Article

Not peer-reviewed version

Towards Developing Big Data Analytics for Machining Decision-Making

Angkush Kumar Ghosh , Saman Fattahi , Sharifu Ura *

Posted Date: 18 July 2023

doi: 10.20944/preprints202307.1199.v1

Keywords: smart manufacturing; big data; manufacturing process; big data analytics; decision-making; uncertainty

Article

Towards Developing Big Data Analytics for Machining Decision-Making

Angkush Kumar Ghosh [1], Saman Fattahi [2] and Sharifu Ura [1,*]

[1] Division of Mechanical and Electrical Engineering, Kitami Institute of Technology, 165 Koen-cho, Kitami 090-8507, Japan; ghosh-ak@mail.kitami-it.ac.jp (A.K.G.); ullah@mail.kitami-it.ac.jp (S.U.)
[2] Advanced Manufacturing Engineering Laboratory, Kitami Institute of Technology, 165 Koen-cho, Kitami 090-8507, Japan; saman.faattahi@gmail.com (S.F.)
* Correspondence: ullah@mail.kitami-it.ac.jp; Tel.: +81-157-26-9207

Abstract: This paper presents a systematic approach to developing big data analytics for manufacturing process-relevant decision-making activities from the perspective of smart manufacturing. The proposed analytics consists of five integrated system components: 1) data preparation system, 2) data exploration system, 3) data visualization system, 4) data analysis system, and 5) knowledge extraction system. The functional requirements of the integrated systems are elucidated. In addition, JAVA™- and spreadsheet-based systems are developed to realize the proposed integrated system components. Finally, the efficacy of the analytics is demonstrated using a case study where the goal is to determine the optimal material removal conditions of a dry electrical discharge machining operation. The analytics identified the variables (among voltage, current, pulse-off time, gas pressure, and rotational speed) that effectively maximize the material removal rate. It also identified the variables that do not contribute to the optimization process. The analytics also quantified the underlying uncertainty. In synopsis, the proposed approach results in transparent, big-data-inequality-free, and less resource-dependent data analytics, which is desirable for small and medium enterprises—the actual sites where machining is carried out.

Keywords: smart manufacturing; big data; manufacturing process; big data analytics; decision-making; uncertainty

1. Introduction

The concept of big data (BD), introduced in the 1990s [1], typically refers to a huge information silo consisting of a vast number of datasets distributed in horizontally networked databases. This concept enriches many sectors, including healthcare [2], banking [3], media and entertainment [4], education [5], and transportation [6]. The same argument is valid for the manufacturing sector, as described in Section 2. Approximately 3 Exabytes (EB) of data existed globally in 1986. By 2011, over 300 EB data were stored in a financial econometric context. Remarkably, it reached more than 1000 EB annually in 2015, and it is expected that the world will produce and consume 94 zettabytes (94000 EB) of data in 2022 [7–9]. Besides its volume, BD is often characterized by a set of Vs, e.g., velocity, variety, volatility, veracity, validity, value, and alike. The rapid speed by which datasets are accumulated in BD determines its velocity. The multiplicity in the contents (text, video, and graphics) and structures (structured, unstructured, and semi-structured) means the variety of BD. As far as variety is concerned, traditional BD-relevant database systems effectively manage only the structured datasets. Handling unstructured datasets is still somewhat challenging since those (unstructured datasets) do not fit inside a rigid data model [10]. The tendency of the data structure to change over time means the volatility of BD. The accuracy of the datasets determines the veracity of BD. The appropriateness of the datasets for their intended use means the validity of BD. Finally, the economic or social wealth generated from BD is referred to as its value. Regarding its economic wealth, it is expected that up to the end of 2022, the big data market will grow to $274.3 billion [11]. The remarkable thing is that the value of BD can be ensured by developing BD analytics (BDA). It (BDA) formally computes the relevant datasets available in BD using statistical or logical approaches [12]. The concerned organizations (e.g., International Organizations for Standardization (ISO) and the National Institute of Standards and Technology (NIST)) have been developing vendor-neutral conceptual definitions, taxonomies, requirements and usages, security and privacy, reference architectures, standardization roadmaps, and adoption and modernization schemes for BD [13–22]. One of the remarkable things is how to mitigate the BD inequality problem [12], [23], ensuring the value (of BD) for all organizations irrespective of their size and information technology enabling resources capacity. The aspect of BD inequality is conceptualized in three dimensions: data access, data representation, and data control inequalities [24–26]. Here, data access-relevant inequality means inaccessibility of the data stored in any data storage infrastructures (e.g., data storage infrastructures of any national or private body) and unavailability of accurate statistics. On the other hand, data representation and control-relevant inequalities mean the lack of infrastructure, human capital, economic resources, and institutional frameworks in developing countries and small firms and organizations, compared to developed countries and big firms and organizations. This generates new dimensions of the digital divide in BDA, knowledge underlying the data, and consequent decision-making abilities. Nevertheless, BDA, by nature, is highly resource-dependent and opaque (acts like a black-box) [27]. Thus, making BDA less resource-dependent and transparent is a challenge. By overcoming this challenge, big data inequality can be resolved.

The remarkable thing is that BD and BDA are valuable constituents of cyber-physical systems of smart manufacturing [28]. This issue is elaborated below using two schematic diagrams (Figures 1, 2). First, consider the smart manufacturing scenario as schematically illustrated in Figure 1. In this scenario, digital manufacturing commons [29–33] integrate the cyber and physical spaces. As seen in Figure 1, different stakeholders (producers, designers, planners, and customers) ubiquitously participated in past manufacturing activities to produce numerous contents such as numerical datasets, digital artifacts (e.g., CAD data), metadata, executable/scripted codes, machine performance statistics, models, algorithms, and so forth. These contents are converted into widely accessible digital format, resulting in digital manufacturing commons. These commons are then stored in a repository called the digital marketplace. The remarkable thing is that the digital marketplace exhibits the Vs of BD described above and acts as BD. Thus, BDA must be installed in the digital marketplace. As such, it (BDA) can operate on digital manufacturing commons and extract the required knowledge. The extracted knowledge helps run the IoT-embedded manufacturing enablers (robots, machine tools, processes, and different planning systems).

Consider the other scenario, as shown in Figure 2. This scenario is similar to the previous one, but this time the aspect of the manufacturing process is focused. As schematically illustrated in Figure 2, a manufacturing process (turning, milling, or electric discharge machining) is controlled by controlling the right set of process variables (e.g., a manufacturing process called turning can be controlled by controlling the cutting conditions (feed rate, cutting speed, depth of cut, cutting tool geometry, coolant, and so on)). Furthermore, this type of control must ensure the desired performance levels in terms of some evaluation variables (e.g., high productivity, low tool wear, low environmental impact, and alike). Thus, two types of parameters entail a manufacturing process. The first type is the parameters that can be set as preferred during a manufacturing process (e.g., cutting conditions in turning). These parameters are denoted as control variables (*CV*s). The other type of parameters is denoted as evaluation variables (*EV*s). Thus, *CV*-*EV*-centric past experimental and analytical results must be transformed into digital manufacturing commons.

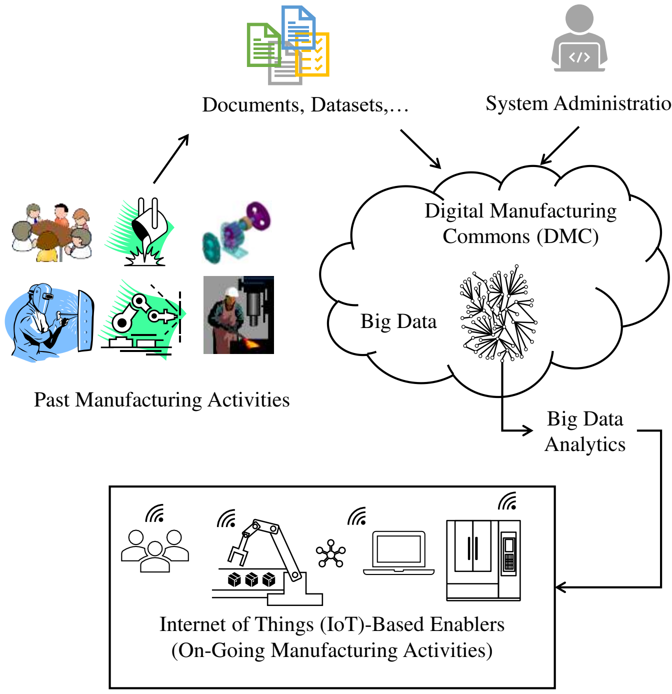

Documents, Datasets,…
System Administration
Digital Manufacturing Commons (DMC)
Big Data
Past Manufacturing Activities
Big Data Analytics
Internet of Things (IoT)-Based Enablers (On-Going Manufacturing Activities)

Figure 1. Context of digital manufacturing commons and big data in smart manufacturing.

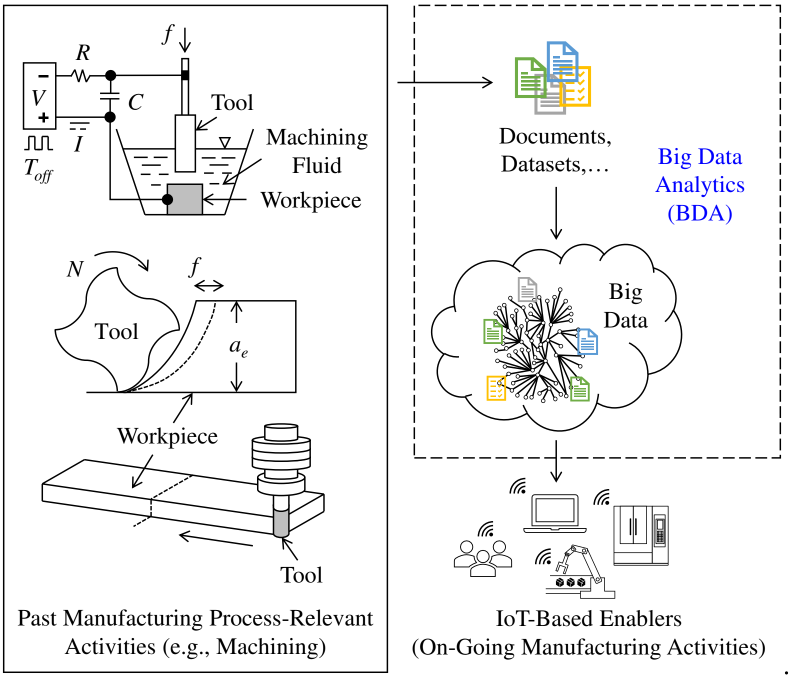


Figure 2. Context of big data analytics for manufacturing processes.

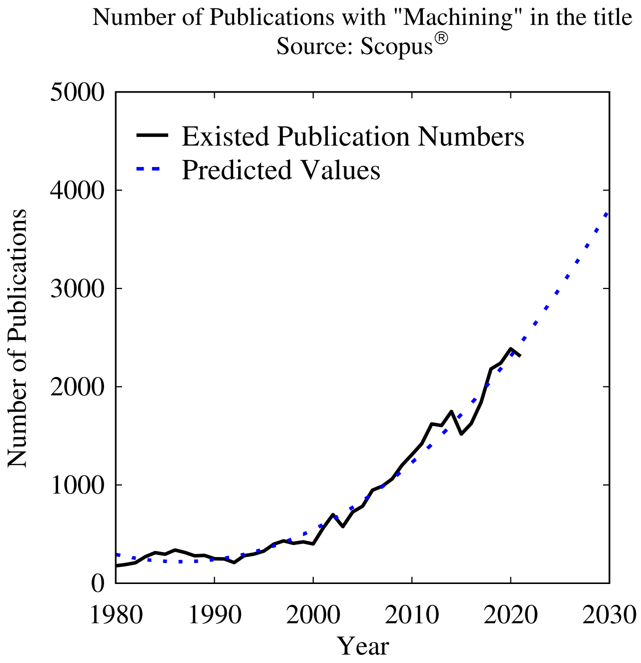


Figure 3. Growth of manufacturing process-relevant documents.

Now, the question is how to develop *CV-EV*-centric BDA. This study answers this question. The rest of this article is organized as follows. Section 2 presents a comprehensive literature review on BDA focusing on smart manufacturing. Section 3 presents the proposed framework of *CV-EV* BDA. This section also elucidates the main modules of the BDA and their functional requirements. Section 4 presents a prototyping BDA developed using JAVA™ platform. Section 5 presents a case study showing the efficacy of the BDA in making decision in a manufacturing process called electrical discharge machining. Finally, Section 6 concludes this article.

2. Literature Review

The following section provides a literature review elucidating the facets of BD and BDA studied by others from the smart manufacturing viewpoint.

Bi et al. [34] proposed an enterprise architecture that combines IoT, BDA, and digital manufacturing to improve a company's ability to respond to challenges. However, they did not provide a detailed outline of the suggested BDA structure.

Wang et al. [35] found that manufacturing cyber-physical systems generate big data from various sources (RFID, sensor, AGV, and alike). They recommend using a BDA approach to analyze the data (data collection, storage, cleaning, integration, analysis, mining, and visualization) but did not provide details on the modules involved.

Kahveci et al. [36] introduced a comprehensive IoT-based BDA platform with five interconnected layers: control and sensing, data collection, integration, storage/analysis, and presentation. In addition, the authors emphasize data retention policies and down-sampling methods to optimize BDA.

Fattahi et al. [37] developed a BDA capable of computing graphs instead of numerical data that helped make the right decisions ensuring Sustainable Development Goal 12 (responsible consumption and production).

Chen and Wang [38] developed a BDA that forecasts cycle time range using data from multiple sources (experts and collaborators). It uses a fuzzy-based deep learning framework to provide predictions learned from relevant datasets. Experts build the computational arrangements, while collaborators interpret the results of the analytics.

Woo et al. [39] developed a BDA platform based on holonic manufacturing systems focusing on object virtualization, data control, and model control. The proposed BDA consists of eight modules:

process data attribute identification, data acquisition, data pre-processing, context synchronization, training dataset preparation, component model computation, model validation, uncertainty quantification, and model composition and use. The analytics can use the Bayesian network, artificial neural network, or statistical analysis, whatever is appropriate.

Bonnard et al. [40] presented a cloud computing-oriented BDA architecture based on three steps (data collection, data storing, and data processing). The BDA gathers data from various levels using technologies such as IoT and ERP, storing it in a distributed database called Cassandra. Finally, machine learning algorithms analyze data, reveal hidden patterns, and predict consequences.

Kozjec et al. [41] introduced a conceptual BDA framework consisting of three levels (implementational level, reference model, and knowledge and skills). The implementation level consists of data and systems (CAD models, CNC programs, quality-control results, test measurements, and alike), hardware and software tools for data management and analysis (NoSQL databases, Scikit-learn, Python, R, Java, and alike), knowledge management, project team, and reference data-analytics solutions.

Jun et al. [42] created a cloud-based BDA framework for manufacturing. It uses a user-defined algorithm template in XML to analyze data related to issues like failure symptoms, RUL prediction, and anomaly detection. The framework selects the appropriate algorithm and visualization technique, such as similarity-based prognostics and time series analysis.

Dubey et al. [43] found that entrepreneurial orientation (EO) traits - proactiveness, risk-taking, and innovativeness - are helpful for decision-making with artificial intelligence-based big data analytics (BDA-AI). Their study used PLS-SEM to integrate entrepreneurship, operations management, and information systems management. As a result, EO can enhance operational performance (OP) in dynamic environments with BDA-AI.

Zhang et al. [44] developed an energy-efficient cyber-physical system to analyze big data and detect production issues in manufacturing workshops. It has three layers: physical energy, cyber energy, and knowledge-driven management. The physical-energy layer includes tools equipped with data acquisition devices. The cyber-energy layer processes data using data cleansing and correlation analysis techniques. Finally, the knowledge-driven management layer uses machine learning algorithms to extract knowledge and make decisions.

Zhong et al. [45] found that incorporating advanced technologies such as DNA-based encryption and self-learning models through deep machine learning can enhance big data analytics in industries like healthcare, finance, economics, supply chain management, and manufacturing. Other technologies include synchronized networks, parallel processing, automatic parallelization, CPL, and cloud computation.

Zhong et al. [46] developed a Big Data Analytics framework using RFID technology in a physical internet-based shop floor setting. The shop floor uses IoT-based tools to turn logistic resources into smart manufacturing objects. A framework has been created to handle the overwhelming amount of data generated by SMOs. It follows five steps: defining data structure, presenting and interpreting data, storing and managing data, processing data with methods like cleansing and classification, and using resulting information for decision-making and predicting.

Zhang et al. [47] proposed the architecture of BDA for product lifecycle (BDA-PL) consisting of four layers: application services of product lifecycle management (PLM), BD acquisition and integration, BD processing and storage, and BD mining and knowledge discovery in database (KDD). It uses RFID and sensors to collect data from multiple sources and process and store using frameworks like Hadoop and SQL. Finally, it analyzes the data using various models to gain knowledge and have a feedback mechanism for sharing.

Lu and Xu [48] introduced a cloud-based manufacturing equipment architecture powered by BDA. It includes sensors, a control module, a monitoring module, and a data processing module. The components interact with a digital twin stored in the cloud, generating data stored in a repository and analyzed through analytics tools. This enables on-demand manufacturing services.

Ji and Wang [49] proposed a framework that uses big data analytics to predict faults. It collects real-time and historical data from the shop floor and performs data cleansing. The framework then uses an analysis algorithm to interpret the data.

Liang et al. [50] found that Big Data (BD) is essential for energy-efficient machining optimization. They developed a system that collects energy consumption data from the shop floor using a wireless system, handles it with a system called Hadoop Hive, and processes it using machine learning algorithms.

Ji et al. [51] presented a machining optimization technique using BDA for distributed process planning. The method uses data attributes to represent machining resources and a hybrid algorithm of Deep Belief Network and Genetic Algorithm for optimization. However, the data analytics structure is not fully explained.

Chen et al. [52] used Hadoop Distributed File System (HDFS) and Spark to extract key characteristics from electric discharge machining (EDM) data. They should have explained how they can be used to predict machining quality. This is an area for potential future research.

To ensure a digital twin function well, Fattahi et al. [53] suggest having a human-cyber-physical system-friendly big data that is easily accessible and understandable for both humans and machines. They also propose a method for preparing the BD dataset, divided into four segments for easy integration with the digital twin's input, modeling, simulation, and validation modules.

Li et al. [54] reviewed industrial Big Data (BD) usage in intelligent manufacturing and found that current Big Data Analytics (BDA) processes face high costs, complex arrangements, and a need for universal frameworks and data-sharing techniques between organizations. The authors presented a conceptual framework for intelligent decision-making based on BD, which includes cyber-physical systems (CPS), digital twins, and BDA. However, further research is needed to determine the feasibility of this framework.

To summarize, no comprehensive guide is available that outlines all the necessary principles, methods, and tools for efficiently constructing and managing BDA for manufacturing decision-making. In order to fill this gap, the following sections present a detailed architecture that can serve as a reference model of BDA for any manufacturing process. In addition, the necessary system components to implement this architecture in real-life scenarios are also developed.

3. Framework of the Proposed BDA

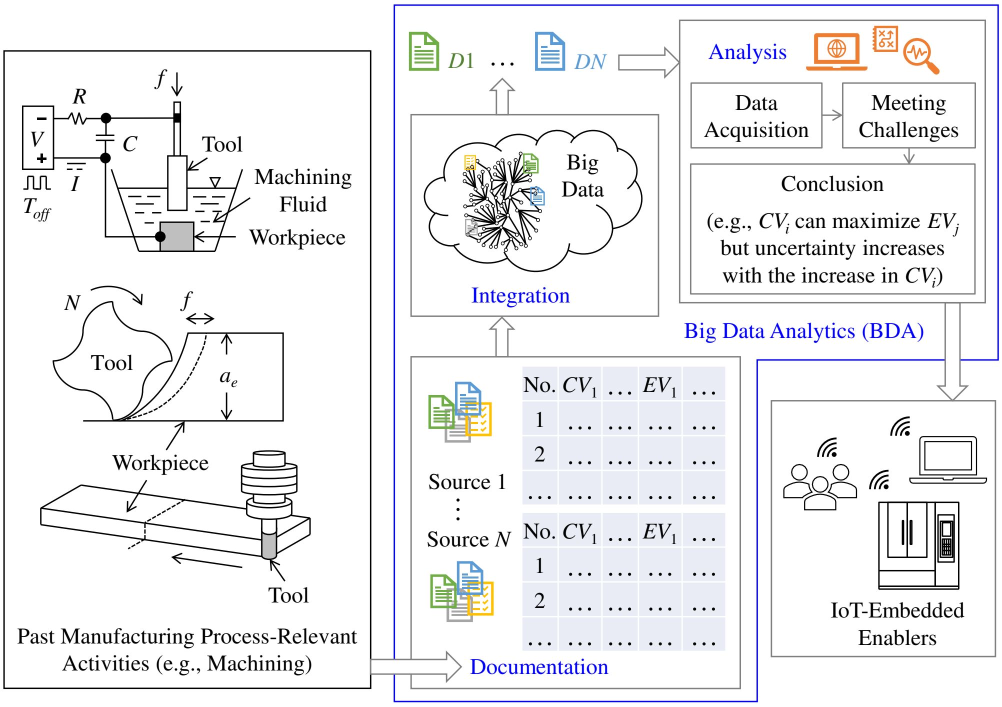
This section presents the proposed big data analytics (BDA). It can be used to making the right manufacturing decisions, and, thereby, to optimize a manufacturing process. In particular, this section first articulates its (BDA) context, basic functionalities, and computational challenges. Afterward, this section presents its (BDA) framework and system architecture.

As mentioned in Section 1, a manufacturing process entails Control Variables (*CV*s) and Evaluation Variables (*EV*s). Here, the *CV*s are some pre-defined variables (e.g., feed rate, depth of cut, cutting speed, and alike) for ensuring the desired performance of the process in terms of some other variables called *EV*s (e.g., low tool wear, high material removal rate, and alike). The knowledge extracted from *CV*-*EV*-centric datasets often drives the relevant process optimization tasks [12]. For example, say a manufacturing process entails a set of *CV*s denoted as $CV_i \mid i = 1,2,\ldots$ and a set of *EV*s denoted as $EV_j \mid j = 1,2,\ldots$. The datasets associated with CV_i and EV_j unfold that “if CV_3 increases, then EV_1 also increases”, meaning that CV_3 is more influential than the other *CV*s as long as EV_1 is concerned. Based on this, one can ensure the desired performance of EV_1 (e.g., maximization, minimization, and alike) by controlling only the CV_3 whenever needed. Therefore, a systematic arrangement (e.g., data analytics) is needed to unfold the knowledge underlying the *CV*-*EV*-centric documents and datasets. As far as smart manufacturing is concerned, *CV*-*EV*-centric past experimental and analytical documents from different sources populate the digital manufacturing commons and generate process-relevant big data, as described in Section 1 (can also be seen in Figure 2). For this, big data analytics (BDA) are a must to extract knowledge from these commons and functionalize optimization of the relevant manufacturing processes. However, in the literature (described in Section 2), no unified framework is found elucidating the systems architecture for

doi:10.20944/preprints202307.1199.v1

developing such BDA. The presented BDA architectures mostly adapt AI-based machine learning/deep learning algorithms (e.g., CANN, GA, DBNN, and alike) and result in black-box systems [27], [55], [56]. Here, black-box systems refer to non-transparent systems where the input and output of the algorithms are known, but internal analysis procedures remain unknown because AI only aims at ultimate goals by its philosophy rather than answering inherent processes [57]. This lack of transparency makes it near impossible to assess the rationality of the algorithm or the BDA system. As a result, such black-box-type BDA frameworks become human-incomprehensible for relevant decision-making [27], [58–60]. One way to solve the above-mentioned issues is to propose a transparent framework elucidating the underlying systems, system-relevant processes, systems integration, and human comprehensibility via human interventions with the systems. In similar context, Ullah and Harib [61] emphasized that a futuristic computer integrated manufacturing system must engage human intelligence (judgment and preference) as straightforwardly as possible for dealing with real-life manufacturing problems. The authors also introduced a human-assisted knowledge extraction system for machining operations. The system utilizes probabilistic reasoning and fuzzy logical reasoning to benefit from the machining data and from the judgment and preference of a user. This contemplation remains valid for BDA as well. Based on this, this study proposes a transparent and human-friendly BDA framework for manufacturing process optimization, as follows.

Figure 4 shows the basic functionalities of the proposed BDA. As seen in Figure 4, the BDA entails three basic functionalities: Documentation, Integration, and Analysis. Here, documentation means documenting a manufacturing process (information related to the process, e.g., machine tool, workpiece, machining conditions, experimental results, *CV*-*EV*-centric datasets, and alike) based on a user-friendly and flexible ontology. It facilitates the structuring of process-relevant documents, eliminating the heterogeneous characteristic due to the involvement of different sources or manufacturing workspaces. Integration means converting the documentation into a machine/human-readable format and integrating it into the process-relevant big data. Finally, analysis means acquiring the desired *CV*-*EV*-centric datasets from big data, meeting computational challenges underlying the datasets, and concluding a set of rules, respectfully. IoT-embedded enablers (machine tools, processes, planning systems, robots, human resources, and alike) residing in a real-life manufacturing environment access the extracted rules and make process optimization-relevant decisions whenever needed. In this regard, one immediate question arises: what sort of computational challenges must be met by the BDA? Figure 5 answers this question schematically.

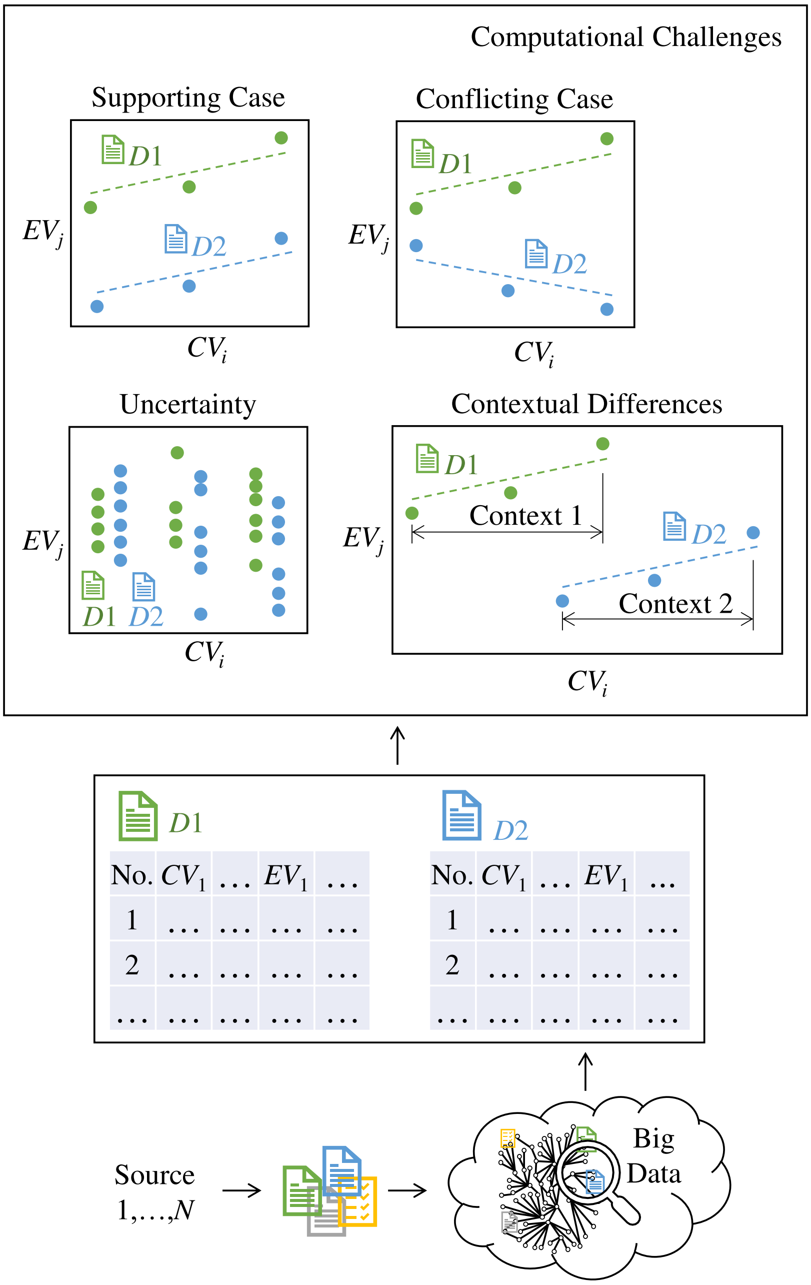


No.	CV_1	…	EV_1	…
1	…	…	…	…
2	…	…	…	…
…	…	…	…	…

No.	CV_1	…	EV_1	…
1	…	…	…	…
2	…	…	…	…
…	…	…	…	…

Figure 4. Basic functionalities of the proposed big data analytics.

Say, as seen in Figure 5, the *CV*-*EV*-centric past experimental and analytical documents from different sources generate manufacturing process-relevant big data. The BDA searches the big data for a user-defined keyword (e.g., process type, workpiece material, and alike) and acquires two relevant documents, denoted as D1 and D2 in Figure 5, that originated from two different sources. As a result, some computational challenges appear for the BDA. In particular, the D1 and D2 may provide supporting or conflicting cases related to *CV*-*EV* relationships. Here, the supporting case means that D1 and D2 reflect the same or similar *CV*-*EV* correlation. For example, as seen in Figure 5, D1 reflects a direct *CV*-*EV* correlation, and D2 does the same. On the other hand, the conflicting case means that D1 and D2 reflect the opposite or dissimilar *CV*-*EV* correlation. For example, as seen in Figure 5, D1 reflects a direct correlation, but D2 reflects an indirect one for the same *CV*-*EV*.



D1

No.	CV_1	…	EV_1	…
1	…	…	…	…
2	…	…	…	…
…	…	…	…	…

D2

No.	CV_1	…	EV_1	…
1	…	…	…	…
2	…	…	…	…
…	…	…	…	…

Figure 5. Computational challenges underlying big data analytics.

Apart from the above-mentioned (supporting and conflicting cases), challenges associated with uncertainty and contextual differences might appear, as shown in Figure 5 schematically. Here, uncertainty means the variability in the acquired *CV-EV*-centric datasets. It reflects how far apart the datasets are. Its quantification helps more accurate predictions in terms of consistency. For example, less uncertainty means the *CV-EV* correlation is consistent, and the correlation might be generalized for other cases. High uncertainty means the *CV-EV* correlation is inconsistent, and the correlation might not be generalized. The remarkable thing is that even though *CV-EV* datasets exhibit a strong direct/indirect correlation, the associated uncertainty can reflect that the correlation is not good enough to be generalized, given that the uncertainty is high. Therefore, identifying only the *CV-EV* correlation is not adequate for effective decision-making. Uncertainty quantification is also needed to get complete knowledge underlying the datasets and make the right decisions. On the other hand, contextual differences appear when *CV-EV*-centric datasets entail different discourses without following a standard one because of the heterogeneity of sources associated with the different *CV* levels and experimental design of the manufacturing process. As such, the BDA must meet the above-mentioned challenges for concluding a set of rules among *CV-EV*-centric datasets from big data. Based on the consideration described above, the following questions arise:

1. How should the documentation process for manufacturing be carried out? (Q1)
2. What should be the process for integrating the prepared documents with process-relevant big data? (Q2)
3. What is the proper procedure for utilizing the relevant dataset (specifically, *CV-EV*-related datasets) found in the shared documents? (Q3)
4. What should be the method for meeting the computational challenges? (Q4)
5. What is the recommended method for extracting rules and drawing conclusions? (Q5)

The answers to the above-mentioned questions lead to a transparent BDA framework, as shown in Figure 6 schematically. As seen in Figure 6, the framework consists of five systems: (1) Big Data Preparation System (BDPS), (2) Big Data Exploration System (BDES), (3) Data Visualization System (DVS), (4) Data Analysis System (DAS), and (5) Knowledge Extraction System (KES).

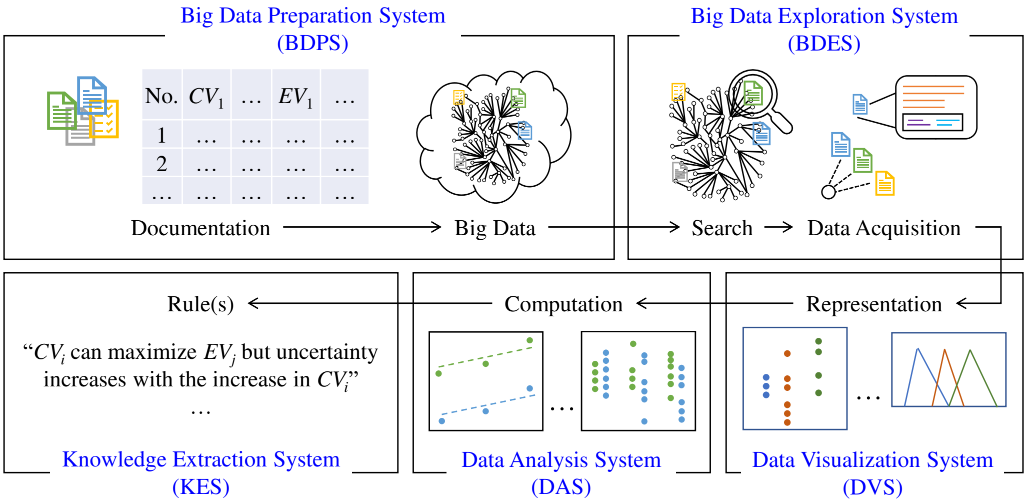

Big Data Preparation System
(BDPS)
No. | CV_1 | … | EV_1 | …
1 | … | … | … | …
2 | … | … | … | …
… | … | … | … | …
Documentation
Big Data
Big Data Exploration System
(BDES)
Search
Data Acquisition
Representation
Computation
Rule(s)
"CV_i can maximize EV_j but uncertainty increases with the increase in CV_i"
…
Knowledge Extraction System
(KES)
Data Analysis System
(DAS)
Data Visualization System
(DVS)

Figure 6. Framework of the proposed big data analytics.

The systems collectively answer the above-mentioned questions (Q1,...,Q5) and functionalize the basic functionalities (Documentation, Integration, and Analysis) of the BDA (can also be seen in Figure 4). In particular, BDPS answers Q1 and Q2. For this, it (BDPS) functionalizes the documentation of the process-relevant information using a flexible ontology, followed by document integration into the process-relevant big data. BDES answers Q3. For this, it (BDES) functionalizes searching for *CV*-*EV*-centric datasets from the big data using a user-defined keyword (e.g., process type, workpiece material, and alike) and acquiring the appropriate *CV*-*EV*-centric datasets from the searched outcomes. DVS and DAS collectively answer Q4. For this, DVS functionalizes the representation of the acquired *CV*-*EV*-centric datasets graphically. DAS functionalizes intelligent computations on the *CV*-*EV*-centric datasets for meeting the computational challenges. Finally, KES answers Q5. For this, it (KES) functionalizes rule extraction based on the outcomes of DAS.

Figure 7 schematically shows the relationships among the above-mentioned systems and their activities. As seen in Figure 7, BDPS provides a facility to create a metafile based on user-defined process-relevant inputs (process type, number of experiments, number of *CV*s/*EV*s, and maximum number of *CV* levels). The metafile can follow any file format. However, this study considers Excel™-based metafile for comprehensibility and availability. It (Excel™-based metafile) provides a user-comprehensible and flexible ontology for documenting a manufacturing process, incorporating process-relevant information such as source, summary, process, machine, tool, workpiece, machining conditions, *CV*s, *EV*s, and results or *CV*-*EV*-centric datasets. After documentation, BDPS provides another facility to convert the document into a machine/human-readable format (e.g., Extensible Markup Language (XML)) and share the XML data through a cloud-based data repository or store it in a local data repository for future use. As a result, the cloud-based data repository contains XML data from different sources and generates process-relevant big data. BDES provides a facility to search the repository using a user-defined keyword (e.g., process type, workpiece material, and alike) and fetch the relevant outcomes (files containing XML data). The outcomes are also presented in a meaningful way by the system. For this, BDES provides a facility to display the contents of the fetched XML data using an appropriate data presentation technique (here, HTML-based presentation). This presentation benefits the user to decide whether or not to adapt the contents for subsequent analysis. If not, the user may re-search the repository. Otherwise, *CV*-*EV*-centric datasets from the XML data are acquired. DVS provides a facility to visualize/represent the acquired *CV*-*EV*-centric datasets using different data visualization techniques (e.g., scatter plots, line graphs, histograms, area charts, heat maps, possibility distribution, and so forth). The user can choose an appropriate technique within the system and visualize the datasets whenever needed. DAS provides a facility to analyze the *CV*-*EV*-centric datasets and identify the *CV*-*EV* relationships using different computational methods (e.g., correlation analysis, uncertainty analysis, and alike). The user can choose appropriate methods whenever needed. As a result, DAS generates a set of analyzed outcomes corresponding to different sources. Finally, KES utilizes the analyzed outcomes to extract underlying knowledge based on user-defined optimization criteria (e.g., maximization/minimization of an *EV*) and concludes optimization rule(s).

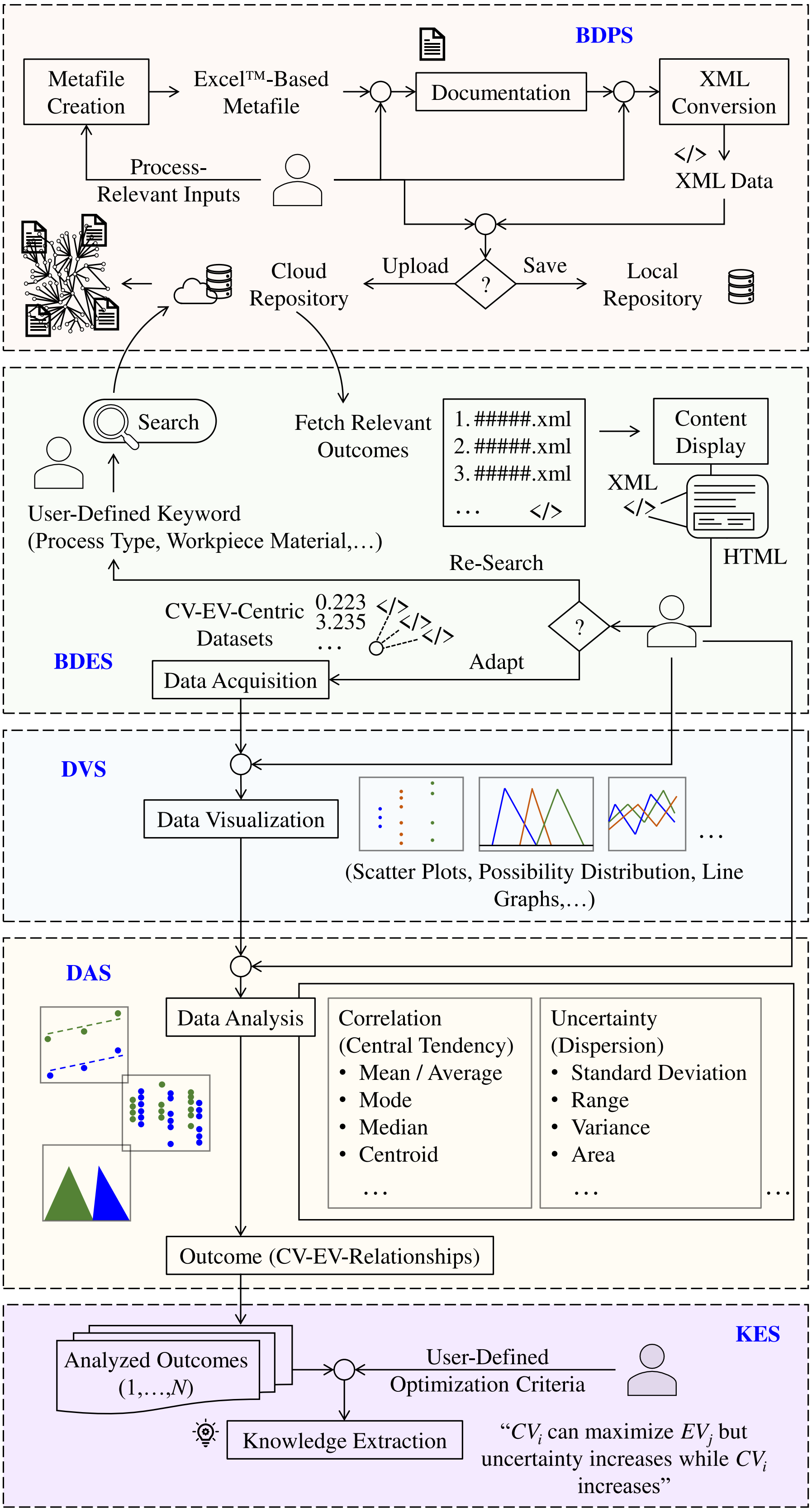


Figure 7. System architecture of the proposed big data analytics.

The remarkable things about the above BDA framework are its transparency and human intervention associated with the underlying systems, functionalities, and methods. It provides the users with a more feasible way of understanding the internal processes, assessing the fairness of the outcomes, and making relevant decisions. Nevertheless, computerized systems are also developed based on the proposed framework and implemented in a case study, as described in the following section.

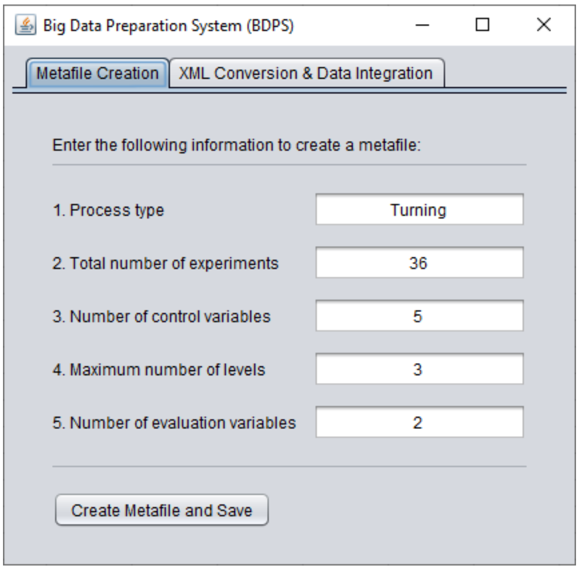
4. Developing BDA

As described in Section 3, the proposed big data analytics (BDA) entail five systems: (1) Big Data Preparation System (BDPS), (2) Big Data Exploration System (BDES), (3) Data Visualization System (DVS), (4) Data Analysis System (DAS), and (5) Knowledge Extraction System (KES). The systems collectively functionalize the three basic functionalities (Documentation, Integration, and Analysis) and drive manufacturing process-relevant optimization tasks. Nevertheless, the systems are developed using a Java™-based platform. This section presents the developed systems in the following sub-sections (Sections 4.1 – 4.5).

4.1. Big Data Preparation System (BDPS)

As described in Section 3, BDPS functionalizes the documentation of the process-relevant information using a flexible ontology-based metafile, followed by document integration into the process-relevant big data in terms of extensible markup language (XML) data. As such, the developed BDPS entails two modules: (1) Metafile Creation and (2) XML Conversion and Data Integration.

Consider the Metafile Creation module of BDPS. Figure 8 shows one instance of it. As seen in Figure 8, the module first takes some user-inputs relevant to a manufacturing process. The user-inputs are: process type, total number of experiments, number of control variables (*CV*s), maximum number of levels, and number of evaluation variables (*EV*s). For the instance shown in Figure 8, the user-inputs are turning, 36, 5, 3, and 2, respectively. The module then provides a facility (a button denoted as ‘Create Metafile and Save’ in Figure 8) by which the user can generate an Excel™-based metafile and save the metafile in a local repository, whenever needed. Note that, the metafile is generated based on the user-inputs. The reasons for considering Excel™-based metafile are its (Excel™) availability and comprehensibility in all sorts of workspaces.


Big Data Preparation System (BDPS)
Metafile Creation
XML Conversion & Data Integration
Create Metafile and Save

Enter the following information to create a metafile:

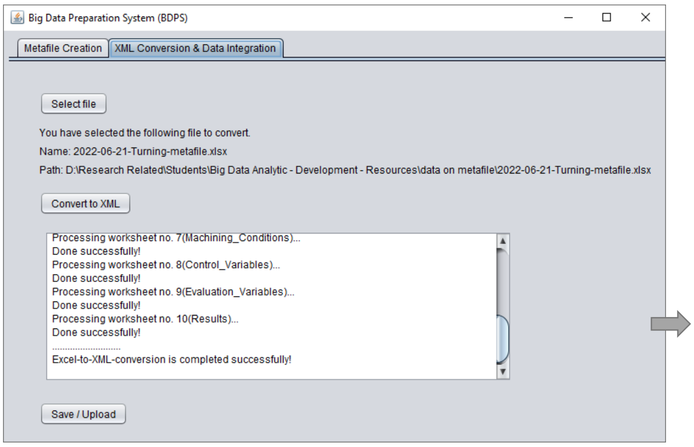
1. Process type: Turning
2. Total number of experiments: 36
3. Number of control variables: 5
4. Maximum number of levels: 3
5. Number of evaluation variables: 2

Figure 8. Screen-print of Metafile Creation module.

Nevertheless, the user can use the generated metafile for documenting the manufacturing process, integrating the relevant attributes such as the source of the experiment (e.g., location, organization, and so forth), a summary of the experiment (e.g., purpose, findings, and so forth), process specifications (e.g., type of process), machine specifications (e.g., type, maker, model, and so forth), tool specifications (e.g., type, material, maker, shape, dimension, and so forth), workpiece specifications (e.g., type, material, shape, size, composition, thermal properties, hardness, tolerance, and so forth), machining conditions, control variables (*CV*s), evaluation variables (*EV*s), and experimental results (*CV*-*EV*-centric numerical datasets).

One remarkable thing about documentation using the metafile is its flexibility. Different users from different workspaces may prefer different ways of documenting a manufacturing process. For example, a user may prefer only documenting the experimental results, whereas another may prefer integrating all or some of the above-mentioned relevant attributes associated with the results (e.g., machining conditions, machine, tool, workpiece, and alike). The metafile is flexible to support such heterogeneity.

Now, consider the XML Conversion and Data Integration module of BDPS. It converts a filled metafile (metafile after documentation) into machine-readable extensible markup language (XML) data and integrates the XML into a cloud-based repository. Figure 9 shows one instance of the module. As seen in Figure 9, the module first provides a facility (a button denoted as 'Select file' in Figure 9) to select a filled metafile. After conversion, the module provides a facility (a button denoted as 'Save/Upload' in Figure 9) to save the XML data on a local repository for future use or upload the XML data to a cloud-based repository for contributing to the digital manufacturing commons. Whenever the user accesses this facility, a separate pop-up window appears (not shown in Figure 9) from where the user chooses the appropriate options, i.e., Save/Upload. When 'Upload' is chosen, the module activates an access token, connects to a cloud-based repository using the token, and uploads the XML data to the repository. This way, users from different workspaces may create the digital manufacturing commons and process-relevant big data for different manufacturing processes, utilizing the BDPS as mentioned above.


Big Data Preparation System (BDPS)
Metafile Creation
XML Conversion & Data Integration
Select file
Convert to XML
Save / Upload

You have selected the following file to convert.
Name: 2022-06-21-Turning-metafile.xlsx
Path: D:\Research Related\Students\Big Data Analytic - Development - Resources\data on metafile\2022-06-21-Turning-metafile.xlsx

```
Processing worksheet no. 7(Machining_Conditions)...
Done successfully!
Processing worksheet no. 8(Control_Variables)...
Done successfully!
Processing worksheet no. 9(Evaluation_Variables)...
Done successfully!
Processing worksheet no. 10(Results)...
Done successfully!
..........................
Excel-to-XML-conversion is completed successfully!
```

XML Data

```
…
</cluster9-Evaluation_Variables>
<cluster10-Results>
<row><cell1>Experiment
no.</cell1><cell2>Insert Radius</cell2>
<cell3>Cutting Speed</cell3><cell4>Depth of
Cut</cell4>
<cell5>Feed Rate</cell5><cell6>Coolant
Flowrate</cell6>
<cell7>Cutting Force</cell7><cell8>Surface
Roughness</cell8>
</row><row>
<cell1>1.0</cell1><cell2>0.4</cell2>
<cell3>40.0</cell3><cell4>0.1</cell4>
<cell5>0.1</cell5><cell6>2.0</cell6>
<cell7>29.42</cell7><cell8>1.9</cell8>
</row>
…
```

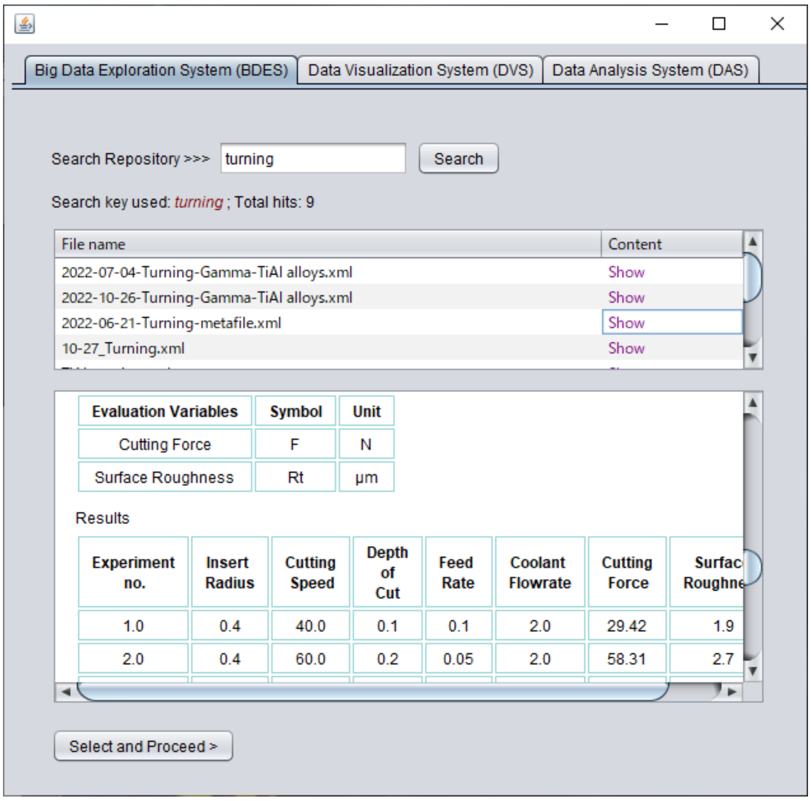
Figure 9. Screen-print of XML Conversion and Data Integration module.

4.2. *Big Data Exploration System (BDES)*

As described in Section 3, BDES functionalizes searching *CV*-*EV*-centric datasets and acquiring the relevant ones from a cloud-based repository that hosts XML data from different sources and generates process-relevant big data. Figure 11 shows one instance of the developed BDES.

As seen in Figure 10, BDES first provides a user-defined search facility, where a user may define a process-relevant search key, such as process type, workpiece material, and alike. For instance, shown in Figure 10, the defined search key is 'turning', a manufacturing process type. Based on the search key, BDES finds all the relevant XML files from the repository and displays the outcomes as a list. For this, the BDES also maintains a connection with the repository, just like the XML Conversion and Data Integration module of BDPS. The BDES then provides a facility (buttons denoted as 'Show' in Figure 10) to present the search outcomes meaningfully. For this, BDES creates an HTML presentation of the contents underlying a specific search outcome (XML data) in the embedded window whenever the user accesses the corresponding 'Show' button. For example, Figure 10 presents the contents (tool, workpiece, machining conditions, *CV*s, *EV*s, *CV*-*EV*-centric datasets, and alike) underlying the XML data created in Section 4.1. This presentation helps the user decide whether or not the search outcomes are appropriate for further analysis, per the user's requirements.

Based on the decision, the user can acquire *CV-EV*-centric datasets from the search outcomes via another BDES facility, a button denoted as 'Select and Proceed' in Figure 10, or re-search.


Big Data Exploration System (BDES)
Data Visualization System (DVS)
Data Analysis System (DAS)
Search
Select and Proceed >

Search Repository >>> turning

Search key used: *turning* ; Total hits: 9

File name	Content
2022-07-04-Turning-Gamma-TiAl alloys.xml	Show
2022-10-26-Turning-Gamma-TiAl alloys.xml	Show
2022-06-21-Turning-metafile.xml	Show
10-27_Turning.xml	Show

Evaluation Variables	Symbol	Unit
Cutting Force	F	N
Surface Roughness	Rt	μm

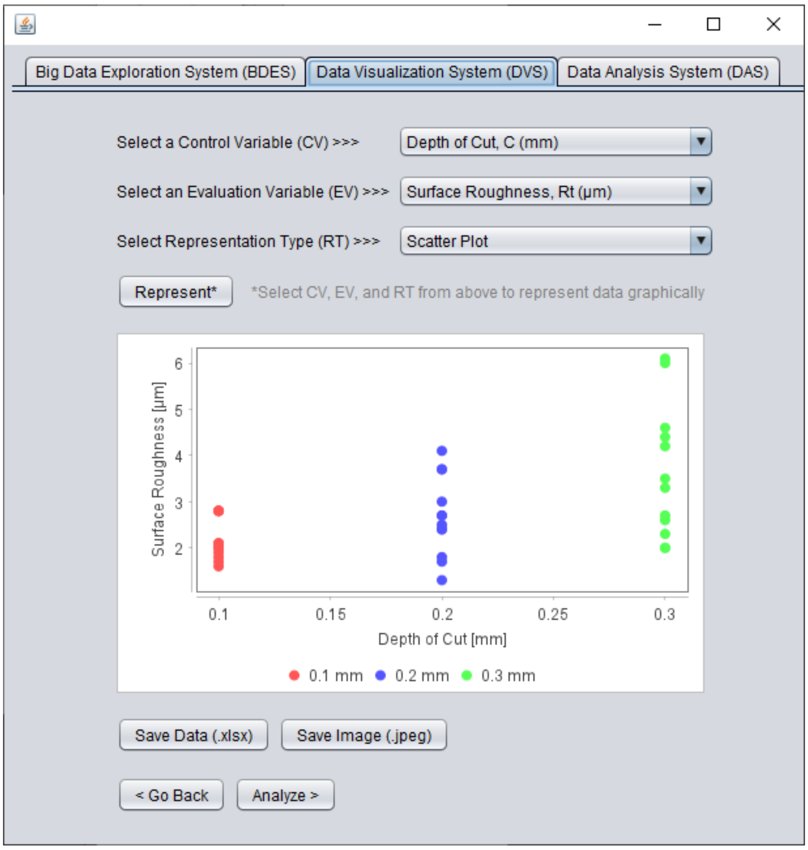
Results

Experiment no.	Insert Radius	Cutting Speed	Depth of Cut	Feed Rate	Coolant Flowrate	Cutting Force	Surfac Roughne
1.0	0.4	40.0	0.1	0.1	2.0	29.42	1.9
2.0	0.4	60.0	0.2	0.05	2.0	58.31	2.7

Figure 10. Screen-print of Big Data Exploration System.

4.3. *Data Visualization System (DVS)*

As described in Section 3, DVS functionalizes visualizing the acquired *CV-EV*-centric datasets from BDES. Figure 11 shows one instance of the developed DVS.



Select a Control Variable (CV) >>> Depth of Cut, C (mm)

Select an Evaluation Variable (EV) >>> Surface Roughness, Rt (µm)

Select Representation Type (RT) >>> Scatter Plot

*Select CV, EV, and RT from above to represent data graphically

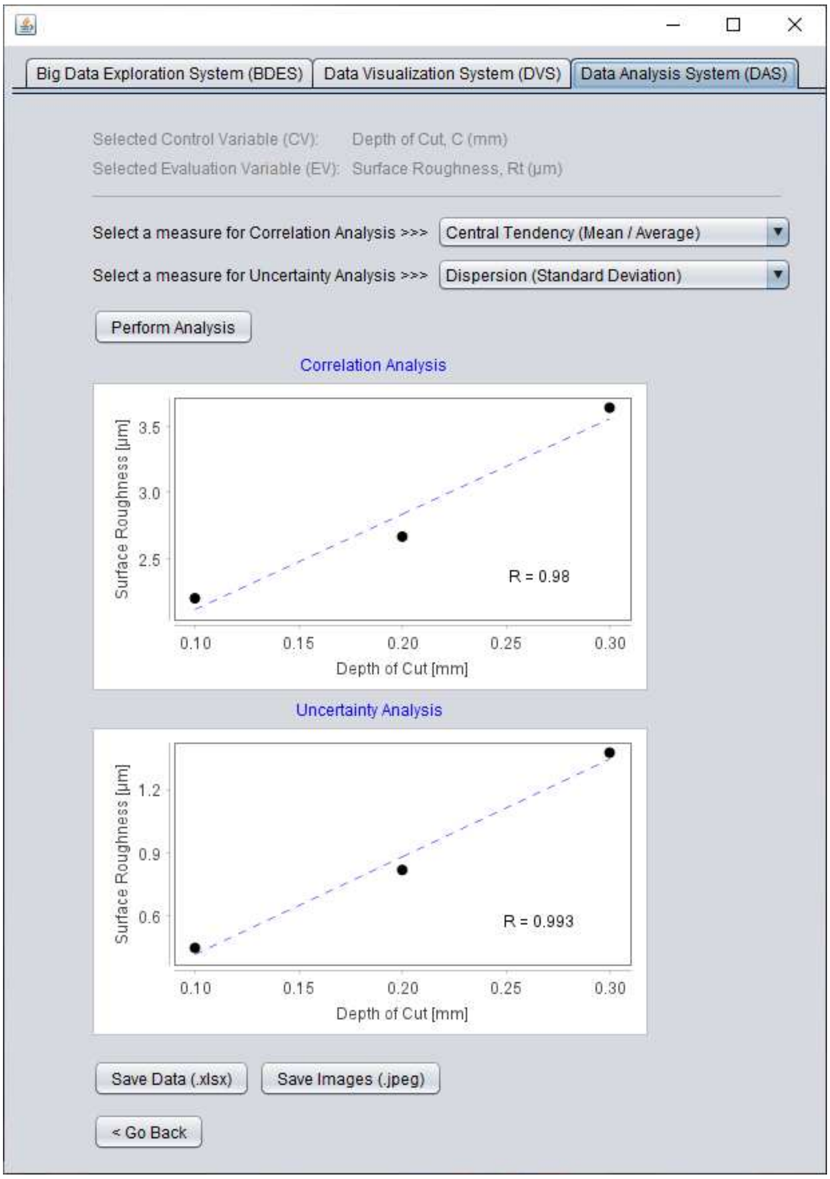
Figure 11. Screen-print of Data Visualization System.

As seen in Figure 11, DVS provides two facilities (drop-down lists denoted as 'Select a Control Variable (*CV*)' and 'Select an Evaluation Variable (*EV*)' in Figure 11) to select a *CV* and *EV* among others from the acquired *CV*-*EV*-centric datasets. For instance, as shown in Figure 11, a *CV* called 'Depth of Cut' and an *EV* called 'Surface Roughness' are selected. Note that the facilities (drop-down lists) update the list of *CV*s and *EV*s based on the acquired datasets whenever needed. DVS then provides another facility (drop-down list denoted as 'Select a Representation Type (RT)' in Figure 11) to select an appropriate representation technique among many such as scatter plots, possibility distributions [62], line graphs, and so forth. For instance, as shown in Figure 11, a scatter plot is selected. When a user sets a *CV*, *EV*, and representation technique, DVS provides another facility (a button denoted as 'Represent' in Figure 11) for visualizing the set *CV*-*EV*-centric datasets in the form of the set representation technique. Figure 11 shows such an instance accordingly. This visualization helps a user understand individual *CV*-*EV* relationships and underlying *CV* levels. This way, DVS aids a user in visualizing the acquired *CV*-*EV*-centric datasets whenever needed. It also provides facilities (buttons denoted as 'Save Data' and 'Save Image' in Figure 11) to store the visualization outcomes in the forms of numeric datasets and images, if needed.

Although DVS helps understand the *CV*-*EV* relationships qualitatively, as described above, the relationships must be quantified for knowledge extraction. For this, DVS provides a facility (a button denoted as 'Analyze' in Figure 11) to transfer the set *CV*-*EV* datasets to the next system called DAS and analyze quantitatively.

4.4. Data Analysis System (DAS)

As described in Section 3, DAS functionalizes analyzing the *CV*-*EV*-centric datasets. For this, it (DAS) receives the DVS-supplied datasets (see Section 4.3), deploys user-defined computational methods (e.g., correlation analysis and uncertainty analysis), and quantifies the *CV*-*EV* relationships. Figure 12 shows an instance of the developed DAS.


Selected Control Variable (CV): Depth of Cut, C (mm)

Selected Evaluation Variable (EV): Surface Roughness, Rt (µm)

Select a measure for Correlation Analysis >>> Central Tendency (Mean / Average)

Select a measure for Uncertainty Analysis >>> Dispersion (Standard Deviation)

Figure 12. Screen-print of Data Analysis System.

As seen in Figure 12, DAS first displays the *CV* and *EV* underlying the DVS-supplied datasets. For the instance shown in Figure 12, the *CV* and *EV* are 'Depth of Cut' and 'Surface Roughness', respectively, supplied from the DVS (see Figure 11). DAS then provides two facilities (drop-down menus in Figure 12) to select measures for analyzing correlation and uncertainty associated with the datasets. For the instance shown in Figure 12, the measures called 'Central Tendency' and 'Dispersion' are selected, respectively. Afterward, DAS provides another facility (a button denoted as 'Perform Analysis') for analyzing the datasets based on the set measures. The outcomes of the analyses (correlation and uncertainty analyses) are displayed graphically, quantified by *R*-values, where $R \in [-1,1]$. An *R*-value closer to '1' indicates a strong direct relationship between the *CV* and *EV*. An *R*-value closer to '-1' indicates a strong indirect relationship between the *CV* and *EV*. For the instance shown in Figure 12, the *R*-values for correlation and uncertainty analyses are 0.980 and 0.993, respectively. This means that the *CV* (here, Depth of Cut) and *EV* (here, Surface Roughness) entail a strong direct correlation associated with high uncertainty. Note that the DAS can be equipped with other computational methods and underlying measures for the sake of analysis, if needed, due to its (DAS) modular architecture. The outcomes from the DAS can also be exported, whenever needed, accessing the in-built facilities (see buttons denoted as 'Save Data' and 'Save Images' in Figure 12).

Nevertheless, a user can explore the manufacturing process-relevant big data and visualize and analyze the *CV*-*EV*-centric datasets using the systems mentioned above: BDES, DVS, and DAS,

doi:10.20944/preprints202307.1199.v1

respectively. The systems are human-comprehensible, offering thorough human interventions compared to the existing black box systems described in Section 3. The analyzed outcomes are processed in the KES for rule(s) extraction, as follows.

4.5. Knowledge Extraction System (KES)

As described in Section 3, KES functionalizes knowledge extraction (or rule(s) extraction) underlying the DAS-supplied analysis outcomes related to *CV-EV*-centric datasets from process-relevant big data (see Section 4.4). Consider the following example for a better understanding.

Say, for a given process type, *CV-EV*-centric datasets from different sources populate the process-relevant big data with the aid of BDPS (see Section 4.1). Next, a user explores the big data based on a search keyword (e.g., process type, workpiece, and alike) and acquires the desired *CV-EV*-centric datasets from all or some of the sources with the aid of BDES (see Section 4.2). The user then visualizes and analyzes the datasets acquired from different sources with DVS and DAS (see Sections 4.3 and 4.4, respectively). As a result, DAS generates analysis outcomes for all the acquired *CV-EV*-centric datasets from different sources. The user finally processes these (analysis outcomes for different sources) for extracting common knowledge with the aid of KES. Of course, the users may deploy different techniques for processing the analysis outcomes and extracting underlying knowledge when the outcomes are gathered in the users' vicinities. This makes the KES highly user-dependent. Nevertheless, the following section presents a case study showing how KES functionalizes knowledge extraction for optimizing a given manufacturing process.

5. Case Study

This case study deals with a manufacturing process called Electrical Discharge Machining (EDM). In order to optimize EDM operations, knowledge is needed, which comes from experimental studies. There are many experimental studies regarding EDM. For example, when Google Scholar was searched using the keyword "EDM," it produced 52900 hits. All these studies constitute big data of EDM. However, datasets specific to a workpiece material are more informative because a process is optimized for a specific workpiece material. Therefore, it would have been advantageous if all the necessary datasets were stored in a machine-readable format. Unfortunately, it is not the case now. However, the authors selected six studies [63-68] on dry EDM where the workpiece materials are high-speed or stainless steels. The datasets are digitized using the systems described previously (see Figure 9). They are also explored and analyzed using the systems described in Figures 10-12. Before describing the results, it is important to see whether or not the issues described in Sections 3 and 4 are relevant here. The description is as follows.

The *CV*s and *EV*s of the studies [63-68] are listed in Table 1. These studies provide a total of 115 datasets. Even though the datasets are limited to 115, they exhibit some of the Vs (see Section 1) of big data. The following visualizations are made to elucidate these.

Table 1. List of *CV*s and *EV*s across different sources for EDM.

Source	Control Variable ($CV_{j=1,\ldots,m}$)		Evaluation Variable ($EV_{k=1,\ldots,n}$)		Datasets
(S_i)	j	CV_j	k	EV_k	($m \times n$)
S_1 [63]	1 2 3 4 5	Gap Voltage Current Pulse Off Time Gas Pressure Rotational Speed	1 2 3 4	Material Removal Rate Tool Wear Rate Radial Over Cut Depth Achieved	20
S_2 [64]	1 2 3 4 5	Gap Voltage Current Pulse Off Time Gas Pressure Rotational Speed	1 2 3 4 5	Material Removal Rate Tool Wear Rate Oversize (entry of hole) Oversize (50% of hole depth) Oversize (90% of hole depth)	30

	6	Shielding Clearance			
S_3 [65]	1 2 3 4	Current Pulse On Time Duty Factor Rotational Speed	1 2	Material Removal Rate Tool Wear Rate	8
S_4 [66]	1 2 3 4 5 6	Gap Voltage Current Pulse Off Time Gas Pressure Rotational Speed Shielding Clearance	1 2 3 4	Material Removal Rate Tool Wear Rate Oversize Depth Achieved	24
S_5 [67]	1 2 3 4 5 6	Current Pulse On Time Duty Factor Gas Pressure Rotational Speed Gas Type	1 2 3	Material Removal Rate Surface Roughness Radial Over Cut	18
S_6 [68]	1 2 3 4 5	Current Pulse On Time Duty Factor Gas Pressure Rotational Speed	1 2 3	Material Removal Rate Surface Roughness Radial Over Cut	15
Total number of *CV-EV*-centric datasets					115

The following *CV*s are used in the said studies: Current (I), Gas Type, Duty Factor (η), Gap Voltage (V), Gas Pressure (P), Pulse Off Time (T_{off}), Pulse On Time (T_{on}), Rotational Speed (S), and Shielding Clearance (C_b). The following *EV*s are used in the said studies: Depth Achieved, Material Removal Rate (*MRR*), Oversize, Oversize (50% of hole depth), Oversize (90% of hole depth), Oversize (entry of hole), Radial Over Cut, Surface Roughness (*Ra*), and Tool Wear Rate.

First, consider whether the datasets exhibit some of the Vs of BD. A series of plots shown in the following diagrams are considered. Consider the bubble plot shown in Figure 13. The plot displays datasets from six different sources in distinct colors. The size of each bubble corresponds to the number of data sets for each possible *CV-EV* combination. However, 61 unique *CV-EV* combinations are exhibited by these 115 datasets, as shown by the plot in Figure 14. This time the bubbles are organized according to their sources.

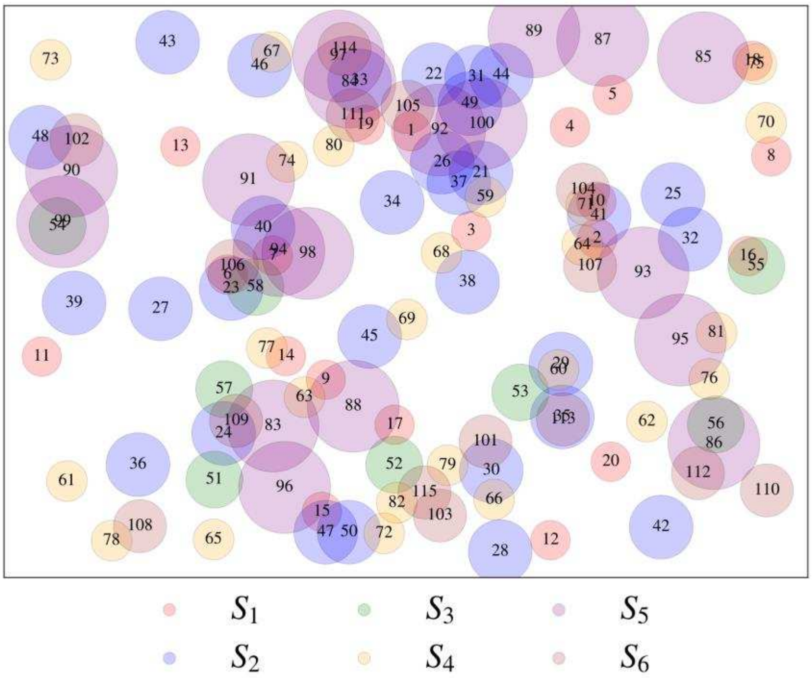


Figure 13. *CV-EV*-centric datasets across relevant to Table 1.

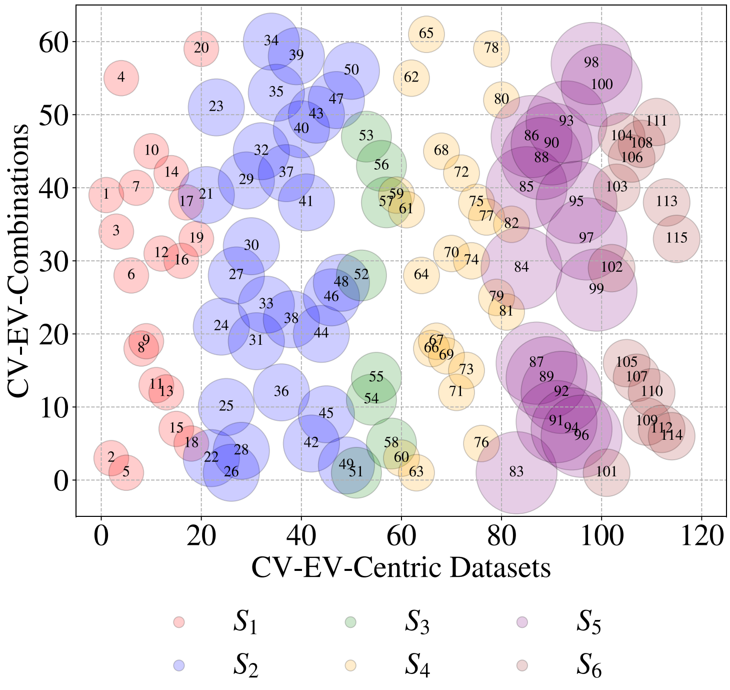


Figure 14. Available combination-centric visualization *CV-EV* datasets.

Let us be more specific. Consider an *EV* called Material Removal Rate (*MRR*). It refers to 32 datasets among115 and 9 unique combinations of *CV-EV*, as shown in Figure 15. As seen in Figure 15, all six sources deal with *MRR*. Consider another *EV* called Tool Wear Rate (*TWR*). It refers to 21 datasets among 115 and 8 unique combinations of *CV-EV*, as shown in Figure 16. Four sources among six provide datasets for this *EV* (i.e., *TWR*), unlike the case shown in Figure 15. Lastly, consider the *EV* called surface roughness measured by the arithmetic average of roughness profile height deviations from the mean line (*Ra*). In this case, only two sources (S_5 and S_6) provide 11 datasets from six unique combinations of *CV-EV*, as shown in Figure 17. This means that a great deal of heterogeneity persists among the datasets, exhibiting some of the characteristics of BD. The author assumes that the heterogeneity level may remain unchanged even though more sources are considered.

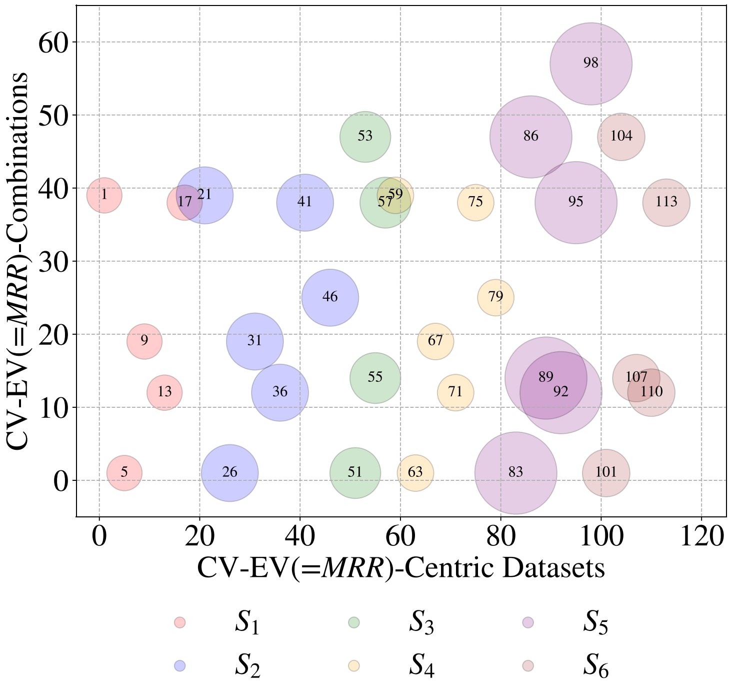


Figure 15. *MRR*-relevant datasets from different sources.

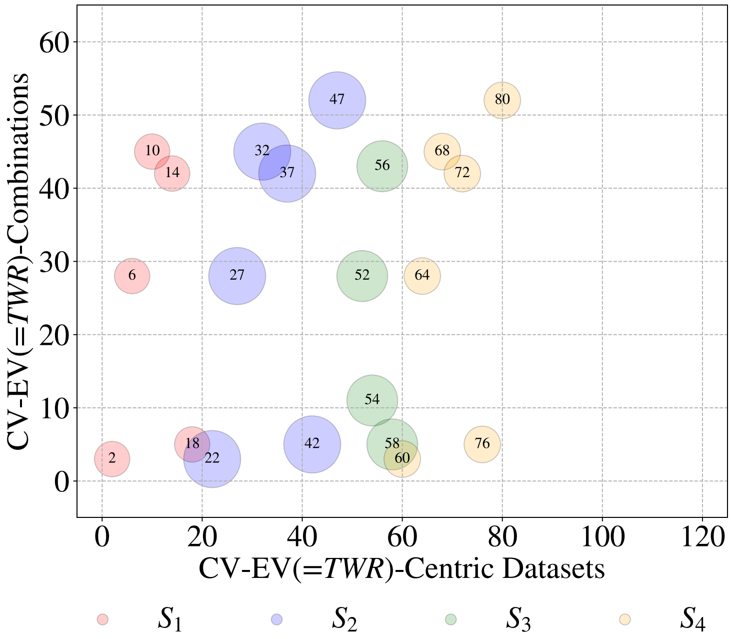


Figure 16. *TWR*-relevant datasets from different sources.

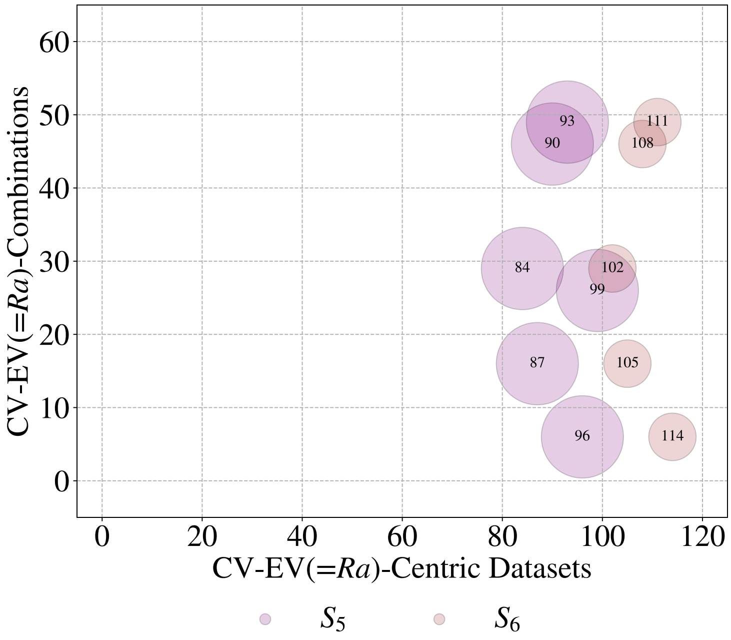
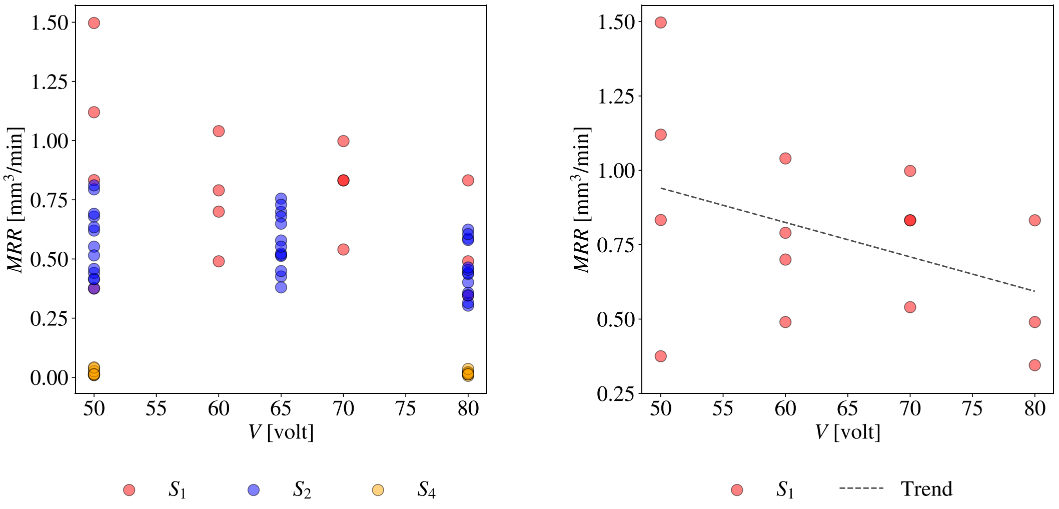


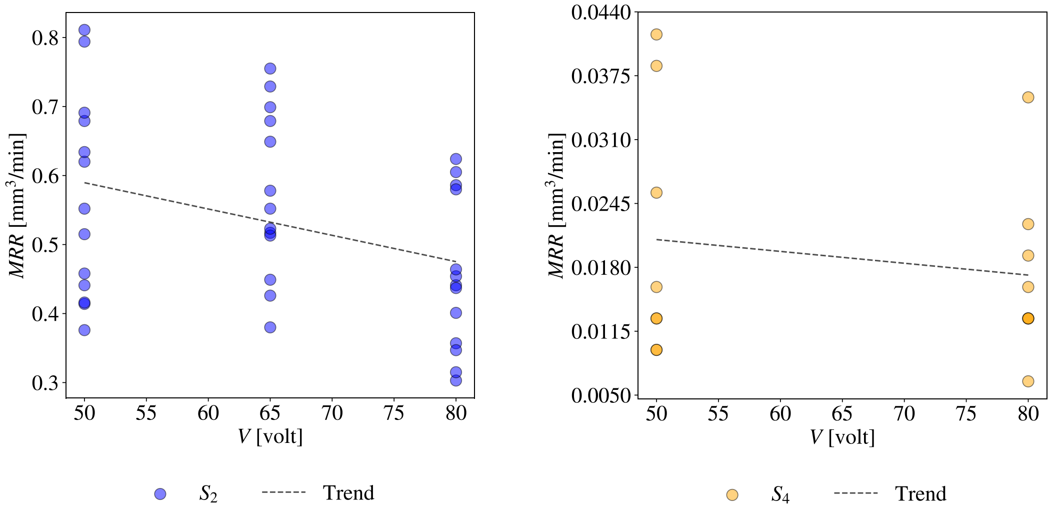
Figure 17. *Ra*-relevant datasets from different sources.

However, when the utility of BD is considered, the characteristics of the validity and value become critical, i.e., whether or not trustworthy knowledge can be extracted to solve problems using the relevant segment of BD. In this particular case, the relationships among *CVs* and *EVs* serves as the primary knowledge. must be quantified. The relationships are established by using the tools available in the system called the Data Analysis System (Figure 12). The tools must be used keeping in mind that there are some computational complexities, as schematically illustrated in Figure 5. Since datasets are collected from multiple sources, source-by-source analysis is a good idea. Otherwise, uncertainty, inconsistency, and other computational complexities cannot be handled with the required transparency and accuracy.

For example, consider the combination (*CV* = Gap Voltage (*V*), *EV* = *MRR*). Figure 18(*a*) shows a scatter plot of all relevant datasets taken from S_1, S_2, and S_4. Figures 18(*b*)-(*d*) show the scatter plots of relevant datasets taken from individual sources, S_1, S_2, and S_4, respectively. The trend lines are also shown in the respective plots. As seen in Figures 18(*b*)-(*d*), even though a consistent trend exists across the sources, a huge amount of uncertainty persists, as well. It is worth mentioning that the values of *MRR* are not consistent across the sources (compare the plots in Figure 18(*b*)-(*d*)).



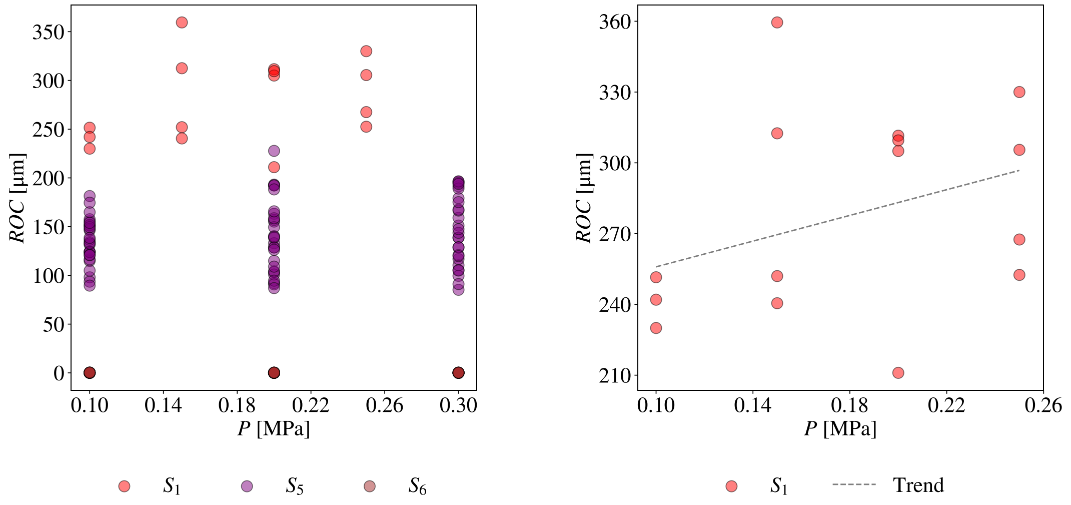
(*a*) (*b*)



(*c*) (*d*)

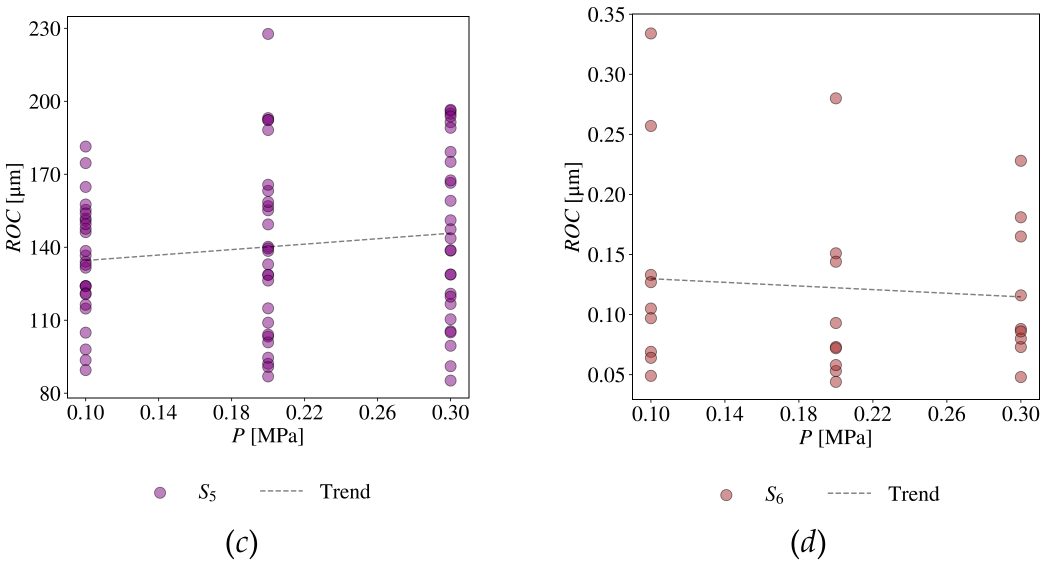
Figure 18. Gap Voltage (*V*) and *MRR*-relevant datasets.

Lastly, consider the *CV-EV* combination of Gas Pressure (*P*) and Radial Over Cut (*ROC*). Figure 19(*a*) shows a scatter plot of all relevant datasets taken from *S*₁, *S*₅, and *S*₆. Figures 19(*b*), (*c*), and (*d*) show the scatter plots of relevant datasets taken from individual sources, *S*₁, *S*₅, and *S*₆, respectively. The trend lines are also shown in the respective plots. As seen in Figures 19(*b*)-(*d*), there is an inconsistency in the trend; *S*₆ exhibits a different trend than the others. Moreover, the values of *ROC* of *S*₆ are significantly different than those of others. Similar to the previous case, a huge amount of uncertainty persists here, too.



(*a*) (*b*)

doi:10.20944/preprints202307.1199.v1



(*c*) (*d*)

Figure 19. Gas Pressure (*P*) and Radial Over Cut (*ROC*)-relevant datasets.

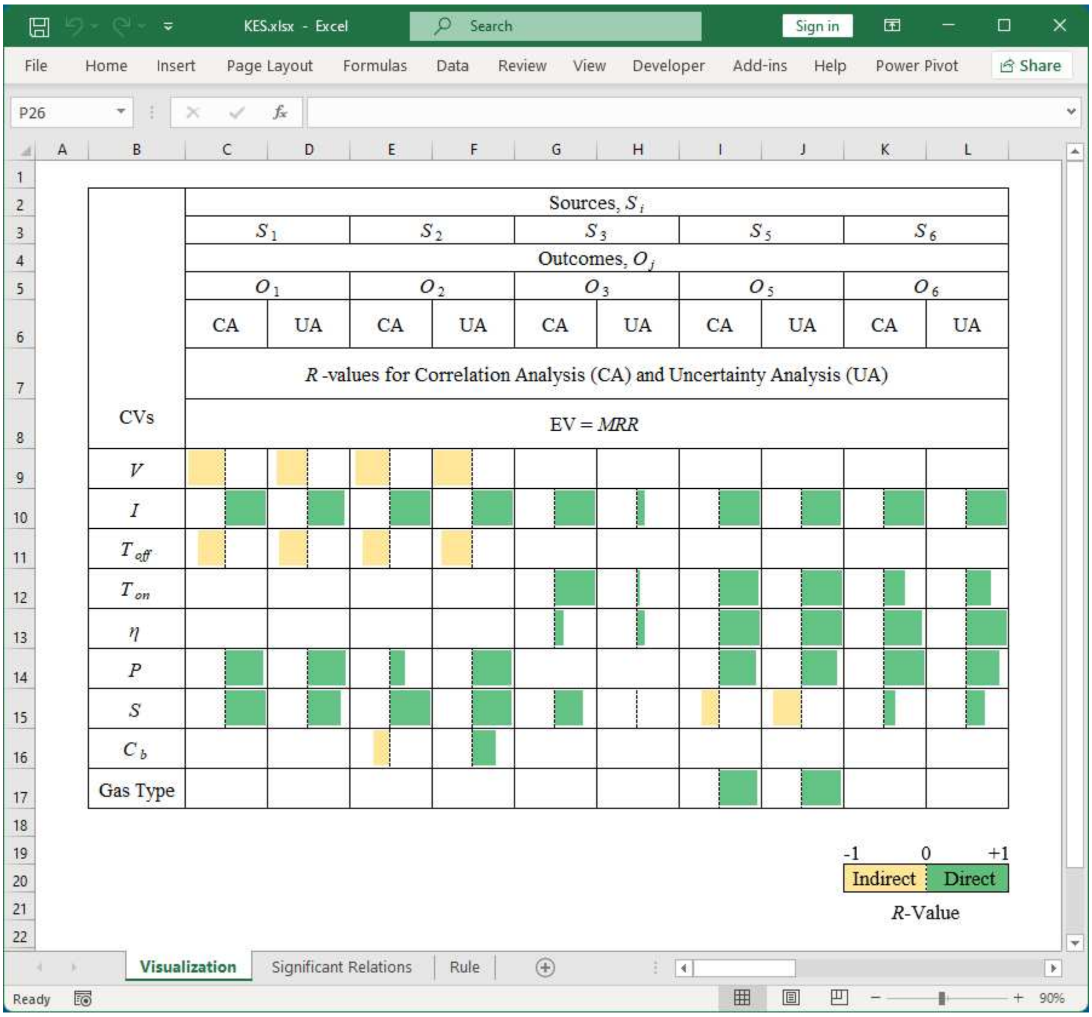
A personalized Knowledge Extraction System (KES) is developed using spreadsheet applications, keeping the computational complexity described above in mind. The system uses the relationships among *CV*s and *EV*s found in the previous system and determines the right set of *CV*s to achieve a given objective (e.g., maximizing *MRR*). The results regarding the optimization of *MRR* are reported below. Table 2 reports the values of correlation coefficient (*R*) (denoted as *R*) in the interval [-1,1]. The values of *R* are calculated for both options, Correlation Analysis (CA) and Uncertainty Analysis (UA), for all possible *CV-EV* combinations (Table 1). The remarkable thing is that *S*4 is kept aside because *CV*s have only two states. As a result, Correlation Analysis (CA) and Uncertainty Analysis (UA) produce only two points for each *CV-EV* combination, and, thereby, *R* = -1 or 1. Thus, including these kinds of datasets may produce a misleading conclusion.

Table 2. Exported outcomes from DAS for knowledge extraction.

	Sources, *S_i*									
	***S*1**		***S*2**		***S*3**		***S*5**		*S*6	
	Outcomes, *O_j*									
	***O*1**		***O*2**		***O*3**		***O*5**		*O*6	
	CA	**UA**	**CA**	**UA**	**CA**	**UA**	**CA**	**UA**	CA	UA
	***R*-values for Correlation Analysis (CA) and Uncertainty Analysis (UA)**									
	EV* = *MRR									
CVs										
V	**-0.904**	**-0.75**	**-0.852**	**-0.957**						
I	0.996	0.917	0.997	0.993	0.989	0.195	0.999	0.972	0.995	0.998
T_{off}	-0.653	-0.69	-0.661	-0.754						
T_{on}					0.989	0.071	0.978	0.993	0.523	0.593
η					0.222	0.181	0.986	0.99	0.944	0.987
P	0.93	0.926	0.369	0.965			0.902	0.863	0.996	0.798
S	0.993	0.826	1	0.957	0.707	-0.009	-0.419	-0.691	0.265	0.452
C_b			-0.397	0.585						
Gas Type							0.934	0.967		

V: Voltage, *I*: Current, *T_{off}*: Pulse off time, *T_{on}*: Pulse on time, *P*: Gas Pressure, *N*: Spindle Rotational Speed, *C_b*: Shielding Clearance

The degree of correlation (given by *R* values) is visualized using an Excel™-based system, as shown in Figure 20. Here, a green-colored box means the corresponding *CV-EV* pair maintains a direct or positive relationship, and a yellow-colored box means the corresponding *CV-EV* pair maintains an indirect or negative relationship. The length of the colored bar indicates the strength of *CV-EV* relationships. The longer the length, the stronger the relationship.

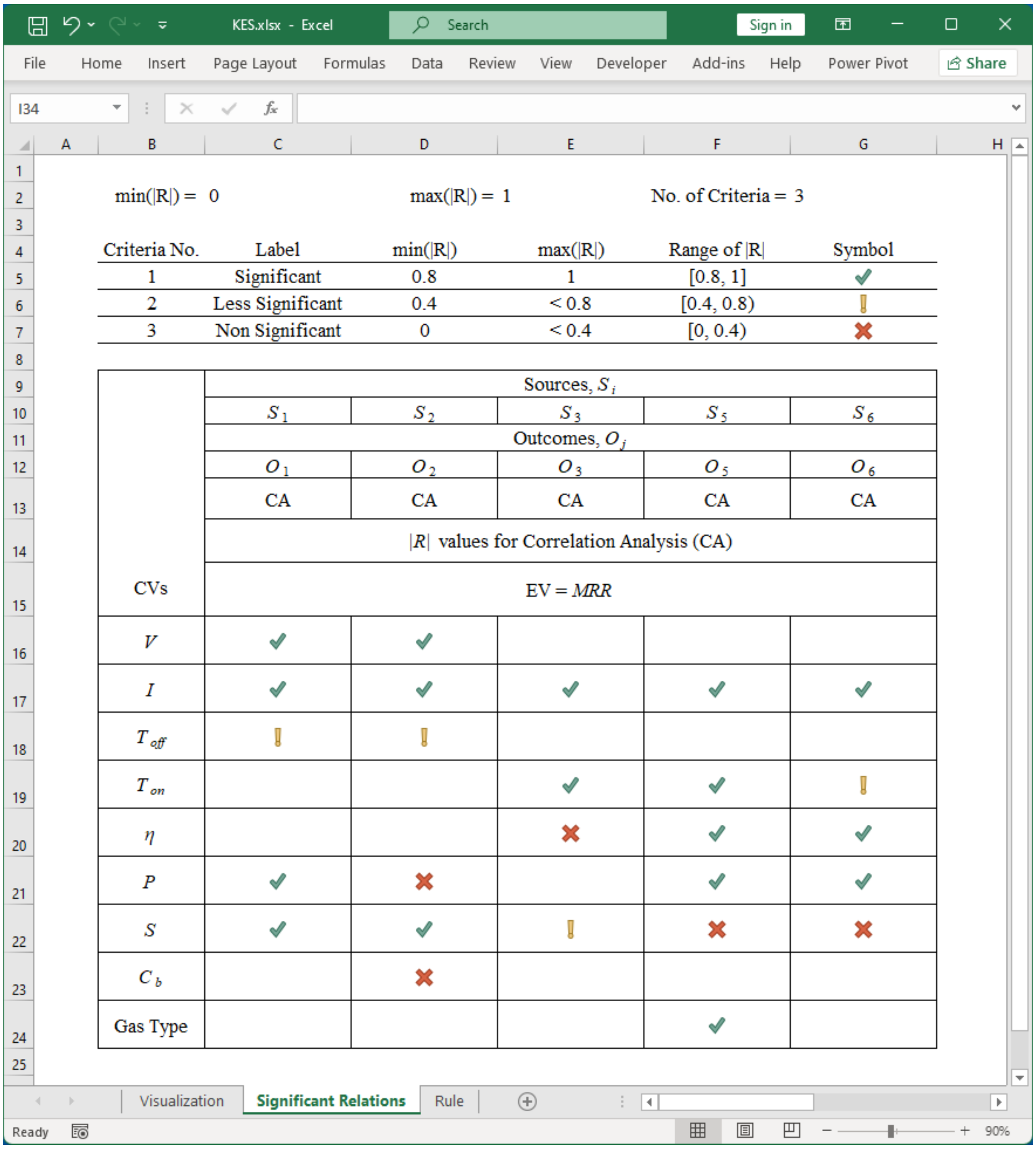


CVs	Sources, S_i									
	S_1		S_2		S_3		S_5		S_6	
	Outcomes, O_j									
	O_1		O_2		O_3		O_5		O_6	
	CA	UA	CA	UA	CA	UA	CA	UA	CA	UA
	R -values for Correlation Analysis (CA) and Uncertainty Analysis (UA)									
	EV = *MRR*									
V	🟨	🟨	🟨	🟨						
I	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
T_{off}	🟨	🟨	🟨	🟨						
T_{on}					🟩	🟩	🟩	🟩	🟩	🟩
η					🟩	🟩	🟩	🟩	🟩	🟩
P	🟩	🟩	🟩	🟩			🟩	🟩	🟩	🟩
S	🟩	🟩	🟩	🟩	🟩		🟨	🟨	🟩	🟩
C_b			🟨	🟩						
Gas Type							🟩	🟩		

-1 0 +1
🟨 Indirect 🟩 Direct
R-Value

Figure 20. Screen-print of the KES for visualizing the analyzed outcomes.

The absolute value of *R*, i.e., |*R*|, can be divided into few states to visualize more clearly the impact of the *CV*s on the given *EV* (this time *MRR*). The results shown in Figure 21 refer to three states, as follows: 1) |*R*| ∈ [0.8, 1] means "significant,"2) |*R*| ∈ [0.4, 0.8) means "less significant," and 3) |*R*| ∈ [0, 0.4) means "non-significant." These states are shown by a green-colored-tick-mark (√), gold-colored-exclamatory-mark (!), and red-colored-cross-mark (X), respectively. Observing the green-colored tick-mark symbols (√) makes it possible to identify a set of rules for maximizing *MRR*. The results are shown in Figure 22 and summarized in Table 3.

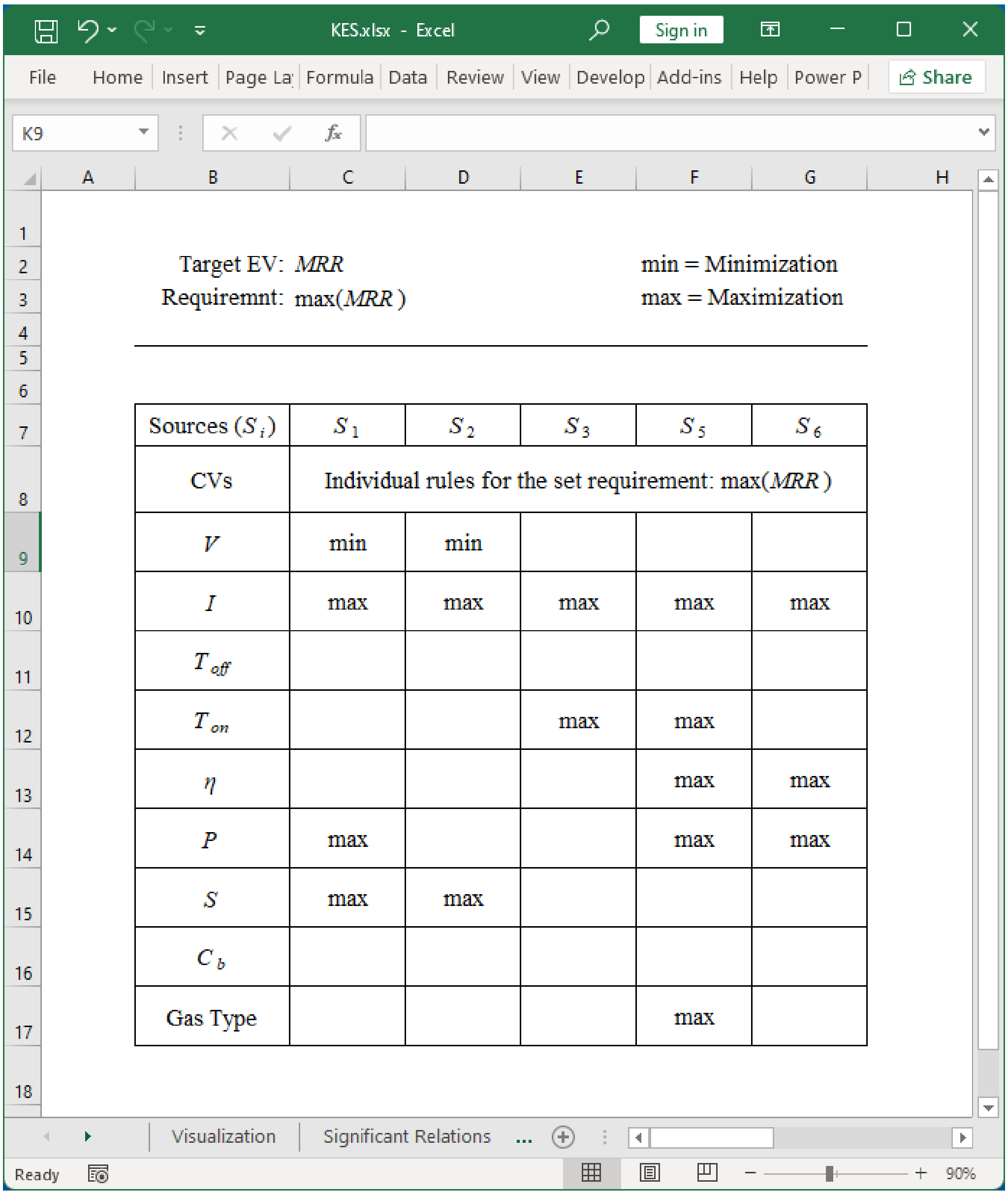


min(|R|) = 0 max(|R|) = 1 No. of Criteria = 3

Criteria No.	Label	min(|R|)	max(|R|)	Range of |R|	Symbol
1	Significant	0.8	1	[0.8, 1]	✓
2	Less Significant	0.4	< 0.8	[0.4, 0.8)	!
3	Non Significant	0	< 0.4	[0, 0.4)	✗

CVs	Sources, S_i				
	S_1	S_2	S_3	S_5	S_6
	Outcomes, O_j				
	O_1	O_2	O_3	O_5	O_6
	CA	CA	CA	CA	CA
	$\|R\|$ values for Correlation Analysis (CA)				
	EV = *MRR*				
V	✓	✓			
I	✓	✓	✓	✓	✓
T_{off}	!	!			
T_{on}			✓	✓	!
η			✗	✓	✓
P	✓	✗		✓	✓
S	✓	✓	!	✗	✗
C_b		✗			
Gas Type				✓	

Figure 21. Screen-print of the KES for identifying the significant relationships.



Target EV: *MRR*
Requiremnt: max(*MRR*)

min = Minimization
max = Maximization

Sources (S_i)	S_1	S_2	S_3	S_5	S_6
CVs	Individual rules for the set requirement: max(*MRR*)				
V	min	min			
I	max	max	max	max	max
T_{off}					
T_{on}			max	max	
η				max	max
P	max			max	max
S	max	max			
C_b					
Gas Type				max	

Figure 22. Screen-print of the KES for rule extraction.

Table 3. Extracted rules.

S_i	R_k	Rules
S_1	R_1	$min(V) \wedge max(I) \wedge max(P) \wedge max(S) \Rightarrow max(MRR)$
S_2	R_2	$min(V) \wedge max(I) \wedge max(S) \Rightarrow max(MRR)$
S_3	R_3	$max(I) \wedge max(T_{on}) \Rightarrow max(MRR)$
S_5	R_5	$max(I) \wedge max(T_{on}) \wedge max(\eta) \wedge max(P) \wedge max(Gas\ Type) \Rightarrow max(MRR)$
S_6	R_6	$max(I) \wedge max(\eta) \wedge max(P) \Rightarrow max(MRR)$

Whether or not the rules produce meaningful results is tested by applying the rules to datasets of $S_1,\ldots,S_6$. Let *MRR′* be the *MRR* corresponding to a rule and let *MRR′′* be the maximum possible *MRR*, for a particular source. The results are summarized in Table 4. As seen in Table 4, for S_1, S_2, and S_6, *MRR′* and *MRR′′* are the same. In particular, for S_1, *MRR′* = *MRR′′* = 1.497 mm3/min. For S_2, *MRR′* = *MRR′′* = 0.811 mm3/min. For S_6, *MRR′* = MRR′′ = 5.31 mm3/min. This suggests that the extracted rules are effective for maximizing the *MRR*. On the other hand, for S_3 and S_5, the rules do not refer to any available datasets. This is perhaps because the rules use many or very few conditions, i.e., a moderate number of *CV*s can be used to achieve the goal (here, maximizing *MRR*).

Table 4. Validation of R_k.

S_i	R_k	MRR' (mm³/min)	MRR'' (mm³/min)
S_1	R_1	1.497	1.497
S_2	R_2	0.811	0.811
S_3	R_3	No conclusion	9.9425
S_5	R_5	No conclusion	3.77
S_6	R_6	5.31	5.31

It is worth mentioning that the proposed BDA and digital twin of a manufacturing process has close connections. Since a digital twin consists of input, modeling, simulation, validation, and output modules (see [69]), the outcomes of BDA, e.g., the rules listed in Table 3, (e.g., $min(V) \wedge max(I) \wedge max(P) \wedge max(S) \Rightarrow max(MRR)$)) can be injected into an appropriate module of digital twin. This way BD and digital twin, two vital constituents of smart manufacturing can work in a synergetic manner.

6. Conclusions

Big data analytics is one of the essential constituents of smart manufacturing. Unfortunately, there is no systematic approach to developing it. This paper sheds some light on this issue. This paper first presents a comprehensive literature review on smart manufacturing-relevant big data analytics. Afterward, this paper presents a systematic approach to developing big data analytics for manufacturing process-relevant decision-making activities.

The proposed analytics consists of five integrated system components:

- Big data preparation system
- Big data exploration system
- Data visualization system
- Data analysis system
- Knowledge extraction system

The functional requirements of the systems are as follows.

First, the big data preparation system must prepare contents to be included in big data. The contents may exhibit the characteristics of the so-called Digital Manufacturing Commons (DMC). Thus, it is desirable that the system supports user-defined ontologies and produces widely acceptable digital datasets using Extensible Markup Language (XML). The big data exploration system can extract relevant datasets prepared by the first system. The system uses keywords derived from the names of manufacturing processes, materials, and analyses- or experiments-relevant phrases (e.g., design of experiment). The third system can help visualize relevant datasets extracted by the second system using suitable methods (e.g., scatter plots and possibility distributions). The fourth system must establish relationships among the relevant control variables (variables that can be adjusted as needed) and evaluation variables (variables that measure the performance) combinations for a given situation. In addition, it must quantify the uncertainty in the relationships. Finally, the last system can extract knowledge from the outcomes of the fourth system using user-defined criteria (e.g., minimize surface roughness, maximize material removal rate, and alike). In addition, JAVA™- and spreadsheet-based systems are developed to realize the proposed integrated systems.

The efficacy of proposed analytics is demonstrated using a case study where the goal is to determine the right states of control variables of dry electrical discharge machining for maximizing the material removal rate. The contents are created from published scientific articles on dry electrical discharge machining that deals with stainless and high-speed steels. In addition, the articles that presented datasets based on the design of experiments are considered. The datasets collectively underlie the following control variables: voltage, current, pulse-off time, pulse-on time, gas pressure, rotational speed, shielding clearance, duty factor, and gas type. The set of control variables differs from article to article.

Consequently, the values of the control variables differ from article to article. In addition, the degree of uncertainty in the datasets differs from article to article. This heterogeneous situation was successfully analyzed using the proposed analytics. The analytics successfully determined which variables among voltage, current, pulse-off time, gas pressure, and rotational speed effectively maximize material removal rate. In addition, the underlying uncertainty is also quantified.

In some cases, scatter plots are effective for the analysis, and in others, possibility distribution is effective. The analytics helps identify the redundant, less effective, and most effective variables by which one can maximize the material removal rate. The knowledge extracted can be used to optimize a dry electrical discharge machining operation and elucidate the research gap in dry electrical discharge machining.

Although the system is implemented for EDM, it can easily be implemented in other manufacturing processes. The reason is that all manufacturing processes are operated by fixing some control variables against some evolution variables. For example, the turning control variable can be feed rate, depth of cut, cutting velocity, and tool nose radius. Likewise, the possible list of evaluation variables is surface roughness, tool wear, and material removal rate. This means the same BDA can be used for turning effortlessly. The user consults the datasets relevant to the control and evaluation variables.

The remarkable thing is that the intervention and settings of a user and underlying computational aspects are transparent. At the same time, it does not require any sophisticated or expensive resources. Thus, the proposed analytics exhibits desirable characteristics regarding big data inequality and transparency issues. This experience can be extended to developing big data analytics and digital twins for smart manufacturing.

Nevertheless, other relevant technical issues can be delved into in the next phase of research. One of them is the issue of security. Consequently, as reviewed in [70], blockchain-based technology can be considered. Particularly, blockchain technology can be integrated with the Big Data Preparation System (BDPS) to make the machine-readable datasets trusted and secured from the very beginning.

Author Contributions: Conceptualization, A.K.G., S.F. and S.U.; methodology, A.K.G., S.F. and S.U.; software, A.K.G. and S.F.; validation, A.K.G. and S.U.; formal analysis, A.K.G., S.F. and S.U.; investigation, A.K.G., S.F. and S.U.; resources, S.U.; data curation, A.K.G. and S.F.; writing—original draft preparation, S.F. and S.U.; writing—review and editing, A.K.G., S.F. and S.U.; visualization, A.K.G. and S.F.; supervision, S.U.; project administration, S.U.; funding acquisition, S.U.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable. We encourage all authors of articles published in MDPI journals to share their research data. In this section, please provide details regarding where data supporting reported results can be found, including links to publicly archived datasets analyzed or generated during the study. Where no new data were created, or where data is unavailable due to privacy or ethical restrictions, a statement is still required. Suggested Data Availability Statements are available in section "MDPI Research Data Policies" at https://www.mdpi.com/ethics.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. F. X. Diebold, "On the Origin(s) and Development of the Term 'Big Data,'" *SSRN Electron. J.*, Sep. 2012. https://doi.org/10.2139/ssrn.2152421.
2. G. Shah, A. Shah, and M. Shah, "Panacea of challenges in real-world application of big data analytics in healthcare sector," *J. Data, Inf. Manag.*, vol. 1, no. 3–4, pp. 107–116, Dec. 2019. https://doi.org/10.1007/s42488-019-00010-1.
3. H. Hassani, X. Huang, and E. Silva, "Banking with blockchain-ed big data," *J. Manag. Anal.*, vol. 5, no. 4, pp. 256–275, Oct. 2018. https://doi.org/10.1080/23270012.2018.1528900.

4. G. G. Hallur, S. Prabhu, and A. Aslekar, "Entertainment in Era of AI, Big Data &amp; IoT," in *Digital Entertainment*, S. Das and S. Gochhait, Eds. Singapore: Springer Nature Singapore, 2021, pp. 87–109. https://doi.org/10.1007/978-981-15-9724-4_5.
5. Y. Wang, "Big Opportunities and Big Concerns of Big Data in Education," *TechTrends*, vol. 60, no. 4, pp. 381–384, Jul. 2016. https://doi.org/10.1007/s11528-016-0072-1.
6. L. Zhu, F. R. Yu, Y. Wang, B. Ning, and T. Tang, "Big Data Analytics in Intelligent Transportation Systems: A Survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 1, pp. 383–398, Jan. 2019. https://doi.org/10.1109/TITS.2018.2815678.
7. N. Henke *et al.*, "The age of analytics: Competing in a data-driven world," Dec. 2016. [Online]. Available: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-age-of-analytics-competing-in-a-data-driven-world
8. S. Yin and O. Kaynak, "Big Data for Modern Industry: Challenges and Trends [Point of View]," *Proc. IEEE*, vol. 103, no. 2, pp. 143–146, Feb. 2015. https://doi.org/10.1109/JPROC.2015.2388958.
9. L. Andre, "53 Important Statistics About How Much Data Is Created Every Day." [Online]. Available: https://financesonline.com/how-much-data-is-created-every-day/
10. R. R. Sreenivasan, "Characteristics of big data - a Delphi study," Memorial University of Newfoundland, Newfoundland, Canada, 2017. Accessed: Sep. 12, 2022. [Online]. Available: https://research.library.mun.ca/13080/
11. T. Verevka, A. Mirolyubov, and J. Makio, "Opportunities and Barriers to Using Big Data Technologies in the Metallurgical Industry," in *Innovations in Digital Economy*, D. Rodionov, T. Kudryavtseva, A. Skhvediani, and M. A. Berawi, Eds. Cham: Springer International Publishing, 2021, pp. 86–102. https://doi.org/10.1007/978-3-030-84845-3_6.
12. S. Fattahi and A. S. Ullah, "Optimization of Dry Electrical Discharge Machining of Stainless Steel using Big Data Analytics," *Procedia CIRP*, vol. 112, pp. 316–321, 2022. https://doi.org/10.1016/j.procir.2022.09.004.
13. W. Chang, "NIST Big Data Reference Architecture for Analytics and Beyond," in *Proceedings of the10th International Conference on Utility and Cloud Computing*, Austin, Texas, USA, Dec. 2017, pp. 3–3. https://doi.org/10.1145/3147213.3155013.
14. D. and T. S. NIST Big Data Public Working Group, "NIST Big Data Interoperability Framework:," Gaithersburg, MD, Oct. 2019. https://doi.org/10.6028/NIST.SP.1500-1r2.
15. NIST Big Data Public Working Group: Definitions and Taxonomies Subgroup, "NIST Big Data Interoperability Framework: Volume 2, Big Data Taxonomies," Gaithersburg, MD, Nov. 2019. https://doi.org/10.6028/NIST.SP.1500-2r2.
16. Wo Chang and Geoffrey Fox, "NIST Big Data Interoperability Framework: Volume 3, Use Cases and General Requirements," Gaithersburg, MD, Oct. 2019. https://doi.org/10.6028/NIST.SP.1500-3r2.
17. Wo Chang, Arnab Roy, and Mark Underwood, "NIST Big Data Interoperability Framework: Volume 4, Security and Privacy," Gaithersburg, MD, Oct. 2019. https://doi.org/10.6028/NIST.SP.1500-4r2.
18. Wo Chang, Sanjay Mishra, and NBD-PWG NIST, "NIST Big Data Interoperability Framework: Volume 5, Architectures White Paper Survey," Gaithersburg, MD, Oct. 2015. https://doi.org/10.6028/NIST.SP.1500-5.
19. Wo Chang, David Boyd, and Orit Levin, "NIST Big Data Interoperability Framework: Volume 6, Reference Architecture," Gaithersburg, MD, Oct. 2019. https://doi.org/10.6028/NIST.SP.1500-6r2.
20. W. Chang and G. von Laszewski, "NIST Big Data Interoperability Framework: Volume 8, Reference Architecture Interfaces," Gaithersburg, MD, Jun. 2018. https://doi.org/10.6028/NIST.SP.1500-9.
21. NIST Big Data Public Working Group, "NIST Big Data Interoperability Framework: Volume 7, Standards Roadmap, Version 2," Gaithersburg, MD, Jun. 2018. https://doi.org/10.6028/NIST.SP.1500-7r1.
22. W. Chang, C. C. Austin, and R. Reinsch, "NIST Big Data Interoperability Framework: Volume 9, Adoption and Modernization," Gaithersburg, MD, Oct. 2019. https://doi.org/10.6028/NIST.SP.1500-10r1.
23. M. Farboodi, R. Mihet, T. Philippon, and L. Veldkamp, "Big Data and Firm Dynamics," *AEA Pap. Proc.*, vol. 109, pp. 38–42, May 2019. https://doi.org/10.1257/pandp.20191001.
24. J. Cinnamon, "Data inequalities and why they matter for development," *Inf. Technol. Dev.*, vol. 26, no. 2, pp. 214–233, Apr. 2020. https://doi.org/10.1080/02681102.2019.1650244.
25. S. Qureshi, "Overcoming Technological Determinism in Understanding the Digital Divide: Where Do We Go From Here?," *Inf. Technol. Dev.*, vol. 20, no. 3, pp. 215–217, Jul. 2014. https://doi.org/10.1080/02681102.2014.930981.
26. M. Hilbert, "Big Data for Development: A Review of Promises and Challenges," *Dev. Policy Rev.*, vol. 34, no. 1, pp. 135–174, Jan. 2016. https://doi.org/10.1111/dpr.12142.
27. M. Favaretto, E. De Clercq, and B. S. Elger, "Big Data and discrimination: perils, promises and solutions. A systematic review," *J. Big Data*, vol. 6, no. 1, p. 12, Dec. 2019. https://doi.org/10.1186/s40537-019-0177-4.
28. D. Shah, J. Wang, and Q. P. He, "Feature engineering in big data analytics for IoT-enabled smart manufacturing – Comparison between deep learning and statistical learning," *Comput. Chem. Eng.*, vol. 141, p. 106970, Oct. 2020. https://doi.org/10.1016/j.compchemeng.2020.106970.

29. P. C. Evans and M. Annunziata, “Industrial Internet: Pushing the Boundaries of Minds and Machines,” Jan. 2012. [Online]. Available: https://www.researchgate.net/publication/271524319_Industrial_Internet_Pushing_the_boundaries_of_minds_and_machines
30. A. G. Banerjee *et al.*, “Cloud Computing-Based Marketplace for Collaborative Design and Manufacturing,” in *Internet of {Things}. {IoT} {Infrastructures}*, B. Mandler, J. Marquez-Barja, M. E. Mitre Campista, D. Cagáňová, H. Chaouchi, S. Zeadally, M. Badra, S. Giordano, M. Fazio, A. Somov, and R.-L. Vieriu, Eds. Cham: Springer International Publishing, 2016, pp. 409–418. https://doi.org/10.1007/978-3-319-47063-4_42.
31. D. Wu, D. W. Rosen, L. Wang, and D. Schaefer, “Cloud-based design and manufacturing: A new paradigm in digital manufacturing and design innovation,” *Comput. Des.*, vol. 59, pp. 1–14, Feb. 2015. https://doi.org/10.1016/j.cad.2014.07.006.
32. B. Beckmann, A. Giani, J. Carbone, P. Koudal, J. Salvo, and J. Barkley, “Developing the Digital Manufacturing Commons: A National Initiative for US Manufacturing Innovation,” *Procedia Manuf.*, vol. 5, pp. 182–194, 2016. https://doi.org/10.1016/j.promfg.2016.08.017.
33. F. Samimi, P. Mckinley, S. Sadjadi, C. Tang, J. Shapiro, and Z. Zhou, “Service Clouds: Distributed Infrastructure for Adaptive Communication Services,” *IEEE Trans. Netw. Serv. Manag.*, vol. 4, no. 2, pp. 84–95, Sep. 2007. https://doi.org/10.1109/TNSM.2007.070901.
34. Z. Bi, Y. Jin, P. Maropoulos, W.-J. Zhang, and L. Wang, “Internet of things (IoT) and big data analytics (BDA) for digital manufacturing (DM),” *Int. J. Prod. Res.*, pp. 1–18, Jul. 2021. https://doi.org/10.1080/00207543.2021.1953181.
35. J. Wang, C. Xu, J. Zhang, and R. Zhong, “Big data analytics for intelligent manufacturing systems: A review,” *J. Manuf. Syst.*, vol. 62, pp. 738–752, Jan. 2022. https://doi.org/10.1016/j.jmsy.2021.03.005.
36. S. Kahveci, B. Alkan, M. H. Ahmad, B. Ahmad, and R. Harrison, “An end-to-end big data analytics platform for IoT-enabled smart factories: A case study of battery module assembly system for electric vehicles,” *J. Manuf. Syst.*, vol. 63, pp. 214–223, Apr. 2022. https://doi.org/10.1016/j.jmsy.2022.03.010.
37. S. Fattahi, S. Ura, and M. Noor-E-Alam, “Decision-Making Using Big Data Relevant to Sustainable Development Goals (SDGs),” *Big Data Cogn. Comput.*, vol. 6, no. 2, p. 64, Jun. 2022. https://doi.org/10.3390/bdcc6020064.
38. T. Chen and Y.-C. Wang, “Hybrid big data analytics and Industry 4.0 approach to projecting cycle time ranges,” *Int. J. Adv. Manuf. Technol.*, vol. 120, no. 1–2, pp. 279–295, May 2022. https://doi.org/10.1007/s00170-022-08733-z.
39. J. Woo, S.-J. Shin, W. Seo, and P. Meilanitasari, “Developing a big data analytics platform for manufacturing systems: architecture, method, and implementation,” *Int. J. Adv. Manuf. Technol.*, vol. 99, no. 9–12, pp. 2193–2217, Dec. 2018. https://doi.org/10.1007/s00170-018-2416-9.
40. R. Bonnard, M. D. S. Arantes, R. Lorbieski, K. M. M. Vieira, and M. C. Nunes, “Big data/analytics platform for Industry 4.0 implementation in advanced manufacturing context,” *Int. J. Adv. Manuf. Technol.*, vol. 117, no. 5–6, pp. 1959–1973, Nov. 2021. https://doi.org/10.1007/s00170-021-07834-5.
41. D. Kozjek, R. Vrabič, B. Rihtaršič, N. Lavrač, and P. Butala, “Advancing manufacturing systems with big-data analytics: A conceptual framework,” *Int. J. Comput. Integr. Manuf.*, vol. 33, no. 2, pp. 169–188, Feb. 2020. https://doi.org/10.1080/0951192X.2020.1718765.
42. C. Jun, J. Y. Lee, and B. H. Kim, “Cloud-based big data analytics platform using algorithm templates for the manufacturing industry,” *Int. J. Comput. Integr. Manuf.*, vol. 32, no. 8, pp. 723–738, Aug. 2019. https://doi.org/10.1080/0951192X.2019.1610578.
43. R. Dubey *et al.*, “Big data analytics and artificial intelligence pathway to operational performance under the effects of entrepreneurial orientation and environmental dynamism: A study of manufacturing organisations,” *Int. J. Prod. Econ.*, vol. 226, p. 107599, Aug. 2020. https://doi.org/10.1016/j.ijpe.2019.107599.
44. C. Zhang, Z. Wang, K. Ding, F. T. S. Chan, and W. Ji, “An energy-aware cyber physical system for energy Big data analysis and recessive production anomalies detection in discrete manufacturing workshops,” *Int. J. Prod. Res.*, vol. 58, no. 23, pp. 7059–7077, Dec. 2020. https://doi.org/10.1080/00207543.2020.1748904.
45. R. Y. Zhong, S. T. Newman, G. Q. Huang, and S. Lan, “Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives,” *Comput. Ind. Eng.*, vol. 101, pp. 572–591, Nov. 2016. https://doi.org/10.1016/j.cie.2016.07.013.
46. R. Y. Zhong, C. Xu, C. Chen, and G. Q. Huang, “Big Data Analytics for Physical Internet-based intelligent manufacturing shop floors,” *Int. J. Prod. Res.*, vol. 55, no. 9, pp. 2610–2621, May 2017. https://doi.org/10.1080/00207543.2015.1086037.
47. Y. Zhang, S. Ren, Y. Liu, and S. Si, “A big data analytics architecture for cleaner manufacturing and maintenance processes of complex products,” *J. Clean. Prod.*, vol. 142, pp. 626–641, Jan. 2017. https://doi.org/10.1016/j.jclepro.2016.07.123.
48. Y. Lu and X. Xu, “Cloud-based manufacturing equipment and big data analytics to enable on-demand manufacturing services,” *Robot. Comput. Integr. Manuf.*, vol. 57, pp. 92–102, Jun. 2019. https://doi.org/10.1016/j.rcim.2018.11.006.

49. W. Ji and L. Wang, "Big data analytics based fault prediction for shop floor scheduling," *J. Manuf. Syst.*, vol. 43, pp. 187–194, Apr. 2017. https://doi.org/10.1016/j.jmsy.2017.03.008.
50. Y. C. Liang, X. Lu, W. D. Li, and S. Wang, "Cyber Physical System and Big Data enabled energy efficient machining optimisation," *J. Clean. Prod.*, vol. 187, pp. 46–62, Jun. 2018. https://doi.org/10.1016/j.jclepro.2018.03.149.
51. W. Ji, S. Yin, and L. Wang, "A big data analytics based machining optimisation approach," *J. Intell. Manuf.*, vol. 30, no. 3, pp. 1483–1495, Mar. 2019. https://doi.org/10.1007/s10845-018-1440-9.
52. C.-C. Chen *et al.*, "A Novel Efficient Big Data Processing Scheme for Feature Extraction in Electrical Discharge Machining," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 910–917, Apr. 2019. https://doi.org/10.1109/LRA.2019.2891498.
53. S. Fattahi, T. Okamoto, and S. Ura, "Preparing Datasets of Surface Roughness for Constructing Big Data from the Context of Smart Manufacturing and Cognitive Computing," *Big Data Cogn. Comput.*, vol. 5, no. 4, p. 58, Oct. 2021. https://doi.org/10.3390/bdcc5040058.
54. C. Li, Y. Chen, and Y. Shang, "A review of industrial big data for decision making in intelligent manufacturing," *Eng. Sci. Technol. an Int. J.*, vol. 29, p. 101021, May 2022. https://doi.org/10.1016/j.jestch.2021.06.001.
55. V. Nasir and F. Sassani, "A review on deep learning in machining and tool monitoring: methods, opportunities, and challenges," *Int. J. Adv. Manuf. Technol.*, vol. 115, no. 9–10, pp. 2683–2709, Aug. 2021. https://doi.org/10.1007/s00170-021-07325-7.
56. M. Hildebrandt and B.-J. Koops, "The Challenges of Ambient Law and Legal Protection in the Profiling Era," *Mod. Law Rev.*, vol. 73, no. 3, pp. 428–460, Sep. 2010, [Online]. Available: http://www.jstor.org/stable/40660735
57. F. Emmert-Streib, O. Yli-Harja, and M. Dehmer, "Artificial Intelligence: A Clarification of Misconceptions, Myths and Desired Status," *Front. Artif. Intell.*, vol. 3, Dec. 2020. https://doi.org/10.3389/frai.2020.524339.
58. D. K. Citron and F. A. Pasquale, "The Scored Society: Due Process for Automated Predictions," *Washingt. Law Rev.*, vol. 89, 2014.
59. B. d'Alessandro, C. O'Neil, and T. LaGatta, "Conscientious Classification: A Data Scientist's Guide to Discrimination-Aware Classification," *Big Data*, vol. 5, no. 2, pp. 120–134, Jun. 2017. https://doi.org/10.1089/big.2016.0048.
60. M. Leese, "The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union," *Secur. Dialogue*, vol. 45, no. 5, pp. 494–511, Oct. 2014. https://doi.org/10.1177/0967010614544204.
61. A. M. M. S. Ullah and K. H. Harib, "A human-assisted knowledge extraction method for machining operations," *Adv. Eng. Informatics*, vol. 20, no. 4, pp. 335–350, Oct. 2006. https://doi.org/10.1016/j.aei.2006.07.004.
62. A. M. M. Sharif Ullah and M. Shamsuzzaman, "Fuzzy Monte Carlo Simulation using point-cloud-based probability–possibility transformation," *Simulation*, vol. 89, no. 7, pp. 860–875, Jul. 2013. https://doi.org/10.1177/0037549713482174.
63. G. Puthumana and S. S. Joshi, "Investigations into performance of dry EDM using slotted electrodes," *Int. J. Precis. Eng. Manuf.*, vol. 12, no. 6, pp. 957–963, Dec. 2011. https://doi.org/10.1007/s12541-011-0128-2.
64. P. Govindan and S. S. Joshi, "Experimental characterization of material removal in dry electrical discharge drilling," *Int. J. Mach. Tools Manuf.*, vol. 50, no. 5, pp. 431–443, May 2010. https://doi.org/10.1016/j.ijmachtools.2010.02.004.
65. R. T. Murickan, L. P. Jakkamputi, and P. Kuppan, "Experimental investigation of Dry Electrical Discharge Machining on SS 316L," *Int. J. Latest Trends Eng. Technol.*, vol. 2, no. 3, pp. 100–107, 2013, [Online]. Available: http://www.ijltet.org/wp-content/uploads/2013/06/14.pdf
66. G. Puthumana, R. Agarwal, and S. S. Joshi, "Experimental investigation on dry electrical discharge machining using helium gas," in *proceedings of the 3rd International & 24th AIMTDR (All India Manufacturing Technology, Design and Research)*, 13-15 December 2010, Visakhapatnam, India. Available: https://backend.orbit.dtu.dk/ws/portalfiles/portal/123939189/AIMTDR2010.pdf
67. S. Fattahi and H. Baseri, "Analysis of dry electrical discharge machining in different dielectric mediums," *Proc. Inst. Mech. Eng. Part E J. Process Mech. Eng.*, vol. 231, no. 3, pp. 497–512, Jun. 2017. https://doi.org/10.1177/0954408915611540.
68. S. Fattahi, I. Shyha, and H. Baseri, "Optimisation of Dry Electrical Discharge Machining of High Speed Steel using Grey-Relational Analysis," *Int. J. Robot. Mechatronics*, vol. 2, no. 4, pp. 132–139, Dec. 2015. https://doi.org/10.21535/ijrm.v2i4.886.
69. A. K. Ghosh, AMM S. Ullah, R. Teti, A. Kubo, Developing sensor signal-based digital twins for intelligent machine tools, *Journal of Industrial Information Integration*, Volume 24, 2021, 100242. https://doi.org/10.1016/j.jii.2021.100242.

70. J. Leng, G. Ruan, P. Jiang, K. Xu, Q. Liu, X. Zhou, C. Liu, Blockchain-empowered sustainable manufacturing and product lifecycle management in industry 4.0: A survey, *Renewable and Sustainable Energy Reviews*, Volume 132, 2020, 110112. https://doi.org/10.1016/j.rser.2020.110112.