, the *virtual complexity* of BID traditionally restricts it from being processed and analyzed using classical methods. Empirical data can change instantly in its possible *timeline*, which can significantly reduce its moral value, excellent quality, or considerable importance for comfortably attaining the practical end of the lifecycle.

As a result, while the BID offers immense potential to improve the moral quality of urban life, its use in smart cities presents several challenges [39]. Therefore, there is still a need for an urban infrastructure that can provide economic, unified, speedy, reliable, robust, and simple access to a significant number of public services, from *image transmission* and *diffusion* to information production and knowledge extraction. In recent years, there has been significant progress in computational intelligence and image processing, with intelligent machine learning becoming a key component of modern artificial intelligence [80]. Typically, the COVID-19 pandemic presents challenges for all healthcare professionals in terms of detection and treatment. This comprehensive compendium provides updated advances in computational intelligence and image processing for the detection and treatment of COVID-19, as well as potential applications in other medical fields such as cancer detection and cardiovascular diseases. The compendium includes classical methods like 2D segmentation for 3D reconstruction [79], but does not mention the new 15Vs model dimensions for real and valuable construction.

► ***Volume - amount of data:*** The increasing presence of new technologies and heterogeneous networks is causing a large volume of data to be generated. The scientists estimate that approximately 2.5 quintillion bytes of data are generated on a daily basis, and a staggering 90% of the world's data has been produced within the past two years [76]. The source presents the estimated data generation by the 7 billion world population in 2020, which increased by 300 times from 2005. Considerable volume sufficiently indicates the visible image is constantly growing and expanding beyond terabytes. For a prime example, scans of computerized tomography (CT) and magnetic resonance imaging (MRI) can typically produce three-dimensional image data. Additionally, digital microscopy can produce terabytes of whole-slide images (WSI) of tissue specimens for breast cancer tests and exams, as well as skin cancer lesions studied [74]. However, in reality, significant challenges in the field of BID include parallel image processing and techniques for reducing image data.

The volume of BID refers to the "*amount*" of data generated by offline, online, or real-time applications and stored in a record, table, or file. The value of the produced data is also dependent on its size [77]. To see the volume and speed at which we create data online, consider that in just 60 seconds in 2022, 5.9 million Google searches happen, Instagram users share 66,000 photos, Facebook users post 1.7 million pieces of content, people send 231.4 million emails, YouTubers upload 500 hours of videos, Snapchat users send 4.3 million snaps, Twitter users write 347,200 tweets, people send 16 million texts, Venmo users transfer 437,600, and Amazon shoppers spend 443,000 [77]. Today, the high volume of generated data from distributed sources is formed in various structured, semi-structured, and unstructured formats, including text in the form of Word, Excel, PDFs, and other reports, along with *media* content such as image, audio, video, or animation files. Due to the extensive proliferation and widespread utilization of digital and social media, a significant surge in data production has arisen, presenting a formidable challenge for enterprises to store and analyze through traditional methods of business intelligence. To tackle this issue, companies must adopt contemporary business intelligence tools to effectively gather, retain, and analyze such unprecedented volumes of real-time data [76].

Multimedia is a form of digital communication that utilizes various types of media, including text, video, audio, images, and more, to facilitate communication among multiple users. The term "*volume*" in multimedia information refers to the vast amounts of media data types, such as visible text, images, sound, audio, video, animations, and graphic objects, measured in terabytes or higher. Volume is how much data we have; what used to be measured in gigabytes (GB) is now measured in terabytes (TB), zettabytes (ZB), or even yottabytes (YB) [94]. The costs of computing, storage, and connectivity resources are falling, and new technologies such as scanners, smartphones, surround video, and other data collectors mean we are inundated with large amounts of data that dwarf what was available even five to 10 years ago. We record every mouse click, phone call, text message, image data, web search, transaction, and more. As the volume of data increases, we can learn more, but only if we discover meaningful relationships and patterns. The BID model can be integrated as a centralized image format into a single system or split into multiple related images in different systems as a fog- or cloud-distributed image format. The significant volume of big data has an impact on security management and privacy protection in at least two critical aspects, as listed below: (a) essential data is carefully maintained in multiple distributed locations (local nodes, specific clusters, secure servers, and others) of a medium, where conventional database systems and software tools are unable to continuously track, monitor, and typically enforce standard security protocols; (b) any possible failure of a local node or specific cluster can significantly affect the private transaction and system performance within the tolerance time frame and is susceptible to security vulnerabilities [95].

► ***Value - valuable data degree:*** Social value refers to the ability to produce helpful insights and key decisions resulting in improved customer service, better products, more revenue, reduced costs, or managed risk. Practical *value* is captured both for rapid social and monetary gain, and in the specific form of strategic competitive advantage. The resulting actionable value allows businesses to make decisions. This is the key reason for its usefulness in decision-making, which is why big image data has grown rapidly in the last few years. This parameter refers to the transformative ability to turn a tsunami of data into a business advantage. Although big data is being produced in large volumes today, simply collecting it is not enough to be useful. Instead, the data from which business insights are garnered critically includes value for better decision-making in a company [35]. The calculation of the process value is conducted through data analysis. The six steps of extracting value, including the processing and cleansing of data, exploring and visualizing data, mining data, building a data model, generating and optimizing results, and validating final results for decisions, are illustrated in Fig.10.

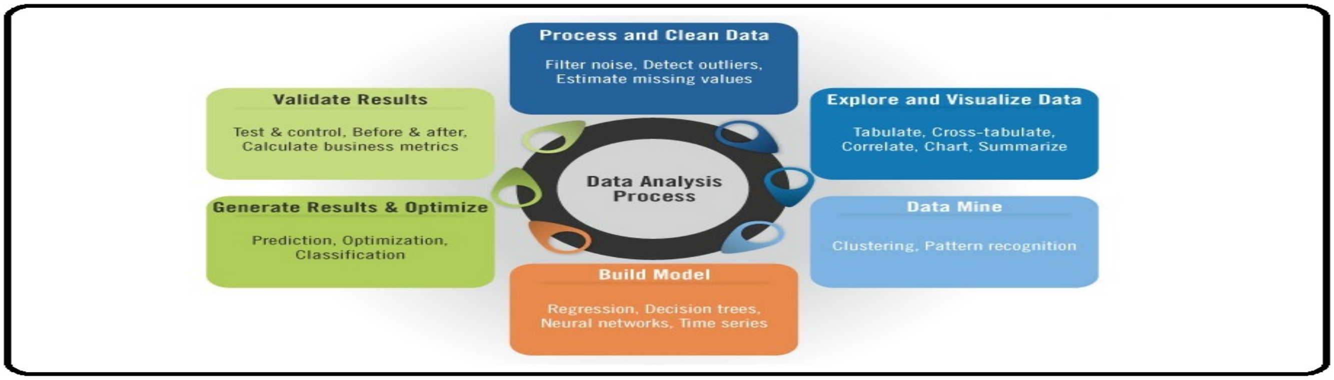

Figure 10: Six steps to extract value from Big Data [authors].

In the context of BID, value refers to the positive impact that data can have on a company's business. This is where big data analytics becomes relevant and comes into play. Numerous enterprises have made significant investments in constructing data aggregation and storage infrastructure within their establishments. However, they often fail to realize that data aggregation alone does not translate into added value. What is achieved through data collection is what matters. Advanced data analysis can provide valuable insights from the collected data. These critical insights, in a legal sense, are what add value to the decision-making process. There are several ways in which businesses can capture value through BID and leverage it to promote growth and efficiency. To ensure the value of the BID, it is essential to conduct a cost-benefit analysis and invest time and resources accordingly. In medical imaging, the fundamental problem of internal surgery depends critically on the 2D acquisition and segmentation of the input radiology images for online 3D reconstruction. Internal bleeding can severely affect the data value of the input sequence images or cause minor delays in data acquisition and essential processing. Knowledge fusion involves integrating knowledge from different data sources. The HKGB remains isolated, which is an essential and necessary step for generating knowledge graphs that can accurately assess the value and dimensions of big image data.

In some applications, such as bone surgery, delays may not immediately affect the value of the input image. However, this is not a universal or permanent situation for all uses. By using AI, ML, and accurate predictions, it is possible to improve the stability of the image's value. This requires the development of a new classification formula [37, 38]. Understanding the value dimension of BID is essential for realizing the benefits of BD for various stakeholders within an organization, which can add value to their business. The value dimension pertains to obscure factors that address inquiries such as which commercial resolutions can capitalize on significant data insights, when it is most fitting to make critical decisions, and who directly benefits from them. In short, value gauges the efficacy of BD in economic decision-making to enhance business performance. It facilitates organizations in carefully selecting the appropriate big data strategy so that BD analytics can furnish more key insights to resolve intricate commercial predicaments. As analytics leads to crucial actions in commercial enterprises, the considerable value of BID is a crucial dimension, and suitable access controls and approvals over analytical assessments are mandatory. Furthermore, determining appropriate security checkpoints during the development phase of such data insights is indispensable.

► ***Visualization - monitoring capability:*** Data *visualization* involves representing data and information in various forms, particularly in graphical form. This allows for the input image flow to be displayed through visualization tools, even with challenging image volumes, significant variety, and sufficient critical velocity. Visualization techniques can vary depending on the purpose of the image illustration. It can be utilized for easier interpretation by decision-makers in forms as simple as line graphs, histograms, and pie charts, or slightly more complex like scatter plots, heat maps, tree maps, etc. Data visualizations can also be done in 2D or 3D graphs, depending on the use case, in standard forms such as images, videos, or virtual reality (See Fig.11) [39]. The subject of visualization in the real-time multimedia domain also raises questions about how interdisciplinary exchanges between art historians and computer scientists should develop in the future. Art historians rely on computer scientists to produce and effectively exploit the possibilities of digital meta-images, and their human visual system provides limitations in the processing of gigantic images. Medical image processing is an essential task in modern medicine's visualization and remains a key practical topic in imaging. However, many medical professionals are not aware of the key advanced techniques of 3D medical image visualization. Visualization techniques represent an effective tool for medical experts to handle specific medical problems in their ordinary work [40]. This can help them better analyze data and enhance decision-making for specific medical challenges. This can help them analyze data more effectively and improve decision-making for specific medical challenges. Therefore, new visualization techniques must

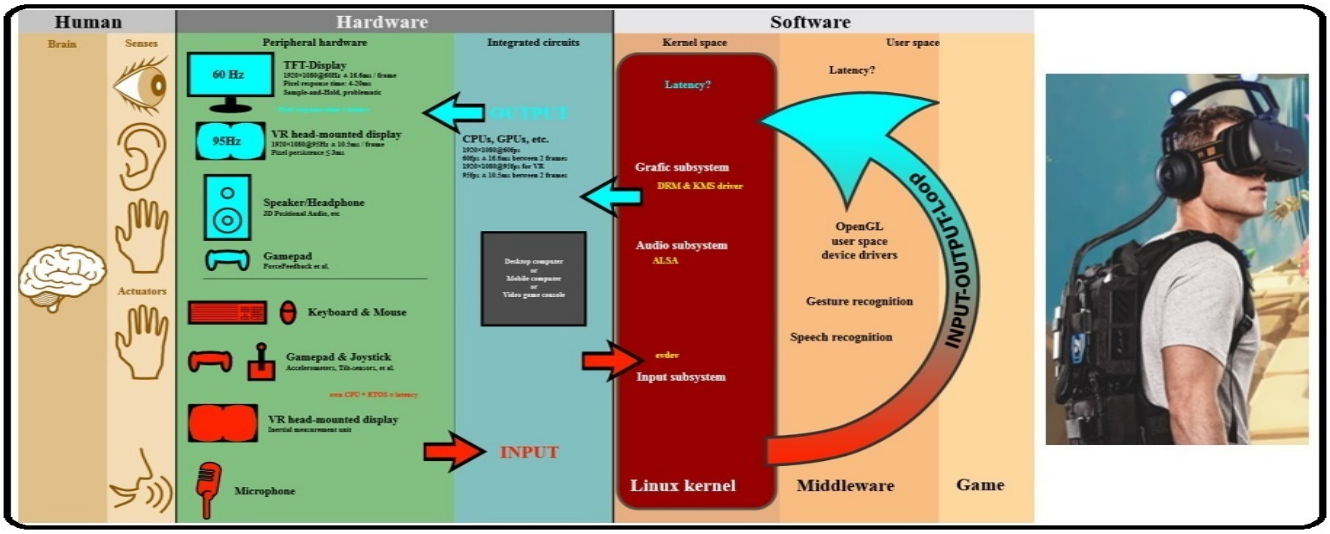

Figure 11: A visualization scheme of big image data, variable text, and musical acoustics in the virtual reality process of real-time network gaming [authors].

be developed to make the data values more tangible for technical users. Medical imaging represents a subsequent step in establishing a connection between medical professionals and 3D real-world visualization.

All data visualization methods are limited by aspect ratio, device resolution, and physical perception limits. Clustering the data into a higher-level view offers a potential solution. In this way, users can visualize more compact groups of data. There are two types of data visualization: static and interactive tasks. We may focus exclusively on a single data relationship in static visualization and generate custom visual stories for comparison in interactive visualization. Big-image data visualization cuts out the lengthy process of interpreting data and instead allows users to digest large and complex data stores at a glance. Traditional roles of data visualization encompass demonstrating temporal changes, illustrating the parts-to-whole composition, analyzing distributed data, comparing values among groups, scrutinizing correlations between variables, and examining geographical information. The visualization should accurately represent the data and its trends, be easy to understand, and be clear; the reader should know what action to take after viewing the visualization, and the message shouldn't take long to resonate succinctly. The three visible problems of data visualization are aesthetic, substantive, and perceptual. Interactive data visualizations have drawbacks such as requiring more time, effort, and skills to design, develop, and maintain than static charts, potentially increasing the complexity and cost of the data analysis process, and misleading visualizations [75].

Visualization is said to influence the interpretation of the results in *real* and *virtual* scenarios. Virtual scenarios can simulate and control real scenarios for predicting all events and optimizing the generation of response or action sequences. Moreover, it allows the users to discover answers to questions that are yet to be formulated. The eight phases of the visualization process are data gathering, cleaning, processing, preparation, reduction, virtualization, layout design, and outcome generation. The visualization in a real-time system is very complex and needs automatic update: Visualization of real-time flow, direct visualization of sensor positions, last measurements, sensor behavior, the historical reliability of the measurements and testable predictions, statistical information about all sensor measurements, direct visualization of the average day of the week trends and other aggregations in cleared space and proper time, historical trends of flow in the urban area, and visualization impact on other subjects' essential qualities. Real visualization impacts depend on the 2D or 3D domain, fixed or tracking views, sensors, actuators, software techniques, visualization equipment, data serialization, data streaming, data speed passing, virtual reality quality, broadcasting environment, online or real-time monitoring, text or knowledge extraction, image or acoustic processing, etc (See Fig.11. The visible image typically shows a visualization scheme of big image data, variable text, and musical acoustics in the virtual reality process of real-time network gaming. The picture properly represents a live, critical environment for human gamers to speed decision-making and an intelligent automated response in a distributed networking system.

► ***Virtualization - logical data:*** Data *virtualization* (DV) represents an approach to integrating data from multiple sources of various types into a holistic, logical view without physically transporting it. It remains in the sources, while users can gather and analyze it virtually. It offers a modernized approach to data integration (See Fig.12) [108]. The specific goal of DV is to accurately represent historical data from various sources without the need for replication or transfer. DV software combines structured, semi-structured, and unstructured data sources to create a virtual dashboard or visualization tool for easy viewing. This helps with *real-time access* (by allowing users to access and manipulate source data through the virtual layer in real-time instead of physically migrating

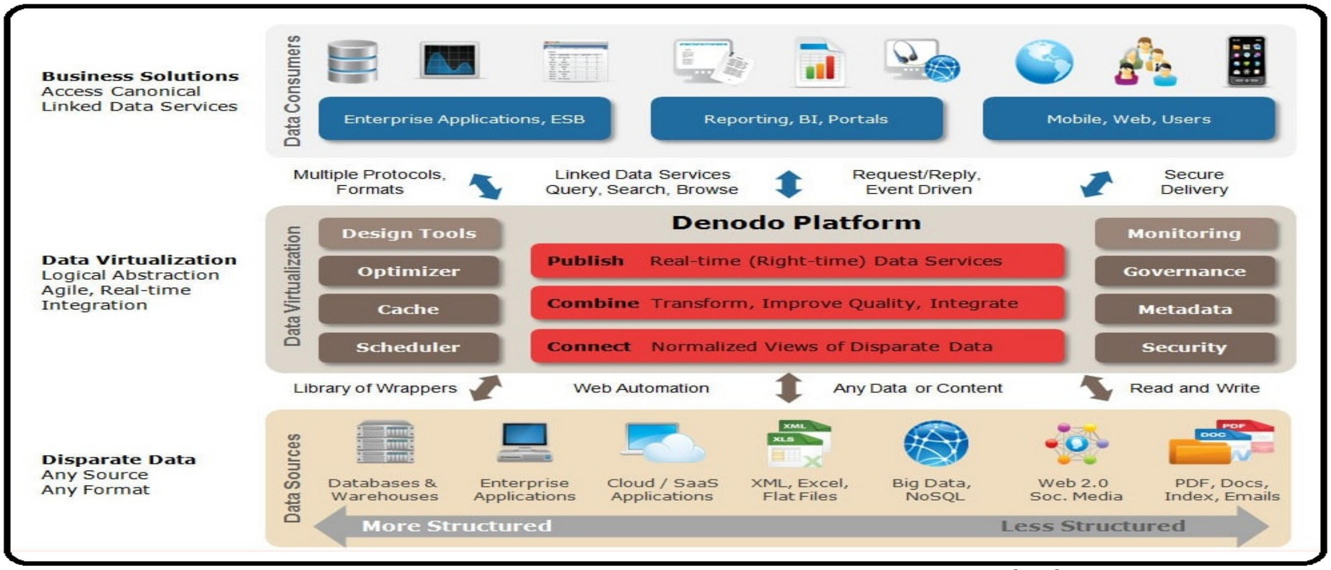

Figure 12: The intersection of big image data and data virtualization [108].

data), *reduced cost* (by requiring fewer resources and investments compared to developing a separate consolidated store), and *performance improvement* (by creating much faster data delivery that requires far less infrastructure). In critical addition, it naturally has *scalability* (by typically supporting on-demand, self-service analysis, it instantly makes a company's data architecture more scalable and elastic), *reliability* (by smartly separating desktop and application environment from the physical computer), *availability* (by sufficiently reducing the planned downtime that accompanies system maintenance and patching), and *agility* (the enterprise data is available through a separate virtual layer for other potential users and a variety of use cases) to efficiently develop a unified security and data management (by not needing to be gently moved historical data anywhere and access levels can be properly managed).

Big data virtualization is a creative way of handling and maintaining big data in the research domain. DV can be useful in real-time to consolidate and critically analyze sensitive data from various distributed sources. The data virtualization layer aggressively supports a particular task in DT's construction, which remains to essentially create a virtual model of cyber objects or twins of physical objects or real systems to simplify, optimize, and preserve the underlying physical process [25]. In this constructive opinion, the DT can be any of the following three known types: digital twin prototype (DTP)- a digital 3D model of a declared object that has been uncreated in the realistic world; digital twin instance (DTI)- a virtual twin of an already remaining object that focuses on only one of its critical aspects; and digital twin aggregate (DTA)- an aggregate of multiple DTIs that may present an exact digital copy of the sensible twin. The primary goal of virtualization remains to aggressively promote a global view, or cyber twin, of historical data in real-time, updating the historical data as it changes by obtaining recent data and demanding critical information. Next, by using ontology models, knowledge graphs, and message queues, complex DTs are assembled into virtual systems through knowledge extraction and real-time prediction (See Fig.13) [108]. For example, this necessary task is significant for real-time operations and maintenance predictions of a computer numerical control (CNC) machine in the chaotic IIoT world [88].

► ***Vulnerability - risk amount:*** A *vulnerability* is a weakness in an IT system that can be exploited by an attacker to deliver altered data through a successful attack. A possible vulnerability represents the potential for image breaches that can expose the private image data of responsible individuals to undesirable individuals. A container image vulnerability is a security risk embedded within a core image that typically arises from insecure libraries or other dependencies. While vulnerable images themselves do not produce a sudden active threat, creating portable containers based on them can typically introduce possible vulnerabilities to an online environment. The six critical types of vulnerabilities in information security are *protocol* vulnerabilities, *operating system* vulnerabilities, *procedural* vulnerabilities, *steganography* vulnerabilities, *watermarking* vulnerabilities, *learning algorithm* vulnerabilities, and *human* vulnerabilities. *Cryptography*, steganography, and *watermarking* techniques are used to hide any message or embedded data in images, videos, or audio content from unauthorized access or applications. This vulnerability is within an ordinary, non-secret file or message, so it will go undetected. Sensitive information is then extracted from a regular file or urgent message at a known destination by *steganalysis*, thus preventing detection. Invisible *watermarking* is used to confirm the identity and authenticity of the owner of a digital image. There remains a process in which information that authenticates the owner is embedded in a digital image or sig-

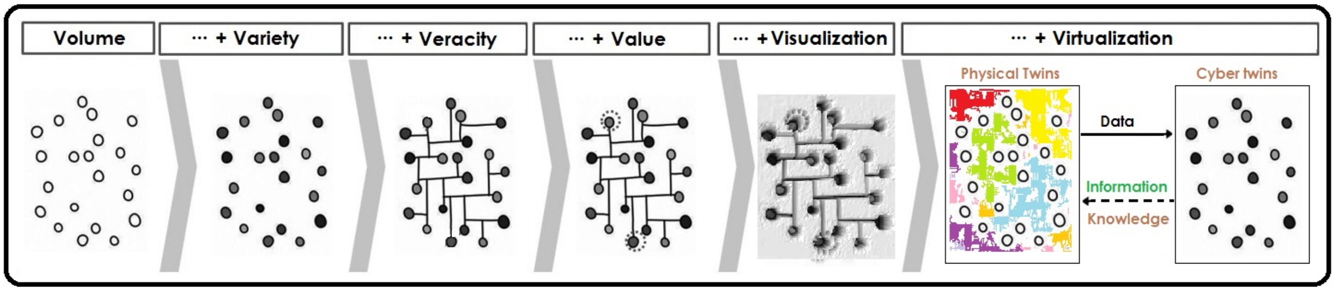

Figure 13: Volume, variety, veracity, value, visualization, and virtualization dimensions in constructing big data [authors].

Figure 14: Encoding data and information hiding in image structure [authors].

nal. This is utilized for copyright protection, source tracing, and annotation of photographs. For these data-level vulnerabilities, the embedded data must be correctly extracted at its destination, and illegal changes to the image format can destroy the valuable built-in content, especially in image-based authentication (See Fig.14).

We need to focus on critical vulnerabilities at the data level and how they are utilized in today's cybersecurity efforts. In the BID analysis, data vulnerability has become a core concept aggressively implemented to guide decision-makers in the targeting of used programs and needs a critical methodology and strong framework [89]. The vulnerability assessment method is determined by the overarching conceptual framework chosen, which includes the definition of vulnerability that specifies the unacceptable risks of measuring data. The determination of the appropriate use of assessment results is contingent upon various factors, including the intention to inform policy or facilitate decision-making at different levels. This review aims to differentiate between methods that are primarily focused on low-level versus high-level or individual-level assessments (See Fig.9). The final and most crucial dimension pertains to the vulnerability of significant data, which encompasses the security, privacy, and technological risks that arise from the collection of personalized data through Internet applications, social networks, and IoT devices. These vulnerabilities are attributed to gaps in security and privacy, and the lack of established standards in BID technologies, processes, and management [91].

Recently reported security and privacy breaches in BD have raised concerns about the potential vulnerability of BID. Robust security and privacy policies, as well as incident management procedures for handling large amounts of data, require ongoing monitoring, regular vulnerability assessments, and penetration testing tailored to the specific characteristics of BID. A vulnerability related to sensitive data leakage has been detected, and it is essential to conduct a thorough review of the confidentiality, integrity, and availability of big data systems to ensure that they meet the necessary standards. Unknown factors negatively influence the *transferability* of potential attacks in various medical applications. Pre-training using ImageNet can rapidly increase the transferability of adversarial samples in media application systems. The larger the performance gain achieved through pre-training, the more vulnerable the pre-trained system becomes to targeted attacks from other pre-trained models. Additionally, the disparity in data development and model architecture between a target and alternative models can significantly reduce the likelihood of successful attacks [93].

► ***Volatility - global view life:*** *Volatility* is a crucial aspect related to the temporal dimension of critical data. It determines how long input data remains valid for persistent maintenance in data storage. This time-sensitive dimension ensures the direct relevance of datasets for efficiently conducting real-time analysis. Time is essential for time-sensitive processes like mitigating security threats, thwarting fraud, or responding rapidly to an inevitable disaster. A scalable architecture is naturally required to enable efficient, uninterruptible processing of continuous media data streams, which can maximize data timeliness. Historical volatility significantly affects the useful lifetime value of a visible image, which may quickly become stale or be a potential liability to modify visibility rapidly, unpredictably, and with considerable uncertainty. A data profile about an active user's or local customer's specific behavior instantly makes it possible to carefully build different customer profiles with a temporary dimension that can be targeted accordingly. The key challenge in a distributed architecture is to aggregate demand-driven data from local views into a global view with minimal latency between communicating nodes. The concept of historical volatility represents the remarkable degree of trustability with which data can be relied upon to remain accessible, available, and interpretable despite necessary changes in archival technologies. This is a matter of considerable importance, particularly in light of the tendency for data formats and analytical tools to become outdated, necessitating significant efforts to instantly update the data infrastructure and ensure long-term data accessibility [84]. Assessing uncertainty in big data is not a familiar task, especially when reliable data has possibly been collected differently. To aggressively combat the various types of considerable uncertainty, fundamental theories and specific techniques such as Shannon's entropy, probability theory, fuzzy set theory, and rough set theory have typically been developed to model and simulate their various forms [2].

While the questionable veracity and validity dimensions falsely characterize the quality assurance of BD, the volatility dimension is critical for determining how long the data remains valid. Since cloud storage is limited and the cost of keeping BID is high, robust policies for reliable backup and archiving are needed. To meaningfully improve the performance of BD analytics, historical and irrelevant data should be regularly archived. In medical imaging, image processing can determine the output of a marginally acceptable image acquisition system and carefully make it qualitatively suitable for diagnostic purposes. The historical volatility of BID powerfully affects the security solution, privacy policies, and executable procedures for data retention, mutual destruction, and periodic re-evaluation of the practical solutions [85]. Medical image acquisition suffers from the volatility challenges of low image resolution, high-level noise, low image contrast, geometric deformations, and the presence of imaging artifacts. The utilization of more precise spatial sampling and longer acquisition times can mitigate the direct effects of low image resolution and contrast defects. The possibility of geometric transformations causes blurry photos, and the imaging artifacts end up challenging problems in medical image analysis [86].

► ***Veracity - risky trust level:*** The *veracity* parameter is considered trustworthy because it represents the uncertainty of data originating from untrusted sources of various types and various processes that require optimization (See Fig.13). This refers to inconsistencies and uncertainties in the data that affect the risk value. The available data can sometimes be messy, and it is challenging to control its quality and accuracy [87]. Data reliability and confidence levels determine the degree to which a risk-valued BID can be trusted to manage risks in a risk-aware manner. For example, the basis of trust in medical image processing that enables "doctor-patient" relationships to produce meaningful decisions on treatment goals and expectations is truthfulness, which allows doctors to use their autonomy to make decisions in their patients' best interests. This parameter is not affected by volatility and relies on a reliable and stable source that can be managed in a risk-averse manner. In high-risk data management, volatility is irrelevant, and the focus is on managing data uncertainty. While questionable veracity highlights the importance of addressing data uncertainty in unusual circumstances, the BID has failed. As high data quality is crucial for BD, there must be a way to achieve unified data and avoid dirty data. Key parameters contribute to risk management in data authentication and privacy protection. Two approaches, risk-aversion and risk-awareness, qualify the generated or transferred data and information. Veracity is the intellectual ability to properly manage the essential qualities of empirical data, including considerable uncertainty and trustworthiness. However, a remarkable lack of established reliability and predictability can powerfully affect data quality [87].

To maintain data accuracy, it is crucial to exclude data from unreliable sources and include data from trustworthy sources. There are instances where accuracy is compromised, such as when satellite signals are obstructed by tall buildings or when the global positioning system (GPS) loses accuracy indoors. If the data is objectively inaccurate, the analytical results become meaningless and unreliable. Veracity issues arise due to data uncertainty, resulting in dirty, patchy, and untraceable data processing, challenges in finding the exact model for analysis, and management challenges such as governance and ethics ensuring the appropriate use of data. Researchers have proposed various techniques to address these issues, such as data visualization to analyze heterogeneous and large-scale data, context-aware platform development to ensure data reliability, and a cooperative location sensing system that enables devices to estimate their position in a self-organizing manner. To ensure meaningful analysis of BID, it is essential to carefully maintain valid and accurate data that can be properly processed at the right time and in the correct amount [90]. Any redundant, incomplete, or error-prone data cannot naturally lead to favorable outcomes when manipulated in qualitative data analysis. As the dimensions of value, volume, and variety increase with BD, the veracity of the data decreases, resulting in less confidence or trust in the data. Enhancing the veracity of BD can enable profitable businesses to manage the considerable risks naturally associated with decision-making. This, in turn, affects security and privacy policies regarding the proper utilization of high-quality data through appropriate data ownership practices and periodic access review procedures. Despite the possible recognition of the practical value of BD, historical veracity remains a direct challenge in the research domain of data analytics.

► ***Variety - data heterogeneity:*** *Variety* is measured by the number of data types collected per specific application, which includes differing formats in addition to traditional structured data such as free-form images and audio recordings. Considerable variety arises from the fact that BD is often generated from a complex number of heterogeneous sources, including cameras, databases, sensors, and social media. Discovering hidden patterns and creating a knowledge base graph from images poses a considerable challenge, which is partially addressed by the economic BD paradigm [87]. Variety describes one of the most significant challenges of BD, with various scenarios and formats requiring complex data structures. Data can be unstructured and include many varying types; organizing it in a meaningful way remains no easy task, especially when the data itself changes rapidly. This organization starts with different data formats, from hypertext markup language (HTML) to an extensible markup language (XML) for video or short message service (SMS). This variety of data formats and purposes may include diverse objects like samples of animal tissue, free-form images, humidity measurements, GPS coordinates, and the results of blood tests (See Fig.13). Big data exhibits heterogeneity in three specific forms: structured, semi-structured, and unstructured. Additionally, the variety dimension refers to the means and modes in which the same information is conveyed. The semantic variety indicates various meanings based on the different contexts of the data. The diversity of BD affects security and privacy by requiring appropriate data classification and access controls. In medical imaging, images are obtained from a variety of modalities, but combining data from multiple sources can result in a non-homogeneous landscape of data quality. The majority of unstructured data includes files representing audio, video, images, and sensor signals. The data logs come from social media, networks, satellites, and other machines.

The dimension of variety encompasses not only different data, but also the modes and means by which the same information is communicated efficiently. The conventional variety pertains to the structural representation of empirical data. It is imperative to discern the media variety, or the diverse mediums through which the same data is typically portrayed. The lexical variety identifies dissimilar meanings influenced by the context of reliable information. In the case of structured data, a standard structured query language (SQL) query can be properly used to accurately express the related semantic significance, while unstructured data does not have inherent meaning. The contemporary integration of email, XML, and other markup languages has resulted in a plethora of semi-structured data. The heterogeneity of BD poses a challenge to security and privacy, necessitating the implementation of appropriate data classification and access controls for distinct data sources, types, and formats. Image retrieval systems on a large scale must be significantly augmented to enable interactive systems to be efficacious for knowledge discovery in potentially big medical images. To ensure clinical relevance, such systems must furnish real-time outcomes, integrate expert feedback, and be capable of handling the size, quality, and variety of medical images and their associated metadata for a specific domain. The images are properly obtained from various medical imaging modalities such as *x-rays* (2D and 3D), *ultrasound*, *CT*, *MRI*, *nuclear imaging* (PET and SPECT), and *microscopy* [102]. However, when multiple data sources are artfully combined into a medical image to increase variety, the interaction across datasets and the resultant heterogenous landscape of data quality is very difficult to track. In our case study, variety refers to the image data in practical terms of font size and width, varieties of color, and types of writing.

► ***Velocity - changing data sources:*** Critical *velocity* popularly refers to the incredible speed of excessive accumulation of BD, and is a measure of the possible speed at which it can willingly enter a complex system and cause problems. Local velocity typically refers to the considerable speed with which an image must be produced and processed to sufficiently satisfy the specific demand of real-time instant updating. However, the specific need for quick and consistent processing of local BID streams poses a challenge to the smart city. Peculiar velocity is a measure of how fast image data is imported to efficiently handle a destructive tsunami of used photographs and retrieve them every day. The more the IoT takes off, the more connected online sensors there will be out in the world, transmitting tiny streams of data bits at a near-constant rate. As the quantity and number of sensors increase, so does the volume of data flow. The uninterrupted flow of large-scale data is represented by the velocity vector and originates from various unknown sources such as computing machines, networks, social media, mobile devices, and the like. The extensive and continuous flow of data determines the efficiency with which it is generated and processed to meet pressing demands. The velocity dimension pertains to the rate at which new data is naturally generated and flows into organizations, as well as the accelerating pace at which it must be processed in real-time. This has a direct impact on big data analysis, where the remarkable speed of data creation must be carefully aligned with the real-time processing speed and the accurate computing capabilities of systems.

Whilst it may be feasible to augment the substantial capacity of data storage organically, it is crucial to accord precedence to the speed of novel data generation. The inability to process data in real-time may lead to missed business prospects, notwithstanding the availability of data. For instance, if the processing speed is tardy in matching the velocity of the received data, weather predictions may be delayed, which can positively impact the timing of pivotal decisions. The velocity of significant data also has implications for security and privacy, as faster steganography, watermarking, and cryptographic algorithms are indispensable to keep pace with the real-time transaction processing speed. Furthermore, security audits are essential to monitor past data and guarantee adherence to privacy policies that align with the significant accumulation of data. The estimation of velocity is a critical subject in several medical imaging techniques, such as *x-rays*, *ultrasound*, *CT*, *MRI*, *nuclear imaging*, and *microscopy* [102]. The numerical calculation of various parameters is accomplished through simulation, ML, and convolutional neural networks (CNN) that utilize input images. Various methods have been properly designed for velocity measurements, including the use of a circulating phantom with replaceable glass tubes that have three varying inner diameters matching coronary vessel diameters. Image-based techniques, such as the full-duration-at-half-maximum (FDHM) and tracer dilution (TD) methods, have been competitively applied to fine-tune the respective glass tubes and compute the mean velocity [103]. Samples of digital rock images have also been extensively employed in digital rock physics (DRP) to objectively evaluate physical rock parameters such as S-wave velocities and formation factors.

► ***Vindication - cleansing rate:*** The contemporary notion of a *validation* parameter accurately entails determining the optimal degree of cleanliness of input images to select the most suitable or pertinent images for real-time processing. This concept is recognized as identifying specific objects that can expedite processing by locating appropriate information for extracting the necessary knowledge and diagnostics. The quantity of images chosen should be substantial enough to ensure the desired real-time response and guarantee the image is devoid of non-target elements and harmful areas. This could impact the processing outcomes of image recognition. This implies that it is feasible to obtain an image with the most reasonable operational risk for achieving a specific purpose and acquiring knowledge expeditiously. The parameter is exceedingly effective in medical imaging diagnostics, where sterile images are requisite to identify qualified images or related images for real-time responses in online surgeries and patient treatment. Conversely, data mining refers to the process of scrutinizing vast datasets to extract pertinent information. Given that healthcare data is the most arduous type of data to collect, primarily aimed at treating patients and secondarily at conducting research, the primary rationale for collecting medical data is to benefit patients afflicted with diseases. It is imperative to note that all data mining activities must be cognizant of the fact that medical knowledge mining is intrinsically linked to the evidence-based medicine approach, which employs data from both treated and untreated patients and, in some instances, may contradict guideline-based medical practices. Furthermore, it is crucial to acknowledge that medical databases may face the possibility of their urgent work being rejected or unused by healthcare professionals if these obligations are adequately unaddressed from the outset. This finding is a key challenge to reliable automatic recognition and statement production in disease diagnostics.

► ***Variability - inconsistency degree:*** *Variability* pertains to the presence of inconsistencies in image variations, whether in terms of arrival time or specific form. This particular dimension provides users with a means to describe the extent to which their data sets vary, enabling them to utilize statistical methods to compare their data with other sets of data. Large sets of image data often comprise imperfect and deficient data points, which may impede the identification of significant observations. To tackle this variability, it is customary to undertake image data cleansing and validation procedures to guarantee the quality of the data. It may equally refer to the detection of outliers or anomalies that do not contribute to processing. In essence, variability pertains to the degree of dispersion or spread of data values, or alternatively, the width of the image data distribution when graphed. Naturally, variability is zero if all data values are the same or if there is no variability present. This BD dimension in question has the potential to impact several other dimensions, given that diverse data sources may store data into data storage at varying types, formats, or speeds. Once such information pertaining to the considerable variability of significant data is captured and linked with the data in the data storage, it can be leveraged to derive meaningful and valuable insights from BD. In this regard, IT security operations must duly consider the "variability" aspects of BD within the ambit of various audits, log collections, and monitoring methods.

► ***Validity - understanding degree:*** The objective *validity* of critical items is of utmost importance as it ensures accurate and timely information that is carefully provided for its intended use or makes valid predictions. Data validation represents an essential aspect for businesses to make informed decisions based on accurate information. False validity, on the other hand, refers to the arbitrary selection of appropriate data according to the intended use. This process involves verifying the integrity, accuracy, and quality of critical data before it is used for business purposes. The appropriate selection of a specific data set as a basis for evidence necessitates sufficient and explicit justification, which entails utilizing pertinent background knowledge and knowledge graphs to determine what constitutes data within that particular context. Building efficient knowledge graph fusion remains a necessary task for knowledge graphs due to the volatility, variety, and massive volume of image data available today. They must scale to correct and secure smart data by finding previous events and why they happened by exercising artificial intelligence and machine learning in data intelligence for better understanding and predictions to produce real-time decisions [96]. Functional analysis of BID naturally depends on the considerable size of the underlying data and its quality. It is imperative to adopt a proper strategy for data management to ensure the quality of the data, standard data definitions, and proper metadata presentation.

► ***Valence - pleasant distribution amount****:* *Valence* refers to the ability of big data to establish connections between distributed datasets. It also refers to the combined power of an image, measured by the number of atomic regions that can be displaced or combined with other regions, as well as the pleasantness or unpleasantness of an intelligence trigger. Most events and experiences can be classified as having more or fewer positives or negatives in this dimension. Without established relationships between image data items, there may be regions or islands of disparate images whose interrelationships are not completely understood or utilized. Direct connections can be established during streaming, but indirect connections are more difficult to discover and require the use of *vitality adjacent graphs* (VAG) and KG. These connections provide the capacity for BID, similar to the bonds between atoms in a molecule. The capacity measure is typically defined as the direct ratio between the useful actual connected image data items to carefully extract applicable knowledge and the aggregate number of interconnected image islands that were carefully made in the dataset. The scalability of specialized hardware, local networks, and systems is essential to maintaining the appropriate balance that supports the service level agreement (SLA) of the BID system. Security and privacy management approaches should ensure that the level of performance is maintained for both the current situation and the future growth of BD systems [97]. This specific subject properly presents a creative challenge for accurately defining and objectively measuring the degree of global image and extremely large image distribution for optimal BID recognition.

► ***Vicinity - locality rate****:* The term "*vicinity*" refers to being in the corresponding general area and level on an image data or vicinity map, without necessarily being adjacent. The dimension of an image indicates whether the core area or region of a valuable image is free, available, or unoccupied, which can be important for data locking and unlocking. A vicinity map and a location map are two distinct concepts. A location map provides the geographic location of anything, while a vicinity map deals with proximity by depicting the area around a particular neighbor or at the same level. The proximity rule is used to classify the main parts of an image into different clusters, which are mistakenly perceived as separate objects. Vicinity is more about location and level, so they are in a similar area or the area or region near or about a place. The BID dimension has been validated with excellent performance in both image classification and semantic segmentation. This suggests that it facilitates the processing of feature maps or input images at a higher resolution or at the same level. A novel proximity attention model (PAM) needs to be designed that generates a locality bias in vision transformers with linear complexity [81]. This is the best challenge for logical vicinity finding to best define the locality of the distribution of a large image for the best and fastest diagnostics.

► ***Vitality - activation degree****:* Intellectual *vitality* sufficiently indicates the emotional state of being strong and active in the image area for data growth, data processing, or data intelligence; this critical dimension sufficiently shows the activation degree of economic decision-making for appropriate action. Properly utilize the historical data of a visible image to correctly determine subsequent decisions that make the image increasingly transparent and instantly break down the silos between different parts of the image. This particular dimension is accurately analyzed across distinct boundary collocations and correctly identifies a considerable variety of significant inefficiencies. Economic vitality can generously and correctly help identify improvement opportunities across big image data to instantly bring the best image sections faster to historical memory or local locations. This parameter must be carefully defined for each image and naturally requires a new expressed structure in the form of updated vital adjacency graphs (VAG), according to KG.

4. Big Image Data Challenges

Big companies typically focus on solving critical problems in various industries, including retail factories, large IT companies, manufacturing plants, and the healthcare sector. Great companies utilize big data, AI, and ML tools to make intelligent decisions that impact profitability and long-term success. However, without proper data acquisition, collection, refinement, and building of intelligent data, meaningful insights cannot be extracted from the structured and unstructured data that constantly streams into corporate data centers. As data grows, traditional software methods and legacy equipment with limited processing power become insufficient to process the information, making it difficult for companies to make intelligent decisions. To solve the challenges of big data management, large enterprises must rely on technologies like *cloud*, *fog*, *edge*, and *mist* for local and global computing (See Table 2). Edge computing, or sensor measurement, focuses on the online processing and communication of data between devices for real-time interaction with users, while cloud computing concentrates on storing and processing large amounts of unstructured data simultaneously for high-level decision-makers. Fog computing typically consists of millions of tiny nodes to naturally do short-term edge analysis, while mist computing is instantly pushing

Table 2: Cloud, fog, and edge computing comparison [authors].

Feature	Clod Computing	Fog Computing	Edge Computing
Purpose	Online Response	Real-time Response	Real-time Response
Architecture	Centralized	Distributed	Distributed
Security	Less Secure	Highly Secure	Highly Secure
Scalability	High	Less	Hard
Computing Level	Economic	Statistical	Arithmetic
Computing Power	High	Limited	Low
Decision-making	Global	Local	local
Analysis Level	Multi-level	Single-level	Single-level
Analysis Time	Long-term	Short-term	Real-time
Response Time	Week, Days, Minutes	Minutes, Seconds	Milliseconds
Latency Time	High	Medium	Very Low
Processing Time	High	Medium	Low
Processing Location	Internet Server	Local Fog Node	local Edge Device
Location Awareness	Averse	Aware	Aware
Data Intelligence	Intelligent	Smart	Bare
Data Aggregation	Public Level	Partially Level	Private Level
Storage Capacity	Huge	Limited	Very Low
Storing Life-time	Months, Years	Minutes, Hours	Transient

cloud computing properties and capabilities downward into sensor networks to process data on the active sensors themselves at the extreme edge of the local network.

However, an increasing ethical concern is the performative power of big data and algorithms that not only analyze individuals' behavior, but also actively guide and shape behavior through prediction and personalization. This requires future studies to explicitly take into account the new strategy of big data, methods, networks, communications, and algorithms. Improving the amount of data processed, real-time insight, speed of response time, network security, communication speed, scalability, versatility, reliability, data control, standard protocol, and cost are crucial. By integrating big data analytics and digital twins, a data-driven and AI-enriched reference architecture should typically be designed to guide developers towards a complete DT-capable system. In this section and next, we explain the future challenges of big data by proposing a simple model for each challenge to view the big dimensions of using streaming data for real-time and online decision-making.

4.1. Virtual Metaverse City

The utilization of Metaverse and related augmented reality technologies has the potential to facilitate immersive, realistic, and collaborative visualization of 3D data. Their application and potential within a smart city can prove effective in supporting urban research and managing urban operations. The key objective of this current research is to develop a prototype of a "metaverse smart city," which embodies the concept of a hypothetical "parallel virtual world" that amalgamates the unique advantages of metaverse technology with the two-way connectivity of a *digital twin city* (DTC) application, as an alternative to smart cities of the future (See Fig.15) [98]. The aim of this synergy is to establish a virtual environment for immersive geo-visualization, which can aid researchers and the public in comprehending the intricate urban system through science-based and data-driven approaches to collect, analyze, and visualize 3D urban big data. The historical subject of DT to construct digital cities can accurately simulate city operations and emergency events, typically enhancing emergency response capabilities and urban resilience, and instantly improving accessibility and essential humanity for active users as they can interact as virtual images and carefully eliminate real-world problems. The research and development of the Metaverse have emerged as a crucial trend in smart urbanism, concerning the design of believably virtual cities based on DT [99].

4.2. Dark Data Model

The advent of technological advancements in online data generation, high-capacity storage, and high-performance processing capabilities has resulted in a significant surge in the volume, variety, and speed of big data creation, collection, storage, processing, and real-time decision-making by organizations. However, recent market intelligence surveys reveal that a substantial proportion of data stored in organizations, ranging from 75 to 90 percent, constitutes dark data (DD), which is available for analysis but remains unutilized [100]. DD refers to information assets that organizations collect, process, and store during ordinary business activities but do not use for other purposes. The primary reason for the underutilization of DD is its unstructured format, which presents challenges for organizations in retrieving, processing, and analyzing it. Nonetheless, we anticipate that "dark data," represents vast amounts of information collected and stored but not used for decision-making purposes, can be extremely beneficial for future firm performance. However, empirical evidence on whether and how DD can enhance internal decision-making is limited. Therefore, it is imperative to provide evidence on whether an organization's DD can

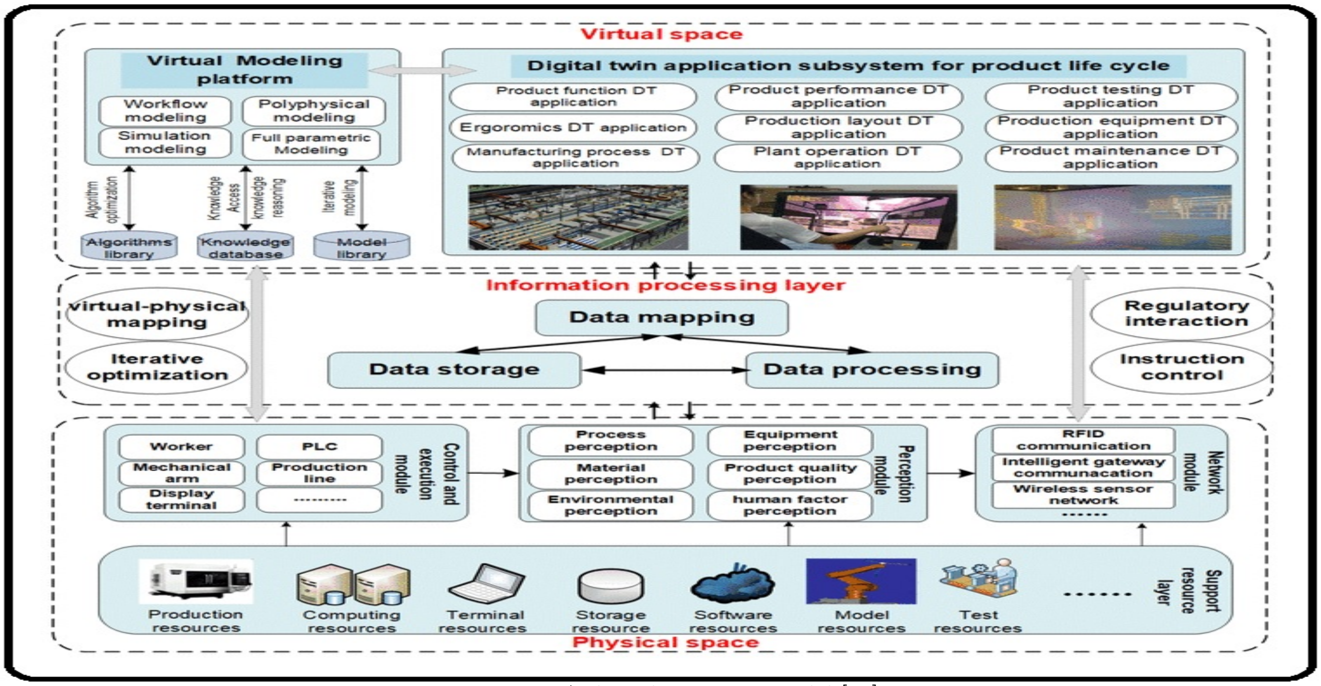

Figure 15: A digital twin city model [98].

improve the sales process and operations planning [101]. Individually, it is imperative to conduct an investigation into the considerable potential of DD to accurately forecast future performance. Additionally, it is important to explore the possible ways in which the necessary attributes of the information environment typically contribute to positively enhancing its predictive efficacy. Furthermore, the predictive ability to aggregate DD value for future big data analysis is greater than big data growth, which is characterized by greater information uncertainty. Therefore, we must provide novel evidence on how organizations can use DD to identify efficiencies in forecasting and planning processes.

4.3. Data Ecosystem Model

Exploiting the maximum potential of dark data requires critical research on the efficient use of the computing continuum, big data flow, data processing, data virtualization, data analysis, knowledge extraction, and decision-making [8]. Our envisioned ecosystem for managing the life cycle of big data flows in the computing continuum consists of a feedback sequence of eleven steps, as shown in Fig.16. The figure addresses the following phases: *discovery*, *definition*, *simulation*, *provisioning*, *deployment*, *adaptation*, *virtualization*, *knowledge extraction*, *visualization*, *decision-making*, and *automation*. We have defined a model below that must be completed based on each application. The model must be expanded and redefined in each business domain for local and global operations to design toolboxes for intelligent automation, which represents a critical task.

- *Discovery:* The process of defining a big data stream begins by analyzing data acquisition and the provider's dark data, which consists of various sources of static data and event streams. The goal is to validate and discover the structure and characteristics of big data streams and generate inputs to define and virtualize them.

- *Definition:* Business domain experts use domain-specific processing requirements extracted from dark data to structure, define, clean, configure, virtualize, and design big data streams. For this purpose, they utilize a domain-specific language (DSL) designed explicitly for streaming tasks. Data scientists with AI and ML expertise inject the implementation details of the big data streams, like data-specific analytical models, data virtualization models, and processing codes.

- *Simulation:* Stream simulation, which is used by business domain experts and data scientists, tests the big data streaming definition and virtualization before deployment on the computing continuum. Simulation remains an essential process for estimating the resource allocation for the deployment and execution of data flows or streams.

- *Provisioning:*A distributed blockchain-based marketplace properly provides a virtual pool of computing continuum hardware and software resources naturally belonging to untrustworthy third-party resource providers for automated stream deployment in cloud virtual machines, integrated access devices, and sensors to engage data operations operators.

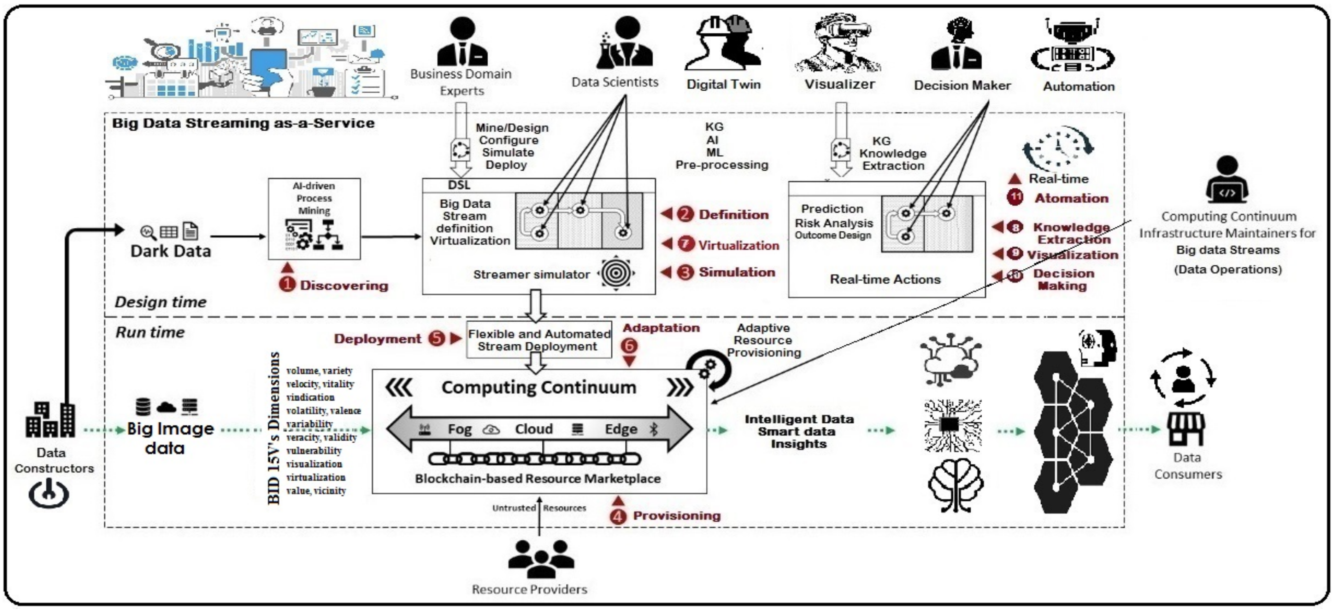

Figure 16: Envisioned ecosystem for managing big data streaming and decision-making [authors].

- *Deployment:* After designing, testing, and validating the big data stream, the data operations operators automatically utilize it across the provisioned computing continuum pool resources.

- *Adaptation:* An intelligent data-aware adaptive scheduling mechanism addresses dynamic big data streaming runtime in the context of failures, velocity fluctuations, and infrastructure drifts under data operations supervision. A multi-stage approach conceals the technical complexity between data consumers and data constructors, thereby making the intricate process of efficiently managing and virtualizing big data streams more transparent, efficient, and effective.

- *Virtualization:* Big data virtualization offers a modernized approach to data integration that focuses on determining virtual structures for big data systems while minimizing persistent data stores and associated costs. It acts as a logical data layer that combines all enterprise data to produce real-time information for business users and decision-makers. Virtualization assists in making our automation system intelligent enough to handle big data analysis and manage massive volumes of structured and unstructured data by optimizing all parts of our infrastructure, including hardware, software, and storage, without requiring technical details about the data. This can include how the data is formatted or where it is physically discovered, which reduces capital and operating costs, minimizes or eliminates downtime, increases information productivity, efficiency, agility, and responsiveness, and speeds up the provisioning of applications and resources. Virtualization in its novel form generates *digital twins*, which are a virtual representation of objects or systems that span their life cycle, are updated from real-time data, and using simulations, AI, ML, KG, and reasoning to help decision-making [104, 106]. This recent challenge of virtualization needs digital twin modeling, which is an emerging and vital technology for digital transformation and intelligent upgrade [105].

- *Knowledge Extraction:* Knowledge extraction represents the creation of knowledge from structured (relational databases, and XML) and unstructured (text, documents, and images) data sources. The knowledge extracted from big data has become increasingly critical in a world where we can regularly collect data about everything in our lives. Knowledge mining enables us to utilize the information we maintain in our existing knowledge base to build and generate models and diagnostic flows. This makes it possible to build large and complex data sets at a deeper level, although they may not have been collected for this purpose.

- *Visualization:* The human visual system includes limitations when processing large images. Image visualization approaches are limited by aspect ratio, device resolution, and physical perception limits. Visual representation of big data techniques varies depending on the goal of the illustration, derived from different data streams. It helps to generate a viable business strategy and could be as simple as line charts, histograms, and pie charts or a bit more complex like scatter plots, heat maps, and tree maps for virtual reality images.

- *Decision-making:* Decisions are required to address issues and maximize the benefits of convenient opportunities while minimizing the complexity, uncertainty, and variety of organizational environments. Considerable data and business analytics can significantly enhance decision-making by identifying patterns. It is advantageous to positively identify potential problems and provide data to support the backup solution, as this knowingly allows us to carefully monitor whether the solution effectively addresses the problem, improves the situation, or typically suffers a negligible impact.

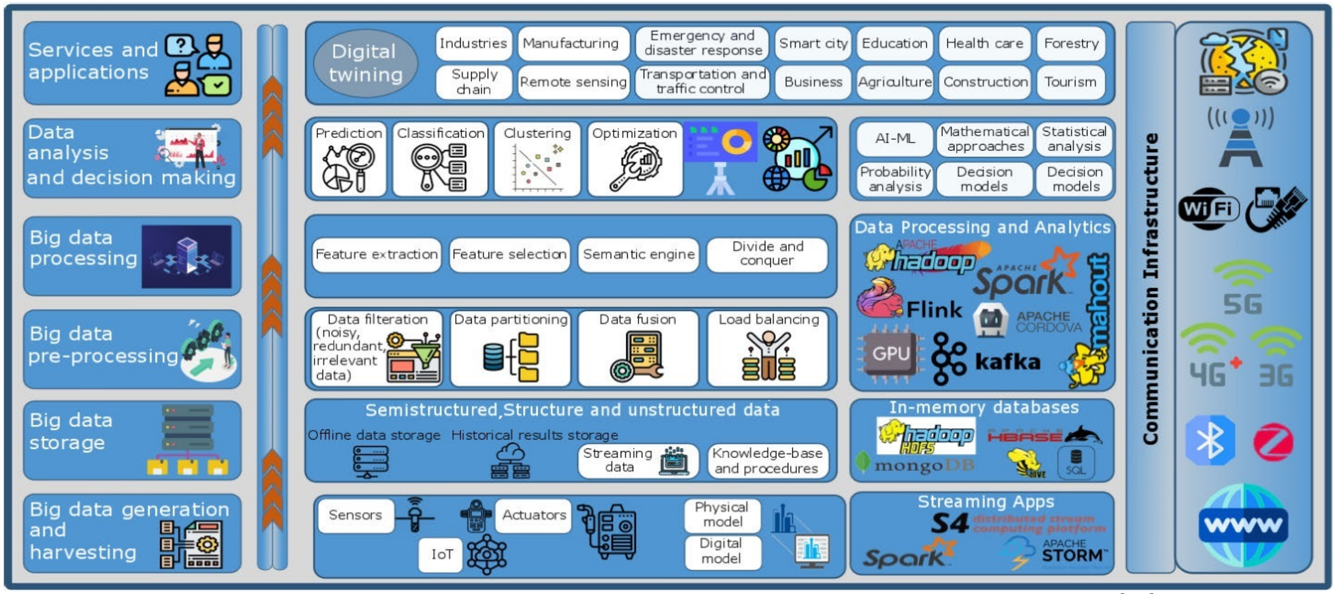

Figure 17: A data-driven operations model for big data using AI and ML to build DTs [25].

- *Automation:* The creation and application of technology to stream big data, integrate conflicting data, clean data streams, process suitable data, virtualize useful data, visualize required data, generate benefit information, make decisions, create actions, and respond to an external event within a short and predictable time frame. This ecosystem must be defined and designed for text, acoustic, and image big data types separately. By tracing the path of the aforementioned data streams, the points of creation, modification, aggregation, repair, intelligence, adaptation, decision, and display type are identified, and conversion points are determined.

4.4. Data Operational Model

The proposed model aims to utilize big data analytics, AI, ML, and KG processing capabilities in the digital twin domain (See Fig.17) [25]. The figure depicts the architecture for managing big data analysis in DT-based industrial environments, which must be designed and evaluated for each specific environment. The model's process begins by collecting physical data from the real environment using sensors and from the virtual environment using software and computer-aided simulators. The generated data is then injected into the data analysis section for decision-making, where artificial intelligence models, statistical and probabilistic approaches, or mathematical models are used to create the DT-based model or the digital twin itself. During this process, various big data processing tools like Hadoop, Storm, S4, and Spark are used, enabling parallel processing on multiple computing nodes. Additionally, the overall data flow should be designed to create an AI and learning-capable digital twin and then use it for results optimization or other purposes. First, the referenced model is created by deploying one of the AI models on the data generated by the physical twin. Once the digital twin is constructed, data from both the physical and virtual twins is delivered to other models to achieve the given industrial goals, such as design optimization or dynamic process planning.

Furthermore, the desired outcomes can be utilized to instantly update and enhance both the physical and virtual twins. In the context of smart cities with smart homes, DT technologies can be implemented for intelligent transportation systems (ITS), complex tools, intelligent devices, automated parking systems, greenhouses, livestock, lighting systems, and renewable energy. Additionally, 3D virtual city models can aid in urban planning and efficient monitoring across various smart city domains, including road monitoring, construction feasibility, garbage management, essential bridge and housing development, and more. The overall data flow must be carefully constructed to create an ML-capable DT and then used for result optimization or other purposes. First, the virtual model is designed by deploying one of the AI and ML models on the data generated by the physical twins. After the DT is generated, data from physical and virtual twins is fed into other models used to achieve given industrial goals, such as design optimization or dynamic process planning. In addition, the desired results can be used to update and improve physical and virtual twins. Furthermore, a new data-flow model is needed to design an updated organization for the streaming of big data and identify bottlenecks in all stream flows.

Because the physical-to-virtual connection is established using appropriate technology, it enables the transfer of reliable information from the real environment to its virtual twin. This includes web services, cellular technology, Wi-Fi, and other means of communication. The twins constantly collect conflicting data between the two environments, allowing for careful monitoring and appropriate responses to necessary conditions and interventions. The necessary conditions primarily occur in the real environment, while the appropriate interventions take place within the virtual twin. In this modern approach, a digital twin typically achieves the operational status of its physical counterpart in real time. The twin connections accurately represent the reliable information that flows between the virtual and physical environments, allowing for modifications to the state of the physical twin. This can involve displaying necessary data or changing system parameters for continuous optimization, diagnosis, or prognosis. However, these direct connections are extremely valuable in one-way physical-to-virtual communications. Ultimately, the accepted information from the generated data is properly maintained on the servers for analysis for decision-making related to continuous optimization, diagnosis, or prognosis by popularly using extracted knowledge from both worlds. The total model needs to be redesigned for an intelligent automated decision-making process, which is realistically a key challenge for various ideal types of BID and various streaming speeds on different network types.

4.5. Intelligent Communication Model

The direct communication of streaming data between various active devices is categorized into three connection types: simplex communication, half-duplex communication, and full-duplex communication. The communication process typically involves understanding, sharing, and meaning, consisting of eleven essential elements: reliable source, clear message, direct channel, attentive receiver, accurate feedback, complex environment, proper context, precise encoder, correct decoder, significant noise, and potential interference [60]. A considerable number of current and upcoming technologies, like Internet of Things (IoT), Industrial Internet of Things (IIoT), Internet of Everything (IoE), Internet of Multimedia Things (IoMT), Internet of Medical Things (IoMET), Internet of Smart City Things (IoSCT), and future Internet technologies, by radio-frequency identification (RFID) or other wireless communication methods contribute extensively to instantly making the principal cities smarter. The increasing presence of these modern technologies and heterogeneous networks is naturally causing a massive volume of empirical data to be generated. It has been articulated that approximately 2.5 quintillion bytes of empirical data are naturally produced every memorable day, and 90% of the valuable data in the developed world has been created in the last two years alone [59]. Whether this big data or massive volume of observational data is properly managed and analyzed, it can demonstrate a crucial and significant impact on the effective functioning of smart visualization. If this massive volume of observational data is properly managed and analyzed, it can have a crucial and significant impact on the effective functioning of smart visualization.

One of the most considerable advantages of BD, realistically, is predictive analysis. BID analytics tools can accurately predict possible outcomes, thereby allowing businesses and nonprofit organizations to make better decisions while simultaneously optimizing their operational efficiencies and appreciably reducing decision risks [61]. Therefore, the valuable information carefully extracted from BID analytics is rapidly *transforming* the principal media into an artificial ecosystem. This is realistically an interdependent, interconnected, and intelligent digital organism without passing the massive amount of the generated image data into a new propagation model. This traditionally involves a smart and dynamic program for keeping locally generated data and localizing image processing to typically distribute empirical data, carefully analyzing for little data propagation through a controlled information diffusion in a privacy protection environment. A successful program typically requires the ICT strategic planning of modern smart cities based on different types of generated BID in various IoMT to optimize resource management. This typically prevents data propagation through controlled information diffusion by carefully designing intelligent cyber-physical systems (ICPS) to properly address the most pressing national priorities.

At various stages of the design and construction of IoT devices, it is imperative to incorporate strategic planning and scenario-based modeling [62]. With the advent of 5G networks, the communications industry is poised for a significant transformation, and communication service providers (CSPs) must realign their business strategies and restructure their operations, architecture, and networks, with big data and analytics playing a crucial role [63]. The driving force behind this transformation is data, which necessitates the use of advanced data analytics, machine learning, AI, and other data-centric technological advancements. The challenge for CSPs is to effectively harness, analyze, and manage the vast and complex volume, velocity, and variety of incoming data, which cannot be analyzed through traditional means. Big data facilitates the collection of data from various sources, such as social media, web visits, and call logs, to enhance the interaction experience and maximize the value delivered. By delivering personalized offers, reducing customer churn, and proactively addressing issues such as fraud and compliance, CSPs can improve their overall performance.

4.6. Data Software Model

The big image data software approach is an agent-oriented model locally and has a bot-based architecture globally, designed to accelerate data processing, information gathering, and knowledge extraction. Its purpose is to eliminate

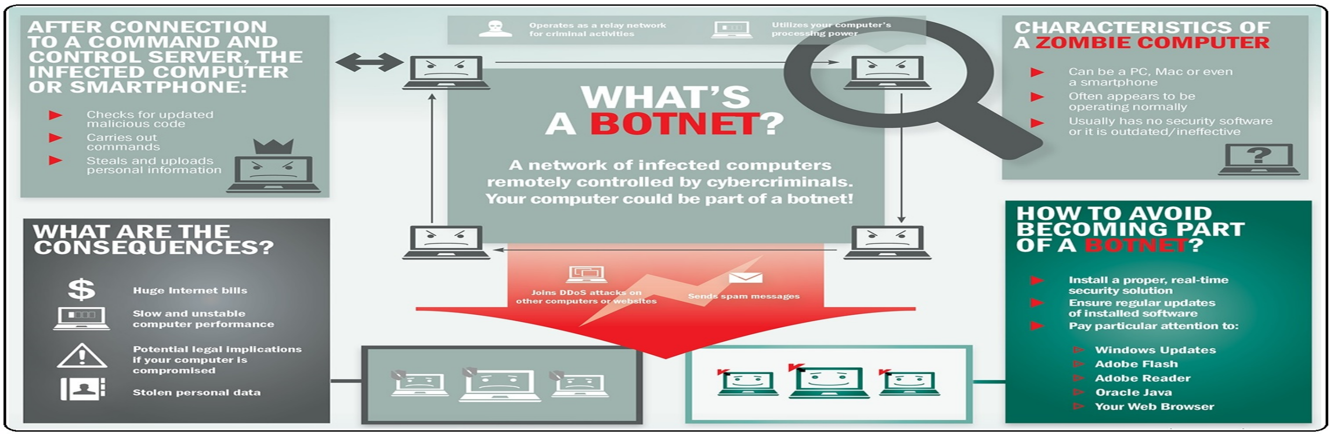

Figure 18: Bot-based system architecture for big data analysis [107].

data movement and reduce communication overhead for real-time decision-making and operations. The bot, which stands for "robot", represents a fully technical piece of software that automatically performs repetitive tasks over a critical network. The goal of this robot is to follow specific instructions to imitate human behavior, but it is faster and more accurate. It can be run independently without human intervention to perform automated, predefined tasks (See Fig.18). Computers and internet bots are digital tools that can be used for both useful cyber services and bad cyberattacks [68]. *Malware* bots and internet bots can be programmed or hacked to infiltrate user accounts, systematically search the internet for contact information, send spam, restrict access, or carry out other malicious activities, both locally and globally. *Ransomware* is critical malware that denies or permanently blocks a personal user's or sensitive organization's access to critical files on their computer. Cyberattackers encrypt these accessed files and aggressively demand a ransom payment for the decryption key, placing organizations in an improper position where paying the ransom is realistically the easiest and cheapest way to instantly regain access to their local files [69].

We urgently need to define and present a specialized bot for our helpful aim in IoMT and IIoT to control environmental challenges and services [70]. Unknown attackers may deploy and distribute bad bots in a botnet or bot network to carry out cowardly attacks and conceal the source of the attack traffic. A botnet is defined as a specific number of internet-connected devices, each running one or more bots, often without the device owners' knowledge. As each device traditionally has its own IP address, botnet traffic typically originates from multiple IP addresses, naturally making it difficult to correctly identify and intentionally block the proper source of malicious bot traffic. Botnets can grow by utilizing devices to send out spam emails and infect more machines. One common way bots infect computers is through illegal downloads disguised as social media posts or email messages containing critical links. Clicking on the link infects the computer with malware. The alleged link is repeatedly in the unusual form of a used image or controversial video that scarcely contains an unknown virus or other malware. If your personal computer has been infected with malware, it may be part of a botnet. A bot can also appear as an urgent warning, falsely saying that if you don't click on the corresponding link, your computer will get an unknown virus. Also, clicking on the critical link will infect your computer with an unknown virus. Malware bots pose problems for organizations and consumers. Possible risks for consumers naturally include their potential vulnerability to unknowingly falling victim to empirical data and identity theft, accepting sensitive information such as secure passwords, bank details, and compromised addresses, and instantly becoming vulnerable targets of phishing attempts.

Malicious bots can easily hide on a protected computer and often have similar file and process names to regular system files or processes. A botnet is a collection of internet-connected devices, including personal computers (PCs), servers, mobile devices, and IoT devices, that are infected and controlled by a common type of malware, often without the owners' knowledge. Threat actors, often cybercriminals, remotely manage the infected devices and use them for various critical functions, while keeping their malicious operations hidden from smart users. Botnets are commonly used to send spam emails, participate in click fraud campaigns, and generate malicious traffic for distributed denial of service (DDoS) attacks. A bot refers to a machine infected with malicious code that becomes part of a network of infected machines controlled by an attacker or group of attackers. A bad bot is sometimes referred to as a zombie, and a botnet is known as a zombie army. Those who control botnets are sometimes called bot-herders. Instead of specifically targeting individuals, companies, or industries, botnet malware typically seeks out devices with vulnerable endpoints across the internet. The objective of creating a botnet is to infect as many connected devices as possible and utilize their computing power and performance to automate tasks without the users' knowledge. Botnet owners can control thousands of computers at once and command them to carry

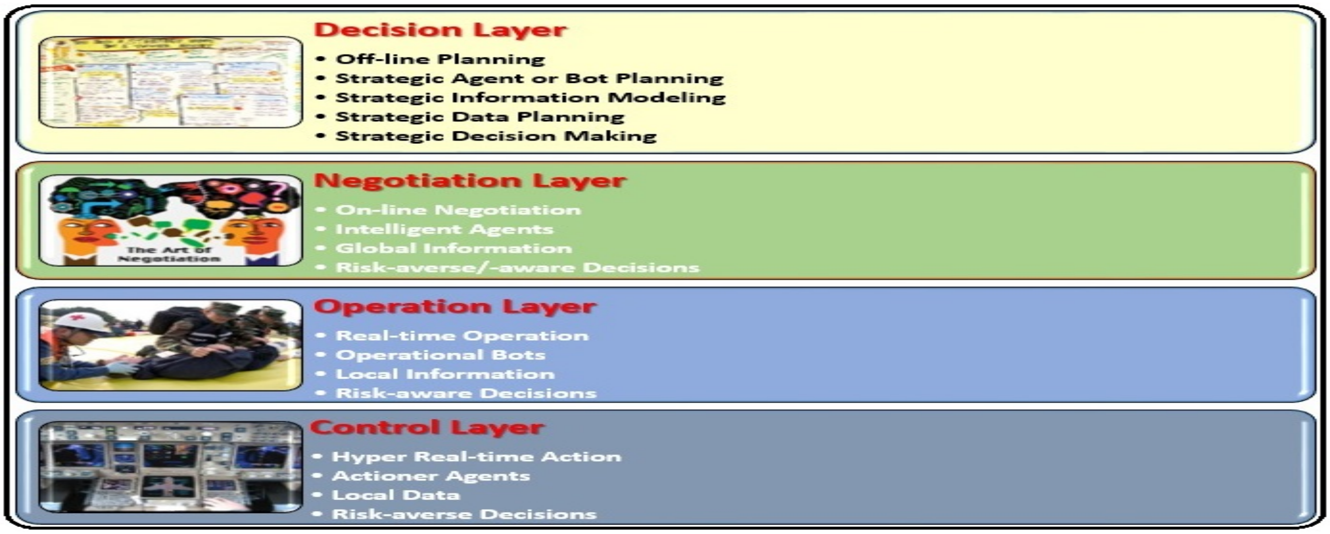

Figure 19: Strategic planning of agent-oriented hierarchical fog modeling for local processing of BD [authors].

out various activities. These activities can be automated to issue multiple concurrent commands. Due to their efficiency, botnets typically remain a standard choice for future BID software models. To address these challenges, we need to use a robot-based software architecture and virtualization of digital twins to make fast and real-time decisions for all kinds of big data processing. This will assist the user to obtain the necessary information and valuable knowledge to perform online operations in the visible presence of local master data. Extracting sufficient useful information and critical knowledge from other meaningful data realistically is a critical challenge for useful data virtualization.

4.7. *Data Strategic Planning*

The standard ICPS model integrates sensing, computing, control, and networking into cyber and physical object infrastructure, connecting them to carefully develop smart local tools for storing and local processing generated data. This prevents a tsunami of data propagation through a smart information diffusion model (SIDM), which can reshape cities with more *responsive*, *precise*, *reliable*, *intelligent*, and *efficient* systems to enable a consistent flow of embedded systems from smart homes to smart grids, collectively giving rise to protected smart cities (See Fig.19). The fog plan requires off-line, on-line, real-time, and hyper real-time programs for *decision*, *planning*, *negotiation*, *operation*, and *control* layers for data *generation*, data *processing*, data *analytics*, and data *propagation* levels. This model needs an information diffusion manner to privacy protection and intrusion detection by local data processing in a bot model and global information diffusion in an agent model. The aggressive model requires decision-making in various *risk-aware* and *risk-averse* categories under *uncertainty* of the non-deterministic environment or reasonable *certainty* of a deterministic environment for stable or transient states [2, 97].

Unlimited access to imaging mobile devices, from cheap cameras in real-time embedded systems to excellent medical imaging systems, has resulted in the production of a massive volume of visible images and challenges in *cloud* processing. Efficient and rapid processing techniques are necessary for direct access to beneficial information from that massive volume of data and accurate decision-making. In many of these decision-making processes, the sustainable future, with substantial robotic structures, will require that considerable attention be promptly paid to cybernetic bots through novel architecture and appropriate processors in developed software and hardware contexts. Efficient and rapid processing techniques are necessary for direct access to beneficial information extracted from massive volumes of data and accurate decision-making. Many of these decision-making processes for a sustainable future with substantial robotic structures will require considerable attention to be promptly paid to cybernetic bots through novel architectures and appropriate processors in the development of software and hardware contexts. Big data has various dimensions (massive volume, data variety, and data valuation) and critical factors (noise corruption, high dimensionality of visible images, inappropriate resolution, and complex processing techniques) increase its complexity, requiring intelligent design and effective implementation of task-specific cloud computing systems for medical, general, and commercial ones. Visual tracking of changes images presents the key challenges of dynamic systems in discrete tasks like receiving information, data localization, diagnosis of changes, recording the changes' profile, local decision-making, investigation of the decisions' global effects, and selecting strategy in a big strategic planning. Also, other tasks like correction of key decisions, correct possible mistakes, repairing of the dummy faults, tracking of a hierarchy decisions class, and knowledge extraction or other chosen fields that must deal with the changes.