

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

² Advanced Manufacturing Engineering Laboratory, Kitami Institute of Technology, 165 Koen-cho, Kitami 090-8507, Japan; saman.faattahi@gmail.com

* 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 consist 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 system components are elucidated. In addition, JAVA™- and spreadsheet-based systems are developed to realize the proposed 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 summary, 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



Citation: Ghosh, A.K.; Fattahi, S.; Ura, S. Towards Developing Big Data Analytics for Machining Decision-Making. *J. Manuf. Mater. Process.* **2023**, *7*, 159. <https://doi.org/10.3390/jmmp7050159>

Academic Editor: Jing Shi

Received: 17 July 2023

Revised: 18 August 2023

Accepted: 31 August 2023

Published: 2 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

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 of 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 (94,000 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, etc. The rapid speed at which datasets are accumulated in BD determines its velocity. The multiplicity of the contents (text, video, and graphics) and structures (structured, unstructured, and semi-structured) leads to the variety of BD. As far as variety is concerned, traditional BD-relevant database systems effectively manage only 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 leads to the volatility of BD. The accuracy of the datasets determines the veracity of BD. The appropriateness of the datasets for their intended use constitutes the validity of BD. Finally, the economic or social wealth generated from BD is

referred to as its value. Regarding its economic value, it is expected that by the end of 2022, the BD market will grow to USD 274.3 billion [11].

The remarkable thing is that the value of BD can be ensured by developing Big Data Analytics (BDA). BDA formally compute the relevant datasets available in BD using statistical or logical approaches [12]. The concerned organizations (e.g., the International Organization 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 BD, and consequent decision-making abilities. Nevertheless, BDA, by nature, are highly resource-dependent and opaque (acting like a black-box) [27]. Thus, making BDA less resource-dependent and transparent is a challenge. By overcoming this challenge, BD inequality can also be resolved.

As seen in Figure 1, BD and BDA are valuable constituents of cyber–physical systems of smart manufacturing [28]. In this scenario, Digital Manufacturing Commons (DMCs) [29–33] integrate the cyber and physical spaces. As seen in Figure 1, different stakeholders (producers, designers, planners, and customers) ubiquitously produce various content such as numerical datasets, digital artifacts (e.g., CAD data), metadata, executable/scripted codes, machine performance statistics, models, algorithms, etc. One example of such content is manufacturing process-relevant documents, such as scientific articles containing process-relevant information and data published in several journals and conference proceedings, as shown in Figure 2. These contents are converted into a widely accessible digital format, resulting in DMCs [32]. These commons (DMCs) 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 [32]. As such, BDA can operate on DMCs and extract the required knowledge. The extracted knowledge helps run the IoT-embedded manufacturing enablers (robots, machine, tools, processes, and different planning systems).

However, there is no steadfast procedure found in the literature (see Section 2) to develop inequality-free and transparent BD and BDA for smart manufacturing. Nevertheless, answers to the following questions are required to develop a BD and BDA framework and the underlying systems (The origins of these questions as well as the answers are described in detail in Section 3).

1. How should the documentation process for manufacturing be carried out?
2. What should be the process for digitizing and integrating the prepared documents with process-relevant BD in terms of DMCs?
3. What is the proper procedure for utilizing the relevant dataset found in the shared documents?
4. What should be the method for meeting the computational challenges underlying the relevant datasets generated from multiple sources?
5. What is the recommended method for extracting rules and drawing conclusions?

Thus, this article addresses the above questions and their answers. The goal is to elucidate a systematic approach for developing a BDA framework from the perspective of

smart manufacturing. The focus is to accelerate manufacturing process-relevant decision-making activities in smart manufacturing.

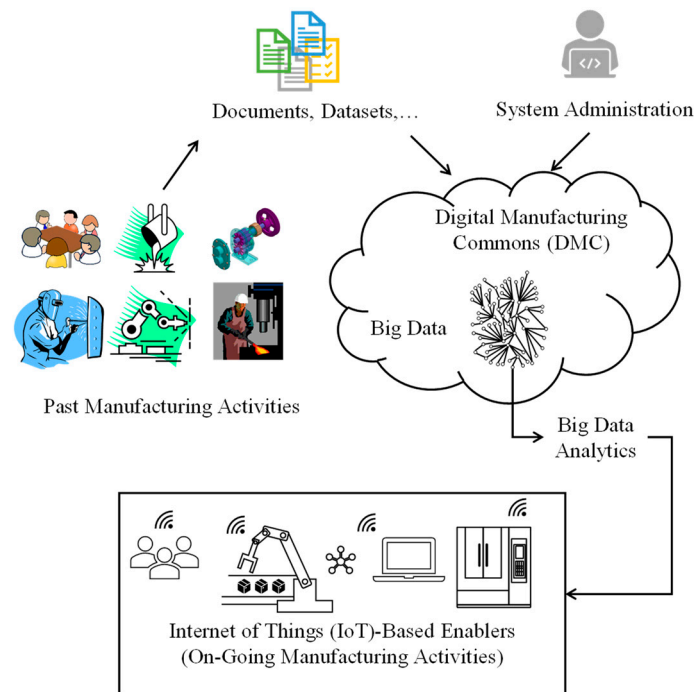


Figure 1. Context of digital manufacturing commons and big data in smart manufacturing.

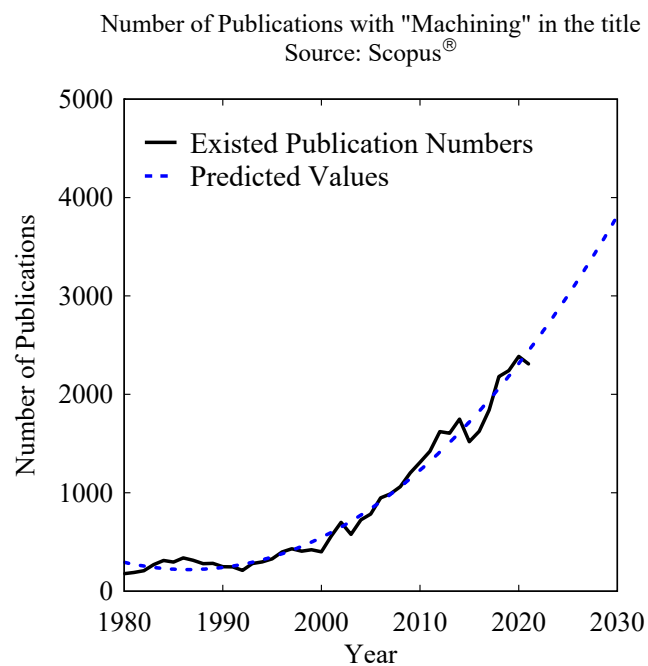


Figure 2. Growth of manufacturing-process-relevant documents.

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 fundamental concepts and the proposed BDA framework. This section also elucidates the main system components of the BDA and their functional requirements. Section 4 presents a prototyping of the proposed BDA developed using the JAVA™ platform and spreadsheet applications. Section 5 presents a case study showing the efficacy of the BDA where the

goal is to determine the optimal material removal conditions of dry Electrical Discharge Machining (EDM). Finally, Section 6 concludes this article.

2. Literature Review

This 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 BD from various sources (RFID, sensor, AGV, etc.). They recommended 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 emphasized data retention policies and down-sampling methods to optimize BDA.

Fattahi et al. [37] developed 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 system 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 a 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 them in a distributed database called Cassandra. Finally, Machine Learning (ML) 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, etc.), hardware and software tools for data management and analysis (NoSQL databases, Scikit-learn, Python, R, Java, etc.), 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 Extensible Markup Language (XML) to analyze data related to issues such as 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 BDA (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 BD 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 ML 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 ML can enhance BDA in industries such as 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 BDA framework using RFID technology in a physical Internet-based shop floor setting. The shop floor used IoT-based tools to turn logistic resources into smart manufacturing objects. A framework was 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 such as cleansing and classification, and using the 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 the database. It uses RFID and sensors to collect data from multiple sources and process and store using frameworks such as 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 BDA 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 BD are essential for energy-efficient machining optimization. They developed a system that collects energy consumption data from the shop floor using a wireless system, handles them with a system called Hadoop Hive, and processes them using ML 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 (DBN) and Genetic Algorithm (GA) for optimization. However, the data analytics structure is not fully explained.

Chen et al. [52] used the 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] suggested having human–cyber–physical system-friendly BD that are easily accessible and understandable for both humans and machines. They also proposed 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 BD usage in intelligent manufacturing and found that current 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, 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. Proposed BDA

This section presents the fundamental concepts and the proposed BDA framework. For the sake of better understanding, this section is divided into two subsections. In particular, Section 3.1 presents the fundamental concepts consisting of context, basic functionalities, and computational challenges of BDA. Lastly, Section 3.2 presents the framework and system components as well as their architecture needed to implement the proposed BDA.

3.1. Fundamental Concepts

First, consider the aspect of a manufacturing process. A manufacturing process entails Control Variables (CVs) and Evaluation Variables (EVs). Here, the CVs are some pre-defined variables (e.g., feed rate, depth of cut, cutting speed, etc.) for ensuring the desired performance of the process in terms of some other variables called EVs (e.g., low tool wear, high material removal rate, etc.). The knowledge extracted from CV–EV-centric datasets often drives the relevant process optimization tasks. Say a manufacturing process entails a set of CVs denoted as $\{CV_i \mid i = 1, 2, \dots\}$ and a set of EVs denoted as $\{EV_j \mid j = 1, 2, \dots\}$. For example, if the datasets associated with the sets of CVs and EVs unfold the following rule “if CV_3 increases, then EV_1 increases”, it means that CV_3 is more influential than the other CVs as far 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 CV_3 whenever needed. One may refer to the work described in [12] for a relevant example, where CV–EV-centric datasets of a machining process called dry EDM are analyzed. The authors articulated that current is the most influential CV (among others such as voltage, pulse-off time, gas pressure, and spindle speed) for controlling (maximizing or minimizing) an EV called material removal rate, as far as dry EDM is concerned. The analysis revealed that if the current increases, then the material removal rate also increases, and if the current decreases, then the material removal rate also decreases. One may use this knowledge for enhancing the productivity of the relevant process, when needed. Therefore, a systematic arrangement (e.g., BDA) 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 DMCs and generate process-relevant BD, as described in Section 1 (can also be seen in Figure 1). For this, BDA is a must to extract knowledge from these commons (DMCs) 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 developing such BDA. The presented BDA architectures mostly adapt AI-based ML/deep learning algorithms (e.g., CANN, GA, DBNN, etc.) and result in opaque/black-box systems [27,55,56].

Black-box systems refer to non-transparent systems where the input and output of the algorithms are known, but internal analyses remain unknown because ML algorithms or AI systems only aim at ultimate goals by their philosophy rather than answering inherent processes [57]. Despite providing convincing results, the lack of an explicit model makes ML solutions somewhat difficult to relate to real-life scenarios. Moreover, compared to statistical analyses, ML demands an *a priori* algorithm and discards human contribution in the subsequent processing of data [58]. As a result, such black-box-type BDA frameworks become incomprehensible to humans [27,59–61]. 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 inter-

ventions with the systems. Note that “human intervention” here means engaging human intelligence (preference and judgment) as straightforwardly as possible. For example, Ullah and Harib [62] demonstrated that the ID3-based decision tree, a popular ML algorithm, although it extracts global knowledge underlying a set of CV–EV-centric data, excludes certain CV(s) while extracting the knowledge. As a result, understanding the relationship between the excluded CV(s) and the target EV remains unknown, which creates operational problems. Therefore, the authors [62] introduced a human-assisted knowledge extraction system for machining operations. The system developed [62] utilizes probabilistic and fuzzy logical reasoning to extract knowledge from machining data. It involves the user through its transparent architecture while performing the formal computation regarding knowledge extraction. Such contemplation of engaging the user with data analysis and knowledge extraction processes also remains valid for BDA. Based on this, this study proposes a transparent and human- and machine-friendly BDA framework for optimizing manufacturing processes.

Figure 3 shows the basic functionalities of the proposed BDA. As seen in Figure 3, the BDA entails three basic functionalities: Documentation, Integration, and Analysis. Here, Documentation means documenting a manufacturing process based on a user-friendly and flexible ontology. It facilitates the structuring of process-relevant information and data, eliminating the heterogeneous characteristic due to the involvement of different sources or manufacturing workspaces. For example, let us assume that different or the same manufacturing processes are carried out in different workspaces (denoted as Source 1, . . . , Source N) as shown in Figure 3. This generates process-relevant information and data (e.g., information related to machine tools, workpiece, machining conditions, experimental results, CV–EV-centric datasets, etc.). These data and information are often documented, depending on practitioners’ preference. As a result, there is a vast heterogeneity in the conventional documentation process. This heterogeneity must be eliminated for the sake of smooth data ingestion from different sources and ease of data analysis. One solution might be documenting process-relevant information and data based on a pre-defined ontology, regardless of the source. As seen in Figure 3, Integration means converting the documentation into a machine/human-readable format (e.g., XML, JSON, etc.) and integrating it into the process-relevant BD. For the sake of integration, one may prefer NoSQL databases (e.g., MongoDB®), cloud-based repositories (e.g., GitHub®, Dropbox™, etc.), or distributed files systems (e.g., Hadoop®, Spark™, etc.) through an appropriate application programming interface (API). Finally, as seen in Figure 3, Analysis means acquiring the desired CV–EV-centric datasets from the BD and extracting knowledge from them (e.g., concluding a set of rules). For this, one may search and access the machine/human-readable data (denoted as D_1, \dots, D_N in Figure 3) generated by different sources (Source 1, . . . , Source N) and residing in the BD and acquire the desired CV–EV-centric datasets. As far as knowledge extraction from the datasets (e.g., concluding a set of rules) is concerned, one may prefer different types of reasoning (e.g., deductive, inductive, cause-and-effect, abductive, etc.) and relevant ML algorithms or statistical analysis. It is worth mentioning that knowledge extraction largely depends on the interest and motive of the end-user. Thus, it is a cumbersome task to provide a universal knowledge extraction approach. Regardless of the knowledge extraction approach followed by a user, IoT-embedded enablers (machine tools, processes, planning systems, robots, human resources, etc.) residing in a real-life manufacturing environment access the knowledge and make process optimization-relevant decisions whenever needed. As schematically shown in Figure 3, computational challenges underlying the CV–EV-centric datasets must be met while extracting the knowledge. One immediate question arises: what sort of computational challenges must be met by the BDA? Figure 4 answers this question schematically.

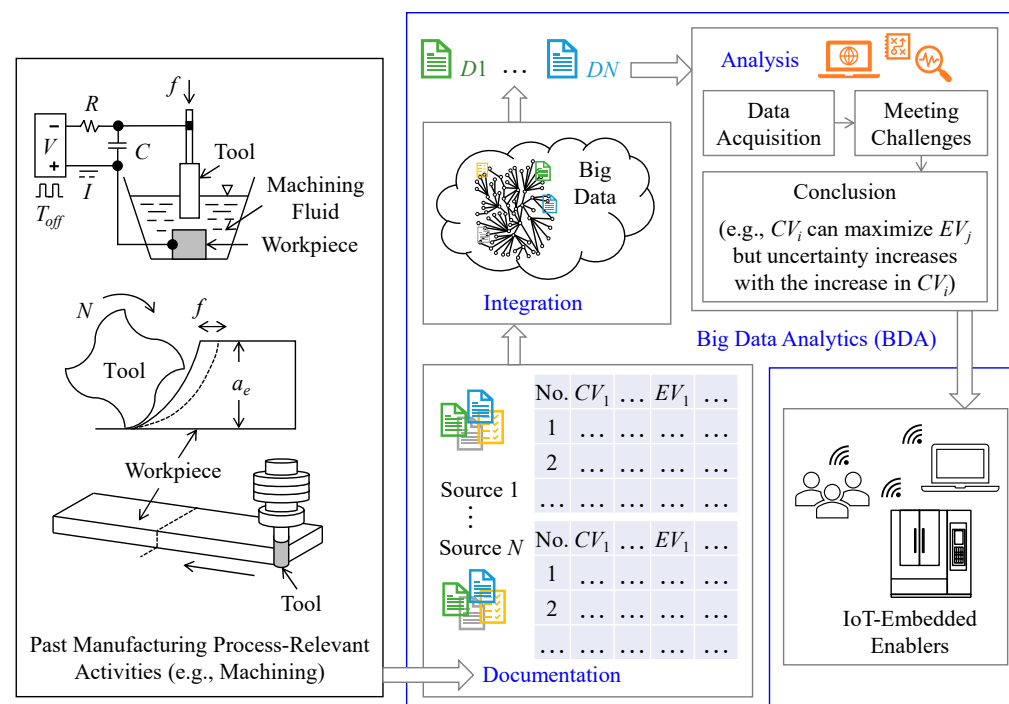


Figure 3. Basic functionalities of the proposed big data analytics.

Let us assume, as seen in Figure 4, the CV–EV-centric past experimental and analytical documents from different sources generate manufacturing process-relevant BD. The BDA searches the BD for a user-defined keyword (e.g., process type, workpiece material, etc.) and acquires two relevant documents, denoted as $D1$ and $D2$ in Figure 4, that originated from two different sources. As a result, some computational challenges appear for the BDA. In particular, $D1$ 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 a 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 correlations. For example, as seen in Figure 4, $D1$ reflects a direct correlation, but $D2$ reflects an indirect one for the same CV–EV.

Apart from the above-mentioned cases (supporting and conflicting cases), challenges associated with uncertainty and contextual differences might appear, as shown in Figure 4 schematically. Here, uncertainty means the variability in the acquired CV–EV-centric datasets. It reflects how far apart the datasets are. Its quantification helps obtain more accurate predictions in terms of consistency. For example, less uncertainty means the CV–EV correlation is consistent, and the correlation may be generalized for other cases. High uncertainty means the CV–EV correlation is inconsistent, and the correlation may 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 obtain 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 the experimental design of the manufacturing process. Therefore, the BDA must meet the above-mentioned challenges for concluding a set of rules among CV–EV-centric datasets from BD. Based on the consideration described above, the following questions arise (already mentioned in Section 1, as well):

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 BD? (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 described in the following subsection.

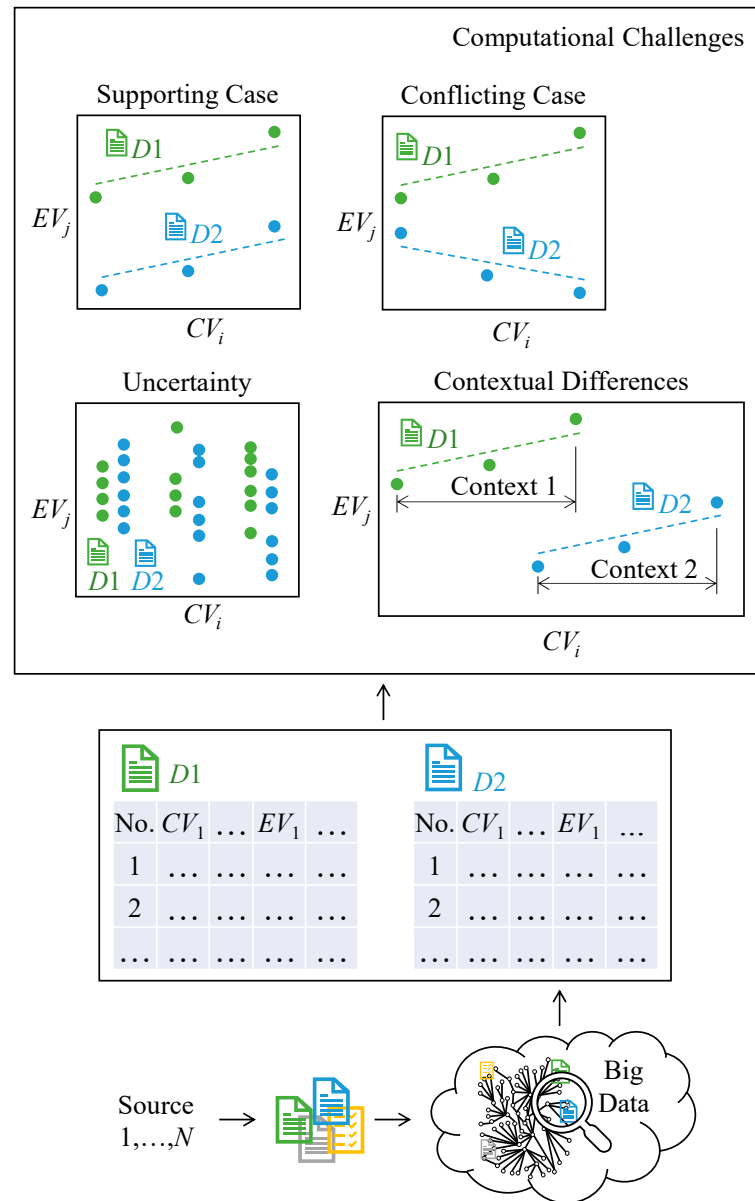


Figure 4. Computational challenges underlying big data analytics.

3.2. BDA Framework

This subsection presents the proposed BDA framework based on the fundamental concepts described in the previous subsection. Figure 5 schematically illustrates the proposed framework. As seen in Figure 5, the framework consists of five system components:

- (1) Data Preparation System (DPS);
- (2) Data Exploration System (DES);

- (3) Data Visualization System (DVS);
- (4) Data Analysis System (DAS);
- (5) Knowledge Extraction System (KES).

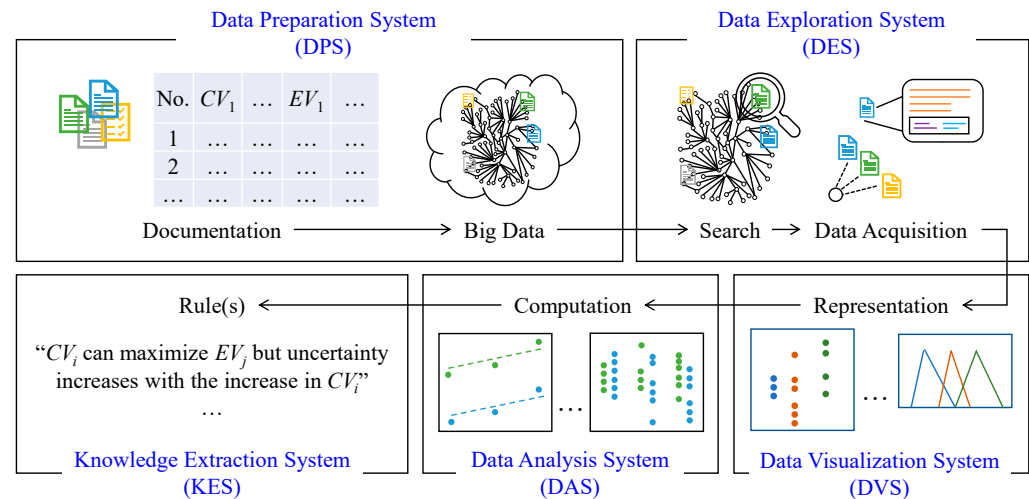


Figure 5. Framework of the proposed big data analytics framework.

The systems collectively answer the above-mentioned questions (Q1, ..., Q5 in Section 3.1) and functionalize the basic functionalities of the BDA (Documentation, Integration, and Analysis, as described in Section 3.1). In particular, the DPS answers Q1 and Q2. For this, it functionalizes the documentation of the process-relevant information using a flexible ontology, followed by document integration into the process-relevant BD. The DES answers Q3. For this, it functionalizes the search for CV–EV-centric datasets from the BD using a user-defined keyword (e.g., process type, workpiece material, etc.) and acquires the appropriate CV–EV-centric datasets from the searched outcomes. The DVS and DAS collectively answer Q4. For this, the DVS functionalizes the representation of the acquired CV–EV-centric datasets graphically. The DAS functionalizes intelligent computations on the CV–EV-centric datasets to meet the computational challenges. Finally, the KES answers Q5. For this, it functionalizes rule extraction based on the outcomes of the DAS. Figure 6 schematically shows the relationships among the above-mentioned system components and their activities. As seen in Figure 6, the DPS provides a facility to create a metafile based on user-defined process-relevant inputs (process type, number of experiments, number of CVs/EVs, and maximum number of CV levels). The metafile can follow any file format. However, this study considers an Excel™-based metafile for comprehensibility and availability. The 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, CVs, EVs, and results or CV–EV-centric datasets. The documented metafiles then need to be digitized following a human/machine-readable data format (e.g., Extensible Markup Language (XML), JavaScript Object Notation (JSON), etc.) for the sake of seamless integration among stakeholders (human and machine). This also promotes interoperability, ease of use, handling a vast amount of data, and efficient data processing across various applications and use cases since Excel™-based metafiles are not adequate. For this, the DPS provides another facility to digitize the documents (filled metafiles) into a machine/human-readable format. This study considers the XML data format for this purpose. The DPS also provides a facility to ingest the XML data to a database or cloud-based repository via an appropriate API. As mentioned in Section 3.1, one may prefer NoSQL databases (e.g., MongoDB®) or cloud-based repositories (e.g., GitHub®, Dropbox™, etc.) for this purpose. This study uses a Dropbox™-based cloud repository for this purpose.

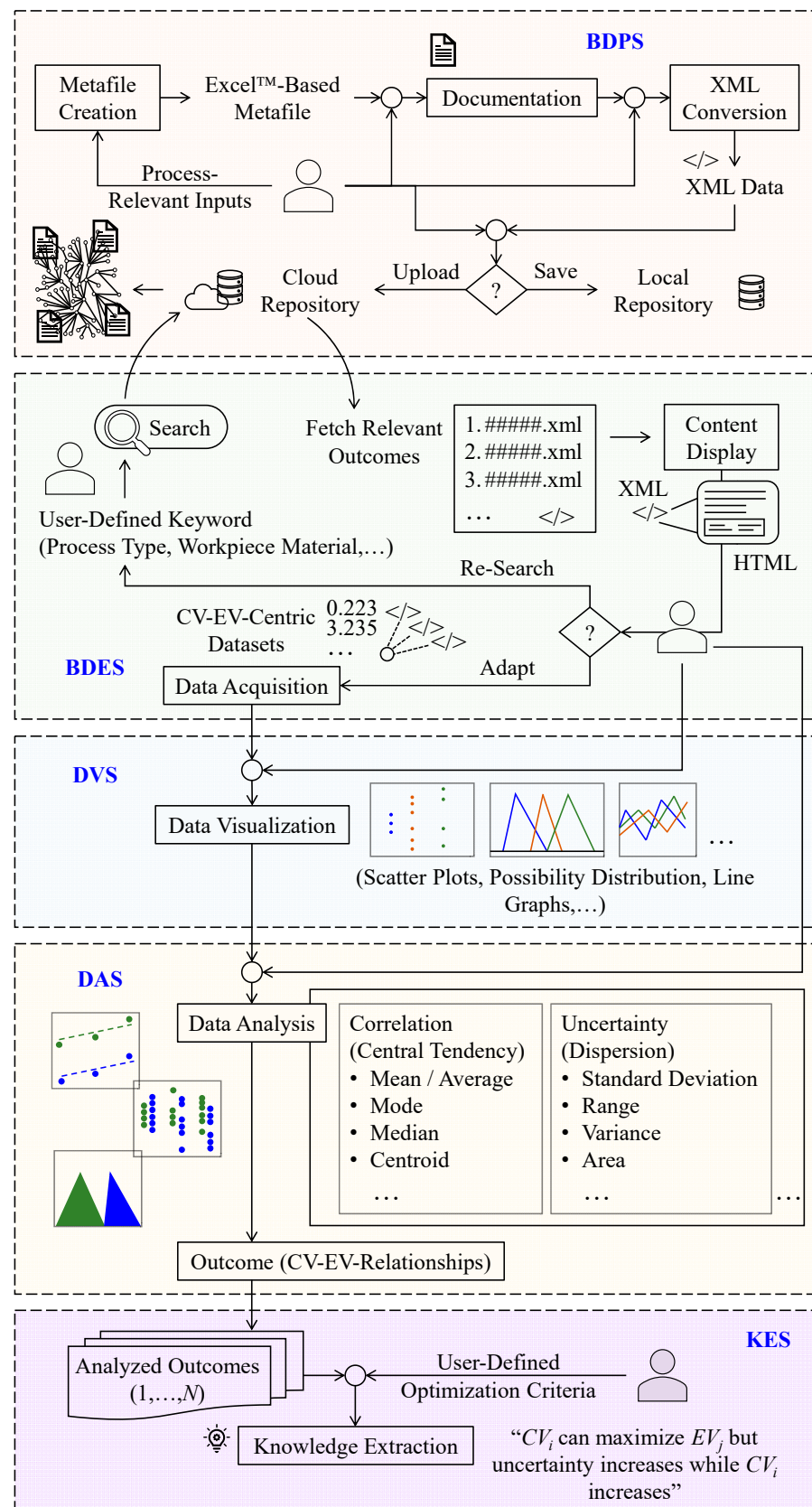


Figure 6. System architecture of the proposed big data analytics framework.

As a result, the repository contains XML data from different sources and generates process-relevant BD. As seen in Figure 6, the DES provides a facility to search the repository

using a user-defined keyword (e.g., process type, workpiece material, etc.) and fetch the relevant outcomes (files containing XML data). The outcomes are also presented in a meaningful way by the system. For this, the DES provides a facility to display the contents of the fetched XML data using an appropriate data presentation technique (here, HTML-based presentation). This presentation helps 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.

As seen in Figure 6, the 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, etc.). The user can choose an appropriate technique within the system and visualize the datasets whenever needed. Note that the modular nature of the DVS allows to incorporate any visualization technique, if needed.

As seen in Figure 6, the 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, etc.). Note that similar to the DVS, the modular nature of the DAS allows to incorporate any computational method, if needed. The user can choose appropriate methods whenever needed. As a result, the DAS generates a set of analyzed outcomes corresponding to different sources. Finally, the 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).

The remarkable features of 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 BDA entail five systems: (1) Data Preparation System (DPS), (2) Data Exploration System (DES), (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 subsections (Sections 4.1–4.5).

4.1. Data Preparation System (DPS)

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

Consider the Metafile Creation module of the DPS. Figure 7 shows one instance of it. As seen in Figure 7, the module first takes some user inputs relevant to a manufacturing process. The user inputs are as follows: process type, total number of experiments, number of CVs, maximum number of levels, and number of EVs. For the instance shown in Figure 7, 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 7) 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 the Excel™-based metafile are its availability and comprehensibility in all sorts of workspaces.

Nevertheless, the user can use the generated metafile for documenting the manufacturing process, integrating relevant attributes such as the source of the experiment (e.g., location, organization, etc.), a summary of the experiment (e.g., purpose, findings,

etc.), process specifications (e.g., type of process), machine specifications (e.g., type, maker, model, etc.), tool specifications (e.g., type, material, maker, shape, dimension, etc.), workpiece specifications (e.g., type, material, shape, size, composition, thermal properties, hardness, tolerance, etc.), machining conditions, CVs, EVs, and experimental results (CV–EV-centric numerical datasets).

Figure 7. Screenshot of Metafile Creation module.

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, etc.). The metafile and relevant systems are flexible to support such heterogeneity.

Now, consider the XML Conversion and Data Integration module of the DPS. It converts a filled metafile (metafile after documentation) into machine-readable data (here, XML) and integrates the XML into a cloud-based repository. Figure 8 shows one instance of the module. As seen in Figure 8, the module first provides a facility (a button denoted as ‘Select file’ in Figure 8) to select a filled metafile. After conversion, the module provides a facility (a button denoted as ‘Save/Upload’ in Figure 8) 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 DMC. Whenever the user accesses this facility, a separate pop-up window appears (not shown in Figure 8) from which 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 DMC and process-relevant BD for different manufacturing processes, utilizing the DPS as mentioned above.

4.2. Data Exploration System (DES)

As described in Section 3, the DES functionalizes the searching of CV–EV-centric datasets and acquisition of relevant ones from a cloud-based repository that hosts XML data from different sources and generates process-relevant BD. Figure 9 shows one instance of the developed DES.

As seen in Figure 9, the DES first provides a user-defined search facility, where a user may define a process-relevant search key, such as process type, workpiece material, etc. For instance, as shown in Figure 9, the defined search key is ‘turning’, a manufacturing process type. Based on the search key, the DES finds all the relevant XML files from the repository and displays the outcomes as a list. For this, the DES also maintains a connection with

the repository, just like the XML Conversion and Data Integration module of the DPS. The DES then provides a facility (button denoted as ‘Show’ in Figure 9) to present the search outcomes meaningfully. For this, the DES 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 9 presents the contents (tool, workpiece, machining conditions, CVs, EVs, CV–EV-centric datasets, etc.) 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 DES facility, a button denoted as ‘Select and Proceed’ in Figure 9, or research.

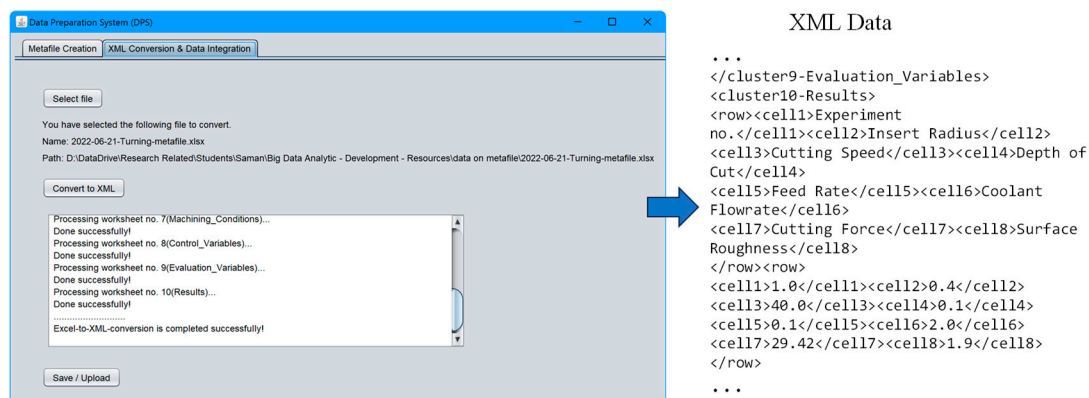


Figure 8. Screenshot of XML Conversion and Data Integration module.

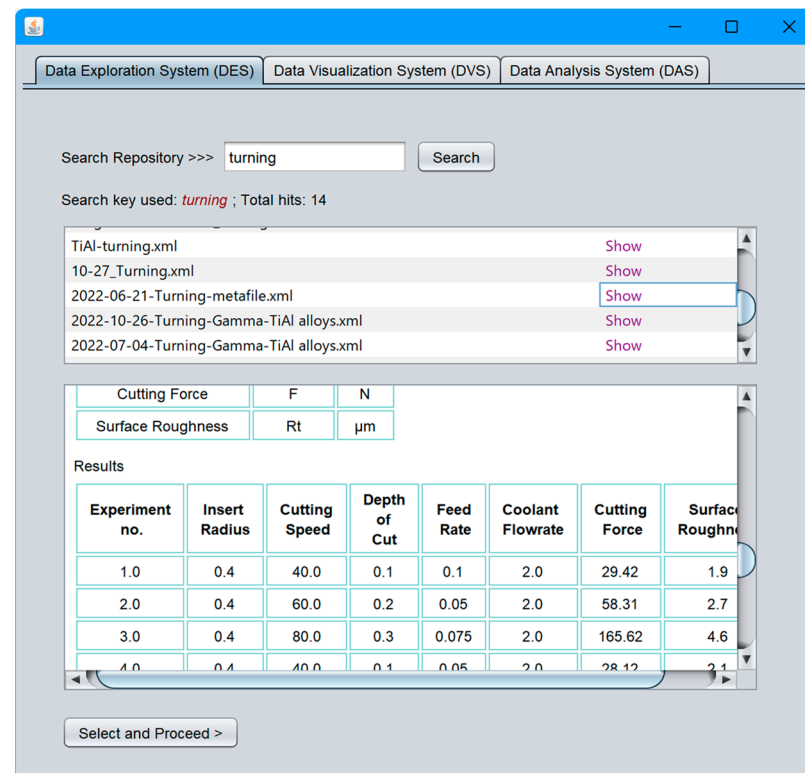


Figure 9. Screenshot of Data Exploration System.

4.3. Data Visualization System (DVS)

As described in Section 3, the DVS functionalizes visualizing the acquired CV–EV-centric datasets from the DES. Figure 10 shows one instance of the developed DVS.

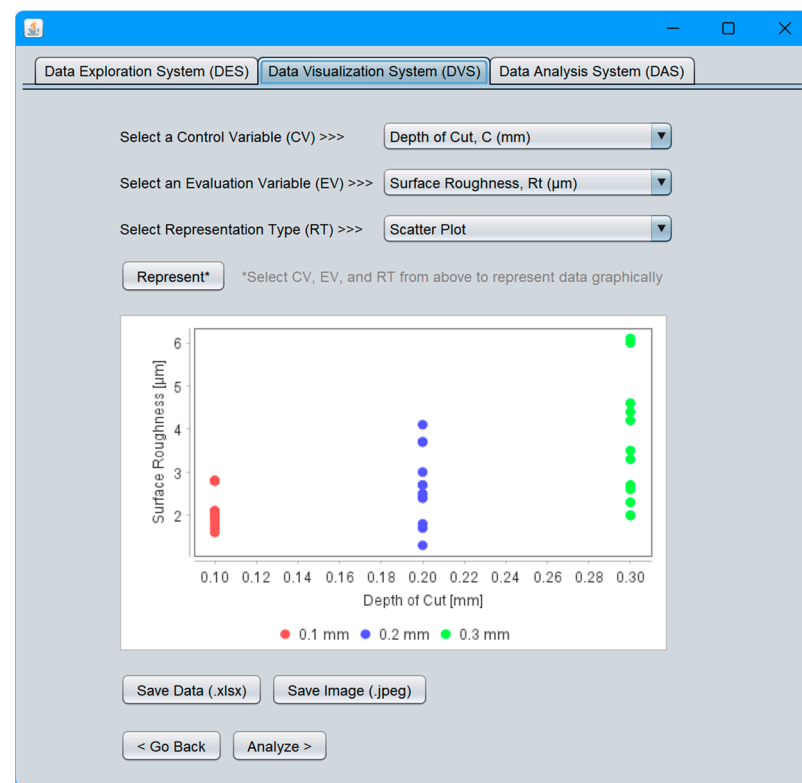


Figure 10. Screenshot of Data Visualization System.

As seen in Figure 10, the DVS provides two facilities (drop-down lists denoted as ‘Select a Control Variable (CV)’ and ‘Select an Evaluation Variable (EV)’ to select a CV and EV among others from the acquired CV–EV-centric datasets. For instance, as shown in Figure 10, 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 CVs and EVs based on the acquired datasets whenever needed. The DVS then provides another facility (drop-down list denoted as ‘Select a Representation Type (RT)’ in Figure 10) to select an appropriate representation technique among many such as scatter plots, possibility distributions [63], line graphs, etc. For instance, as shown in Figure 10, a scatter plot is selected. When a user sets a CV, EV, and representation technique, the DVS provides another facility (a button denoted as ‘Represent’ in Figure 10) for visualizing the set CV–EV-centric datasets in the form of the set representation technique. Figure 10 shows such an instance accordingly. This visualization helps a user understand individual CV–EV relationships and underlying CV levels. This way, the 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 10) to store the visualization outcomes in the forms of numeric datasets and images, if needed.

Although the DVS helps understand the CV–EV relationships qualitatively, as described above, the relationships must be quantified for knowledge extraction. For this, the DVS provides a facility (a button denoted as ‘Analyze’ in Figure 10) to transfer the set CV–EV datasets to the next system called the DAS and analyze them quantitatively.

4.4. Data Analysis System (DAS)

As described in Section 3, the DAS functionalizes the analysis of the CV–EV-centric datasets. For this, it 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 11 shows an instance of the developed DAS.

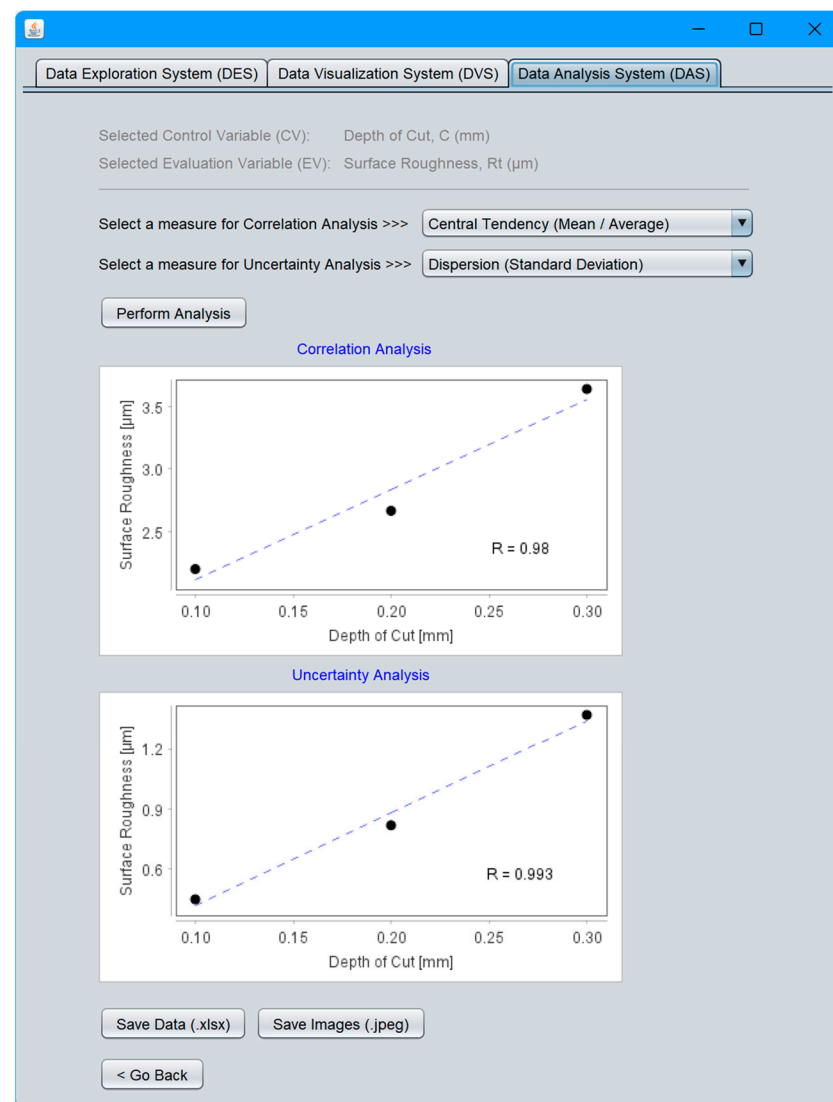


Figure 11. Screenshot of Data Analysis System.

As seen in Figure 11, the DAS first displays the CV and EV underlying the DVS-supplied datasets. For the instance shown in Figure 11, the CV and EV are 'Depth of Cut' and 'Surface Roughness', respectively, supplied from the DVS (see Figure 10). The DAS then provides two facilities (drop-down menus in Figure 11) to select measures for analyzing correlation and uncertainty associated with the datasets. For the instance shown in Figure 11, the measures called 'Central Tendency' and 'Dispersion' are selected, respectively. Afterward, the 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 11, 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 modular architecture. The outcomes from the DAS can also be exported, whenever needed, by accessing the in-built facilities (see buttons denoted as 'Save Data' and 'Save Images' in Figure 11).

Nevertheless, a user can explore the manufacturing process-relevant BD and visualize and analyze the CV–EV-centric datasets using the systems mentioned above: the DES, the DVS, and the DAS, respectively. The systems are human-comprehensible, offering thorough human interventions compared to the existing black-box systems described in Section 3. The analyzed outcomes from the DAS are processed in the KES for rule(s) extraction, as follows.

4.5. Knowledge Extraction System (KES)

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

Let us assume for a given process type, CV–EV-centric datasets from different sources populate the process-relevant BD with the aid of the DPS (see Section 4.1). Next, a user explores the BD based on a search keyword (e.g., process type, workpiece, etc.) and acquires the desired CV–EV-centric datasets from all or some of the sources with the aid of the DES (see Section 4.2). The user then visualizes and analyzes the datasets acquired from different sources with the DVS and the DAS (see Sections 4.3 and 4.4, respectively). As a result, the 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 the 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, as described in Section 3.1 as well. Nevertheless, the following section presents a case study showing how the KES functionalizes knowledge extraction for optimizing a given manufacturing process.

5. Case Study

This case study applies the proposed BDA shown in the previous section. It considers EDM as an illustrative example. In order to optimize EDM operations, knowledge is needed, which comes from experimental studies. There are many experimental studies regarding EDM. For example, when a popular search engine (Google Scholar™) was searched using the keyword “EDM”, it produced 52,900 hits. All these studies thus constitute the BD 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 under different workpiece materials. Unfortunately, it is not the case now. However, the authors selected six studies [64–69] on dry EDM where the workpiece materials were high-speed or stainless steels, reporting relevant CV–EV datasets. First, the datasets collected from the selected studies are digitized using the system components described previously (see Figures 7 and 8). They are also explored and analyzed using the system components described in Figures 9–11. 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 CVs and EVs of the studies [64–69] are listed in Table 1, showing that the following CVs were used in the said studies collectively: 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). As seen in Table 1, the following EVs were used in the said studies collectively: 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 (R_a), and Tool Wear Rate. Note that the CVs and EVs used in these studies are not exactly the same as each other. Table 1 demonstrates this aspect clearly and indicates that 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 BD. The following visualizations elucidate these.

Table 1. List of CVs and EVs across different sources for EDM.

| Source (S_i) | Control Variable ($CV_{j=1, \dots, m}$) | | Evaluation Variable ($EV_{k=1, \dots, n}$) | | Datasets ($m \times n$) |
|--|---|---------------------|--|------------------------------|------------------------------|
| | j | CV_j | k | EV_k | |
| S_1 [64] | 1 | Gap Voltage | 1 | Material Removal Rate | 20 |
| | 2 | Current | 2 | Tool Wear Rate | |
| | 3 | Pulse Off Time | 3 | Radial Over Cut | |
| | 4 | Gas Pressure | 4 | Depth Achieved | |
| | 5 | Rotational Speed | | | |
| S_2 [65] | 1 | Gap Voltage | 1 | Material Removal Rate | 30 |
| | 2 | Current | 2 | Tool Wear Rate | |
| | 3 | Pulse Off Time | 3 | Oversize (entry of hole) | |
| | 4 | Gas Pressure | 4 | Oversize (50% of hole depth) | |
| | 5 | Rotational Speed | 5 | Oversize (90% of hole depth) | |
| | 6 | Shielding Clearance | | | |
| S_3 [66] | 1 | Current | 1 | Material Removal Rate | 8 |
| | 2 | Pulse On Time | 2 | Tool Wear Rate | |
| | 3 | Duty Factor | | | |
| | 4 | Rotational Speed | | | |
| S_4 [67] | 1 | Gap Voltage | 1 | Material Removal Rate | 24 |
| | 2 | Current | 2 | Tool Wear Rate | |
| | 3 | Pulse Off Time | 3 | Oversize | |
| | 4 | Gas Pressure | 4 | Depth Achieved | |
| | 5 | Rotational Speed | | | |
| | 6 | Shielding Clearance | | | |
| S_5 [68] | 1 | Current | 1 | Material Removal Rate | 18 |
| | 2 | Pulse On Time | 2 | Surface Roughness | |
| | 3 | Duty Factor | 3 | Radial Over Cut | |
| | 4 | Gas Pressure | | | |
| | 5 | Rotational Speed | | | |
| | 6 | Gas Type | | | |
| S_6 [69] | 1 | Current | 1 | Material Removal Rate | 15 |
| | 2 | Pulse On Time | 2 | Surface Roughness | |
| | 3 | Duty Factor | 3 | Radial Over Cut | |
| | 4 | Gas Pressure | | | |
| | 5 | Rotational Speed | | | |
| Total number of CV–EV-centric datasets | | | | | 115 |

First, consider the bubble plot shown in Figure 12. The plot displays datasets from the six different studies or sources (S_1, \dots, S_6) in distinct colors. The size of each bubble corresponds to the number of datasets 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 13. This time the bubbles are organized according to their sources.

Let us be more specific. Consider the EV called Material Removal Rate (MRR). It is referred to in 32 datasets among the 115 and nine unique combinations of CV–EV, as shown in Figure 14. As seen in Figure 14, all six sources deal with the MRR. Consider another EV called Tool Wear Rate (TWR). It is referred to in 21 datasets among the 115 and eight unique combinations of CV–EV, as shown in Figure 15. Four sources among six provide datasets for this EV (i.e., TWR), unlike the case shown in Figure 14. Lastly, consider the EV called Surface Roughness measured by the arithmetic average of roughness profile height deviations from the mean line (R_a). 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 16. 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.

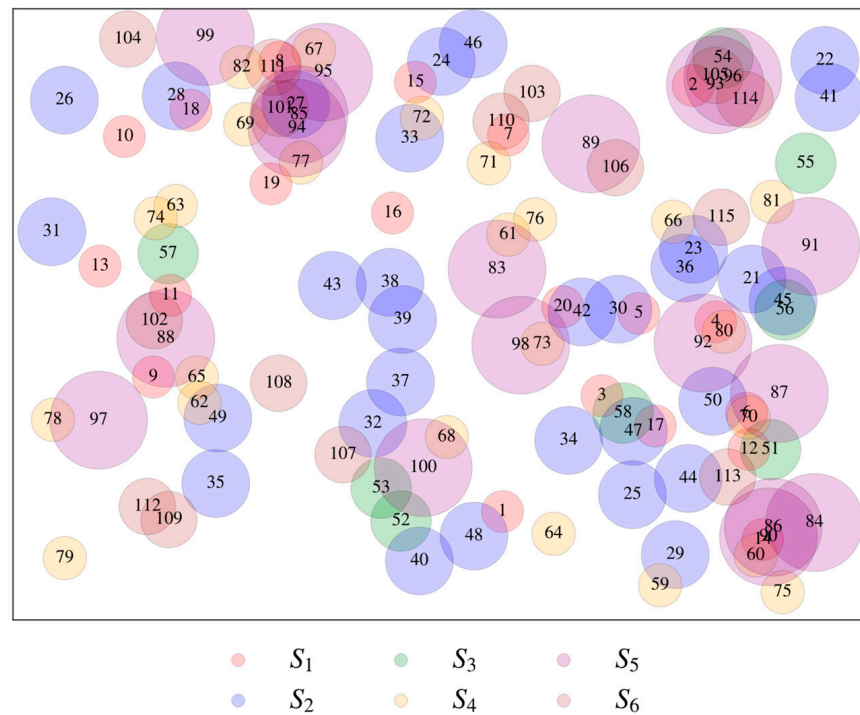


Figure 12. CV-EV-centric datasets relevant to Table 1.

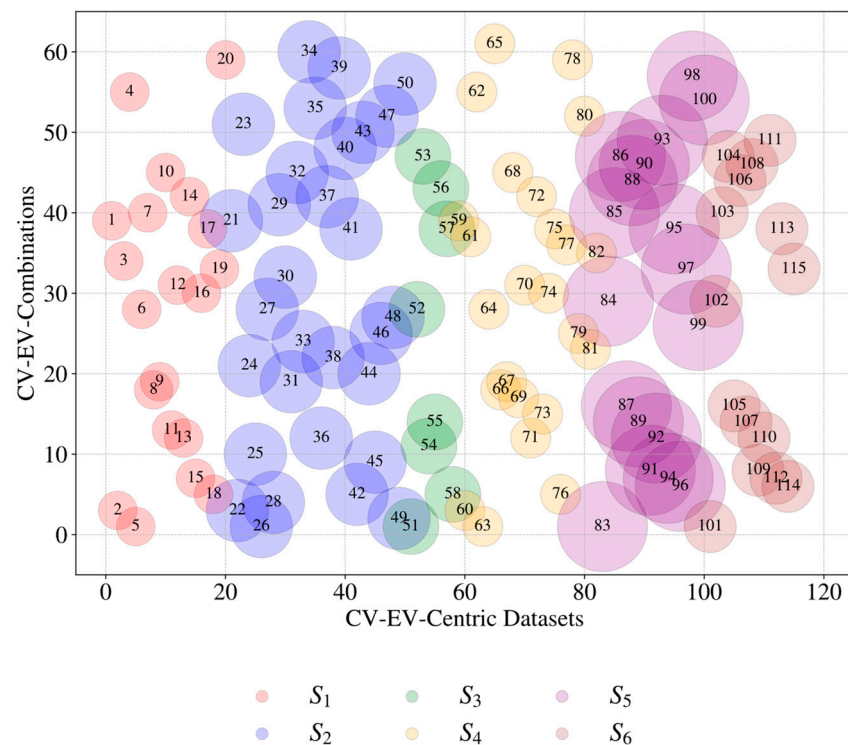


Figure 13. Available combination-centric visualization of CV-EV-centric datasets.

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 between CVs and EVs serve as the primary knowledge, which must be quantified. The relationships are established by using the tools available in the DAS (Figure 11). The tools must be used keeping in mind that there are some computational complexities, as schematically illustrated in Figure 4. 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.

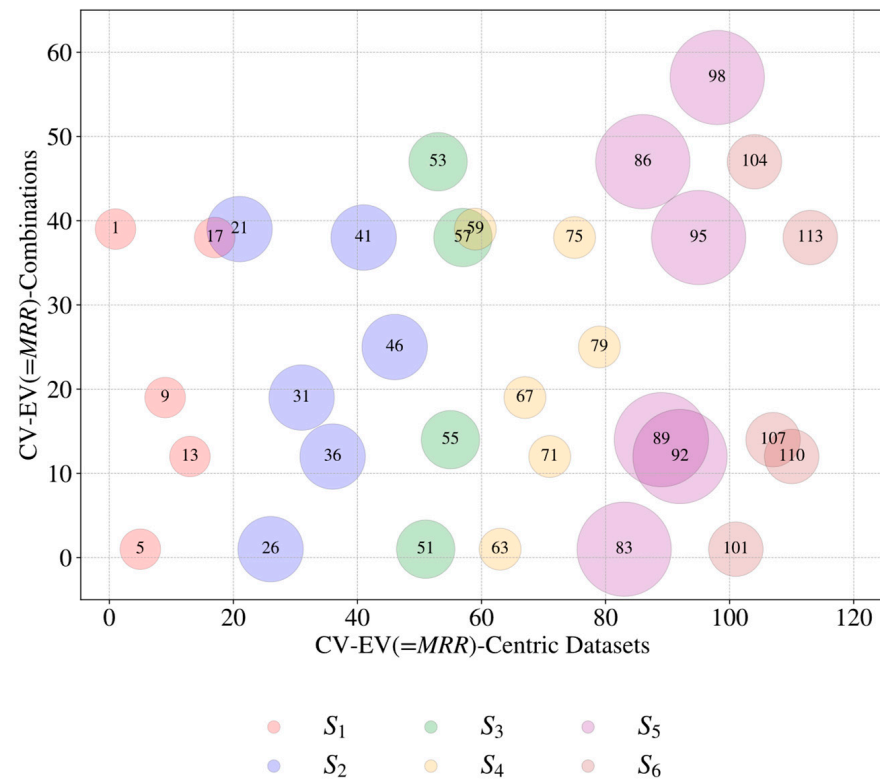


Figure 14. MRR-relevant datasets from different sources.

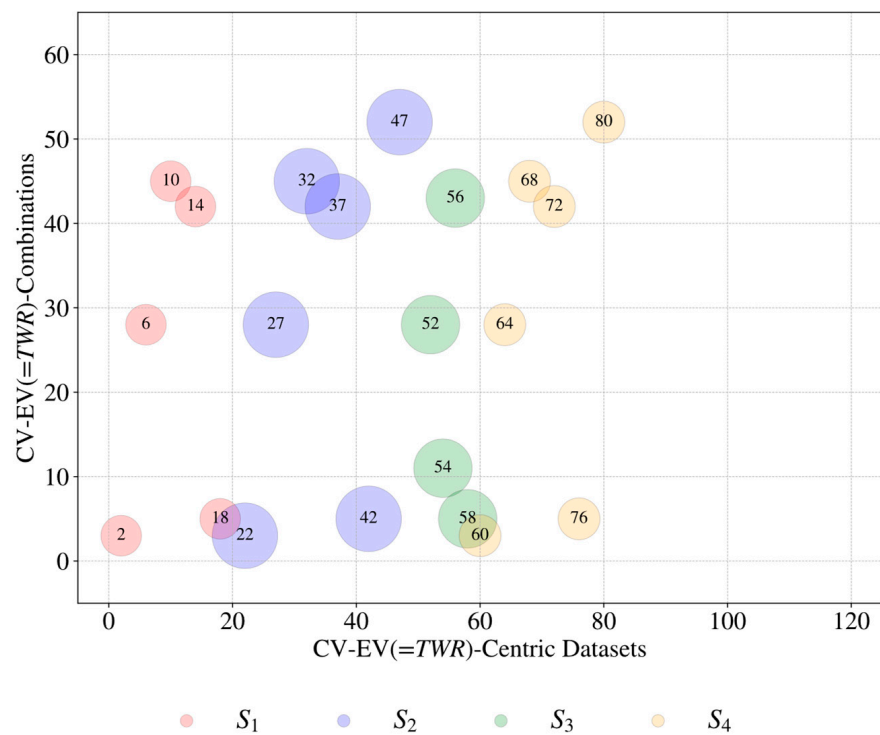


Figure 15. TWR-relevant datasets from different sources.

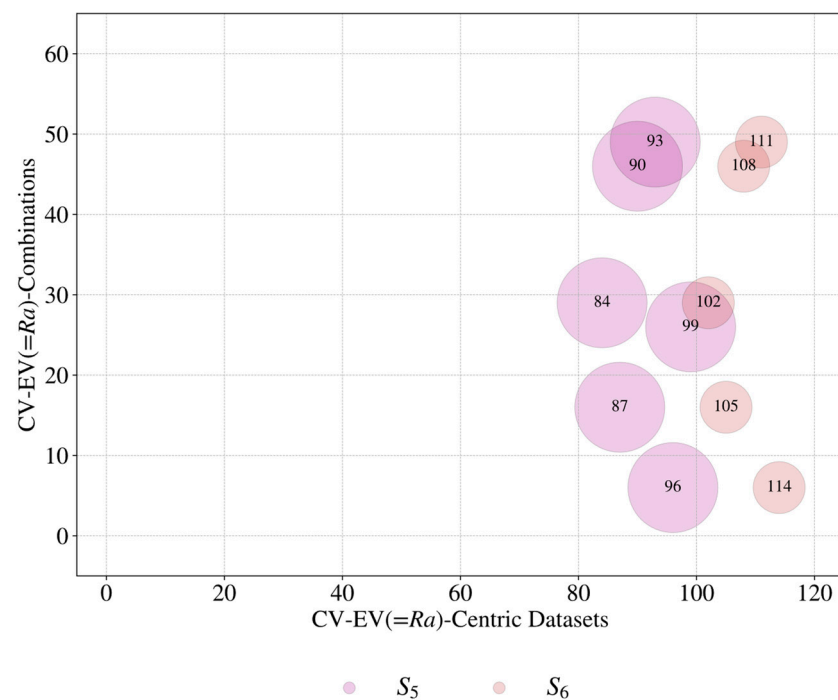


Figure 16. *Ra*-relevant datasets from different sources.

For example, consider the combination ($CV = \text{Gap Voltage (V)}$, $EV = MRR$). Figure 17a shows a scatter plot of all relevant datasets taken from S_1 , S_2 , and S_4 . Figure 17b–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 Figure 17b–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 17b–d).

Lastly, consider the CV – EV combination of Gas Pressure (P) and Radial Over Cut (ROC). Figure 18a shows a scatter plot of all relevant datasets taken from S_1 , S_5 , and S_6 . Figure 18b, Figure 18c, and Figure 18d 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 Figure 18b–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 the others. Similar to the previous case, a huge amount of uncertainty persists here, too.

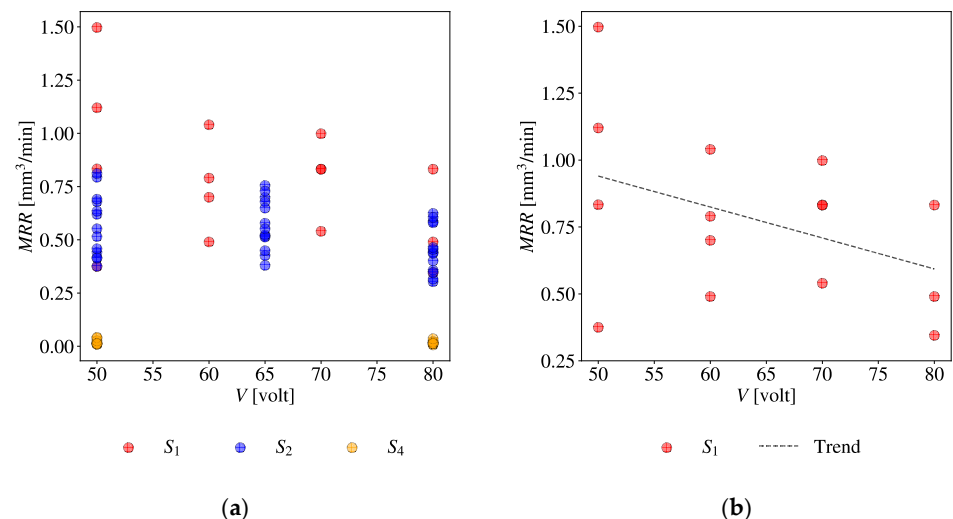


Figure 17. *Cont.*

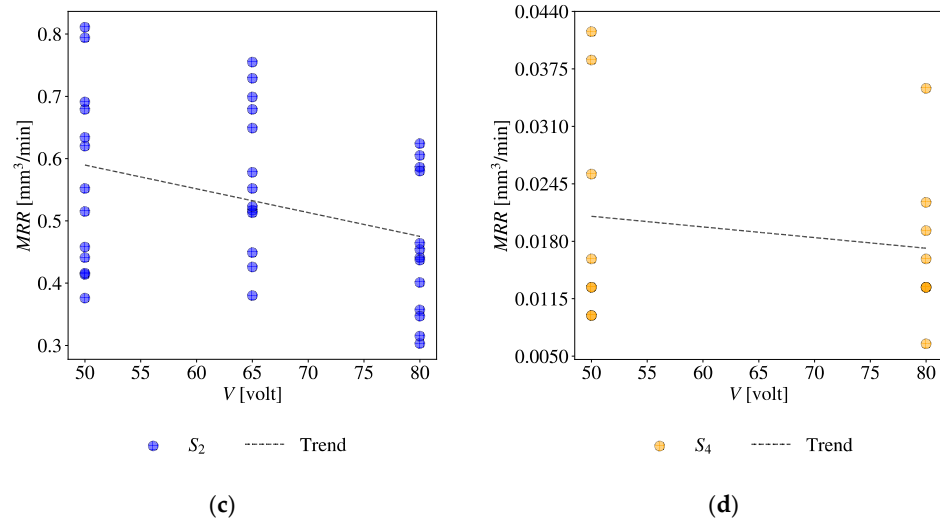


Figure 17. Gap Voltage (V) and MRR -relevant datasets for: (a) S_1 , S_2 , and S_4 , (b) S_1 only, (c) S_2 only, and (d) S_4 only.

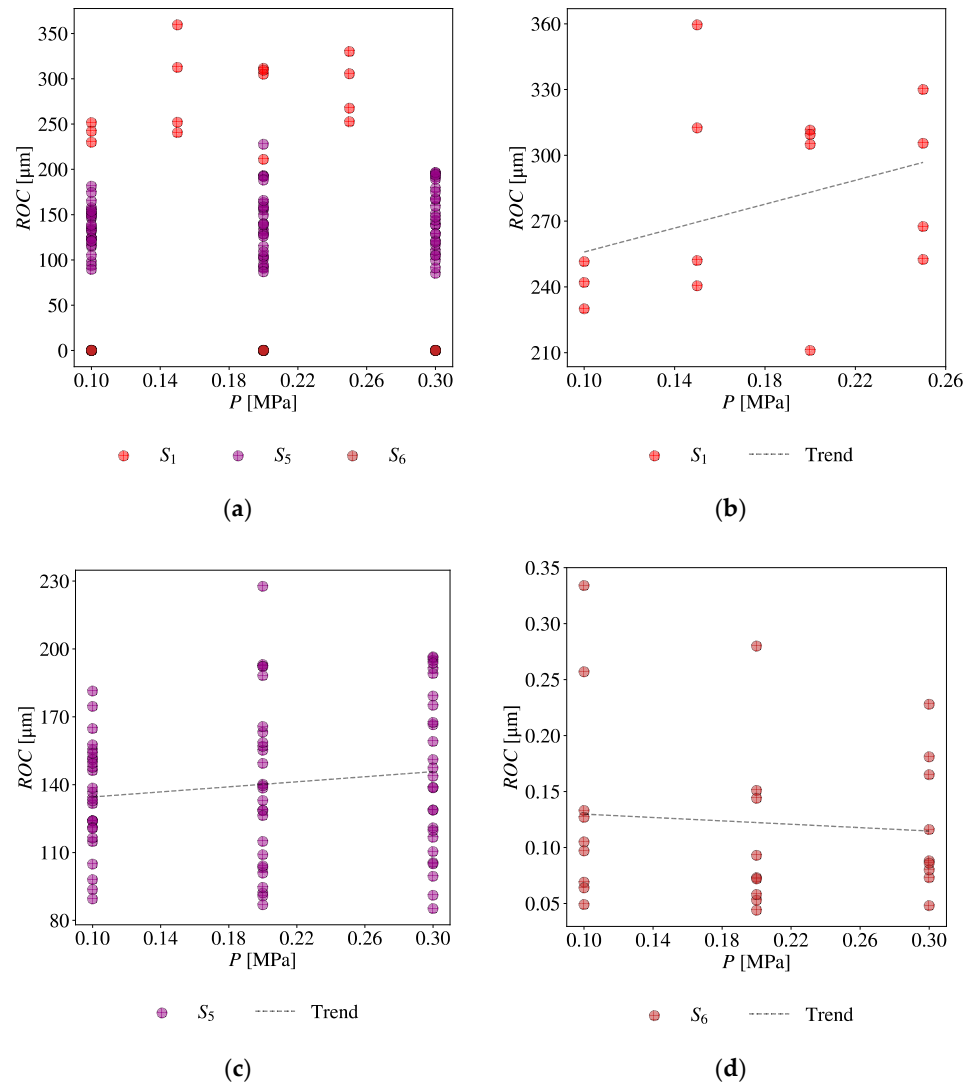


Figure 18. Gas Pressure (P) and Radial Over Cut (ROC)-relevant datasets for: (a) S_1 , S_5 , and S_6 , (b) S_1 only, (c) S_5 only, and (d) S_6 only.

Nevertheless, a personalized KES is developed using spreadsheet applications, keeping the computational complexity described above in mind. The system uses the relationships among CVs and EVs found in the previous system and determines the right set of CVs 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) 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 (see Table 1). The remarkable thing is that the study denoted as S_4 was kept aside because the CVs in S_4 have only two states. As a result, CA and UA produce only two points for each CV–EV combination, and therefore, $R = -1$ or 1 . Thus, including these kinds of datasets may produce a misleading conclusion.

Table 2. Exported outcomes from DAS for knowledge extraction.

| 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 | −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 19. 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.

The absolute value of R , i.e., $|R|$, can be divided into a few states to visualize more clearly the impact of the CVs on the given EV (this time MRR). The results shown in Figure 20 refer to three states, as follows: (1) $|R| \in [0.8, 1]$ means “significant”; (2) $|R| \in [0.4, 0.8)$ means “less significant”; and (3) $|R| \in [0, 0.4)$ means “non-significant”. These states are shown by a green-colored tick mark (\checkmark), a gold-colored exclamation mark (!), and a red-colored cross mark (X), respectively. Observing the green-colored tick mark symbols (\checkmark) makes it possible to identify a set of rules (denoted as $R_k \mid k = 1, 2, \dots$) for maximizing MRR. The results are shown in Figure 21 and summarized in Table 3.

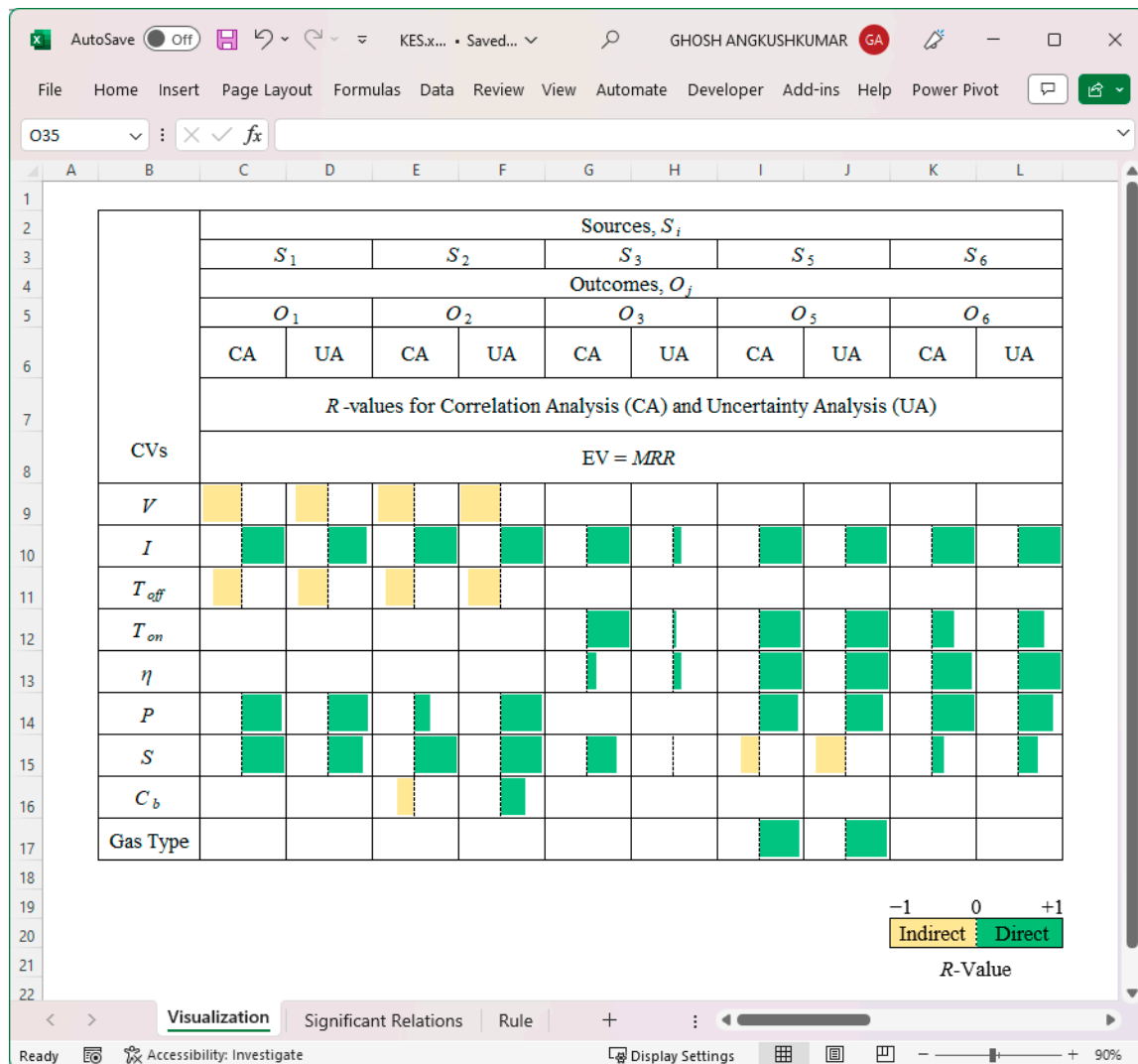


Figure 19. Screenshot of the KES for visualizing the analyzed outcomes.

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 | $\max(I) \wedge \max(T_{on}) \wedge \max(\eta) \wedge \max(P) \wedge \max(\text{Gas Type}) \Rightarrow \max(MRR)$ |
| S_6 | R_6 | $\max(I) \wedge \max(\eta) \wedge \max(P) \Rightarrow \max(MRR)$ |

Whether or not the rules produce meaningful results was tested by applying the rules to datasets S_1, \dots, 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 , $MRR' = MRR'' = 1.497 \text{ mm}^3/\text{min}$; for S_2 , $MRR' = MRR'' = 0.811 \text{ mm}^3/\text{min}$; and for S_6 , $MRR' = MRR'' = 5.31 \text{ mm}^3/\text{min}$. 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 CVs can be used to achieve the goal (here, maximizing MRR).

Table 4. Validation of R_k .

| S_i | R_k | MRR' (mm ³ /min) | MRR'' (mm ³ /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 have close connections. Since a digital twin consists of input, modeling, simulation, validation, and output modules (see [70]), 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.

The screenshot shows the KES interface with a spreadsheet titled 'KES.xlsx'. The spreadsheet contains the following data:

| 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 20. Screenshot of the KES for identifying significant relationships.

Target EV: MRR min = Minimization
 Requirement: $\max(MRR)$ 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 21. Screenshot of the KES for rule extraction.

6. Conclusions

Big data analytics (BDA) is one of the essential constituents of smart manufacturing. Unfortunately, there is no system-wise 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 BDA. Afterward, this paper presents a systematic approach to developing BDA for manufacturing process-relevant decision-making activities.

The proposed analytics consists of five integrated system components:

- Data Preparation System (DPS);
- Data Exploration System (DES);
- Data Visualization System (DVS);
- Data Analysis System (DAS);
- Knowledge Extraction System (KES).

The functional requirements of the systems are as follows.

First, the DPS must prepare contents to be included in BD. The contents may exhibit the characteristics of the so-called Digital Manufacturing Commons (DMCs). Thus, it is desirable that the system supports user-defined ontologies and produces widely acceptable digital datasets using Extensible Markup Language (XML) or any other human/machine-readable format (e.g., JSON). The DES can extract relevant datasets prepared by the first system. The system uses keywords derived from the names of manufacturing processes, materials, and analysis- or experiment-relevant phrases (e.g., design of experiment).

The third system (DVS) can help visualize relevant datasets extracted by the second system (DES) using suitable methods (e.g., scatter plots and possibility distributions). The fourth system (DAS) must establish relationships among the relevant CV (control variables that can be adjusted as needed) and EV (evaluation variables that measure the performance of a process) combinations for a given situation. In addition, it must quantify the uncertainty in the relationships. Finally, the last system (KES) can extract knowledge from the outcomes of the fourth system (DAS) using user-defined criteria (e.g., minimize surface roughness, maximize material removal rate, etc.). In addition, JAVA™- and spreadsheet-based systems are developed to realize the proposed integrated systems.

The efficacy of the proposed analytics has been demonstrated using a case study where the goal was to determine the right states of CVs of dry Electrical Discharge Machining (EDM) for maximizing the Material Removal Rate (MRR). The contents were created from published scientific articles on dry EDM that deal with stainless and high-speed steels. In addition, articles that presented datasets based on the design of experiments were considered. The datasets collectively underlie the following CVs: voltage, current, pulse-off time, pulse-on time, gas pressure, rotational speed, shielding clearance, duty factor, and gas type. The set of CVs differs from article to article.

Consequently, the values of the CVs 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 MRR. 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 MRR. The knowledge extracted can be used to optimize a dry EDM operation and elucidate the research gap in dry EDM.

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

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 BD inequality and transparency issues. This experience can be extended to develop BDA 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. For this, as reviewed in [71], blockchain-based technology can be considered. Particularly, blockchain technology can be integrated with the Data Preparation System (DPS) 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.

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

References

1. Diebold, F.X. On the Origin(s) and Development of the Term ‘Big Data’. *SSRN Electron. J.* **2012**. [CrossRef]
2. Shah, G.; Shah, A.; Shah, M. Panacea of challenges in real-world application of big data analytics in healthcare sector. *J. Data Inf. Manag.* **2019**, *1*, 107–116. [CrossRef]
3. Hassani, H.; Huang, X.; Silva, E. Banking with blockchain-ed big data. *J. Manag. Anal.* **2018**, *5*, 256–275. [CrossRef]
4. Hallur, G.G.; Prabhu, S.; Aslekar, A. Entertainment in Era of AI, Big Data & IoT. In *Digital Entertainment*; Das, S., Gochhait, S., Eds.; Springer Nature Singapore: Singapore, 2021; pp. 87–109. [CrossRef]
5. Wang, Y. Big Opportunities and Big Concerns of Big Data in Education. *TechTrends* **2016**, *60*, 381–384. [CrossRef]
6. Zhu, L.; Yu, F.R.; Wang, Y.; Ning, B.; Tang, T. Big Data Analytics in Intelligent Transportation Systems: A Survey. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 383–398. [CrossRef]
7. Henke, N.; Bughin, J.; Chui, M.; Manyika, J.; Saleh, T.; Wiseman, B.; Sethupathy, G. The Age of Analytics: Competing in a Data-Driven World. December 2016. Available online: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-age-of-analytics-competing-in-a-data-driven-world> (accessed on 15 July 2023).
8. Yin, S.; Kaynak, O. Big Data for Modern Industry: Challenges and Trends [Point of View]. *Proc. IEEE* **2015**, *103*, 143–146. [CrossRef]
9. Andre, L. 53 Important Statistics about How Much Data Is Created Every Day. Available online: <https://financesonline.com/how-much-data-is-created-every-day/> (accessed on 15 July 2023).
10. Sreenivasan, R.R. Characteristics of Big Data—A Delphi Study. Master’s Thesis, Memorial University of Newfoundland, St. John’s, NL, Canada, 2017. Available online: <https://research.library.mun.ca/13080/> (accessed on 12 September 2022).
11. Verevka, T.; Mirolyubov, A.; Makio, J. Opportunities and Barriers to Using Big Data Technologies in the Metallurgical Industry. In *Innovations in Digital Economy*; Rodionov, D., Kudryavtseva, T., Skhvediani, A., Berawi, M.A., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 86–102. [CrossRef]
12. Fattahi, S.; Ullah, A.S. Optimization of Dry Electrical Discharge Machining of Stainless Steel using Big Data Analytics. *Procedia CIRP* **2022**, *112*, 316–321. [CrossRef]
13. Chang, W. NIST Big Data Reference Architecture for Analytics and Beyond. In Proceedings of the 10th International Conference on Utility and Cloud Computing, Austin, TX, USA, 5–8 December 2017; p. 3. [CrossRef]
14. NIST Big Data Public Working Group. *NIST Big Data Interoperability Framework: Volume 1, Definitions*; NIST: Gaithersburg, MD, USA, October 2019. [CrossRef]
15. NIST Big Data Public Working Group; Subgroup, T. *NIST Big Data Interoperability Framework: Volume 2, Big Data Taxonomies*; NIST: Gaithersburg, MD, USA, November 2019. [CrossRef]
16. Chang, W.; Fox, G. *NIST Big Data Interoperability Framework: Volume 3, Use Cases and General Requirements*; NIST: Gaithersburg, MD, USA, October 2019. [CrossRef]
17. Chang, W.; Roy, A.; Underwood, M. *NIST Big Data Interoperability Framework: Volume 4, Security and Privacy*; NIST: Gaithersburg, MD, USA, October 2019. [CrossRef]
18. Chang, W.; Mishra, S. *NIST Big Data Public Working Group. NIST Big Data Interoperability Framework: Volume 5, Architectures White Paper Survey*; NIST: Gaithersburg, MD, USA, October 2015. [CrossRef]
19. Chang, W.; Boyd, D.; Levin, O. *NIST Big Data Interoperability Framework: Volume 6, Reference Architecture*; NIST: Gaithersburg, MD, USA, October 2019. [CrossRef]
20. Chang, W.; von Laszewski, G. *NIST Big Data Interoperability Framework: Volume 8, Reference Architecture Interfaces*; NIST: Gaithersburg, MD, USA, June 2018. [CrossRef]
21. NIST Big Data Public Working Group. *NIST Big Data Interoperability Framework: Volume 7, Standards Roadmap, Version 2*; NIST: Gaithersburg, MD, USA, June 2018. [CrossRef]
22. Chang, W.; Austin, C.C.; Reinsch, R. *NIST Big Data Interoperability Framework: Volume 9, Adoption and Modernization*; NIST: Gaithersburg, MD, USA, October 2019. [CrossRef]
23. Farboodi, M.; Mihet, R.; Philippon, T.; Veldkamp, L. Big Data and Firm Dynamics. *AEA Pap. Proc.* **2019**, *109*, 38–42. [CrossRef]
24. Cinnamon, J. Data inequalities and why they matter for development. *Inf. Technol. Dev.* **2020**, *26*, 214–233. [CrossRef]
25. Qureshi, S. Overcoming Technological Determinism in Understanding the Digital Divide: Where Do We Go from Here? *Inf. Technol. Dev.* **2014**, *20*, 215–217. [CrossRef]
26. Hilbert, M. Big Data for Development: A Review of Promises and Challenges. *Dev. Policy Rev.* **2016**, *34*, 135–174. [CrossRef]
27. Favaretto, M.; De Clercq, E.; Elger, B.S. Big Data and discrimination: Perils, promises and solutions. A systematic review. *J. Big Data* **2019**, *6*, 12. [CrossRef]
28. Shah, D.; Wang, J.; He, Q.P. Feature engineering in big data analytics for IoT-enabled smart manufacturing—Comparison between deep learning and statistical learning. *Comput. Chem. Eng.* **2020**, *141*, 106970. [CrossRef]
29. Evans, P.C.; Annunziata, M. Industrial Internet: Pushing the Boundaries of Minds and Machines. January 2012. Available online: https://www.researchgate.net/publication/271524319_Industrial_Internet_Pushing_the_boundaries_of_minds_and_machines (accessed on 15 July 2023).

30. Banerjee, A.G.; Beckmann, B.; Carbone, J.; DeRose, L.; Giani, A.; Koudal, P.; Mackenzie, P.; Salvo, J.; Yang, D.; Yund, W. Cloud Computing-Based Marketplace for Collaborative Design and Manufacturing. In *Internet of Things. IoT Infrastructures*; Mandler, B., Marquez-Barja, J., Campista, M.E.M., Cagáñová, D., Chaouchi, H., Zeadally, S., Badra, M., Giordano, S., Fazio, M., Somov, A., et al., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 409–418. [\[CrossRef\]](#)
31. Wu, D.; Rosen, D.W.; Wang, L.; Schaefer, D. Cloud-based design and manufacturing: A new paradigm in digital manufacturing and design innovation. *Comput. Des.* **2015**, *59*, 1–14. [\[CrossRef\]](#)
32. Beckmann, B.; Giani, A.; Carbone, J.; Koudal, P.; Salvo, J.; Barkley, J. Developing the Digital Manufacturing Commons: A National Initiative for US Manufacturing Innovation. *Procedia Manuf.* **2016**, *5*, 182–194. [\[CrossRef\]](#)
33. Samimi, F.; McKinley, P.; Sadjadi, S.; Tang, C.; Shapiro, J.; Zhou, Z. Service Clouds: Distributed Infrastructure for Adaptive Communication Services. *IEEE Trans. Netw. Serv. Manag.* **2007**, *4*, 84–95. [\[CrossRef\]](#)
34. Bi, Z.; Jin, Y.; Maropoulos, P.; Zhang, W.-J.; Wang, L. Internet of things (IoT) and big data analytics (BDA) for digital manufacturing (DM). *Int. J. Prod. Res.* **2021**, *61*, 4004–4021. [\[CrossRef\]](#)
35. Wang, J.; Xu, C.; Zhang, J.; Zhong, R. Big data analytics for intelligent manufacturing systems: A review. *J. Manuf. Syst.* **2022**, *62*, 738–752. [\[CrossRef\]](#)
36. Kahveci, S.; Alkan, B.; Ahmad, M.H.; Ahmad, B.; Harrison, R. 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.* **2022**, *63*, 214–223. [\[CrossRef\]](#)
37. Fattahi, S.; Ura, S.; Noor-E-Alam, M. Decision-Making Using Big Data Relevant to Sustainable Development Goals (SDGs). *Big Data Cogn. Comput.* **2022**, *6*, 64. [\[CrossRef\]](#)
38. Chen, T.; Wang, Y.-C. Hybrid big data analytics and Industry 4.0 approach to projecting cycle time ranges. *Int. J. Adv. Manuf. Technol.* **2022**, *120*, 279–295. [\[CrossRef\]](#)
39. Woo, J.; Shin, S.-J.; Seo, W.; Meilanitasari, P. Developing a big data analytics platform for manufacturing systems: Architecture, method, and implementation. *Int. J. Adv. Manuf. Technol.* **2018**, *99*, 2193–2217. [\[CrossRef\]](#)
40. Bonnard, R.; Arantes, M.D.S.; Lorbieski, R.; Vieira, K.M.M.; Nunes, M.C. Big data/analytics platform for Industry 4.0 implementation in advanced manufacturing context. *Int. J. Adv. Manuf. Technol.* **2021**, *117*, 1959–1973. [\[CrossRef\]](#)
41. Kozjek, D.; Vrabčič, R.; Rihtaršič, B.; Lavrač, N.; Butala, P. Advancing manufacturing systems with big-data analytics: A conceptual framework. *Int. J. Comput. Integr. Manuf.* **2020**, *33*, 169–188. [\[CrossRef\]](#)
42. Jun, C.; Lee, J.Y.; Kim, B.H. Cloud-based big data analytics platform using algorithm templates for the manufacturing industry. *Int. J. Comput. Integr. Manuf.* **2019**, *32*, 723–738. [\[CrossRef\]](#)
43. Dubey, R.; Gunasekaran, A.; Childe, S.J.; Bryde, D.J.; Giannakis, M.; Foropon, C.; Roubaud, D.; Hazen, B.T. 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.* **2020**, *226*, 107599. [\[CrossRef\]](#)
44. Zhang, C.; Wang, Z.; Ding, K.; Chan, F.T.S.; Ji, W. An energy-aware cyber physical system for energy Big data analysis and recessive production anomalies detection in discrete manufacturing workshops. *Int. J. Prod. Res.* **2020**, *58*, 7059–7077. [\[CrossRef\]](#)
45. Zhong, R.Y.; Newman, S.T.; Huang, G.Q.; Lan, S. Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives. *Comput. Ind. Eng.* **2016**, *101*, 572–591. [\[CrossRef\]](#)
46. Zhong, R.Y.; Xu, C.; Chen, C.; Huang, G.Q. Big Data Analytics for Physical Internet-based intelligent manufacturing shop floors. *Int. J. Prod. Res.* **2017**, *55*, 2610–2621. [\[CrossRef\]](#)
47. Zhang, Y.; Ren, S.; Liu, Y.; Si, S. A big data analytics architecture for cleaner manufacturing and maintenance processes of complex products. *J. Clean. Prod.* **2017**, *142*, 626–641. [\[CrossRef\]](#)
48. Lu, Y.; Xu, X. Cloud-based manufacturing equipment and big data analytics to enable on-demand manufacturing services. *Robot. Comput. Integr. Manuf.* **2019**, *57*, 92–102. [\[CrossRef\]](#)
49. Ji, W.; Wang, L. Big data analytics based fault prediction for shop floor scheduling. *J. Manuf. Syst.* **2017**, *43*, 187–194. [\[CrossRef\]](#)
50. Liang, Y.C.; Lu, X.; Li, W.D.; Wang, S. Cyber Physical System and Big Data enabled energy efficient machining optimisation. *J. Clean. Prod.* **2018**, *187*, 46–62. [\[CrossRef\]](#)
51. Ji, W.; Yin, S.; Wang, L. A big data analytics based machining optimisation approach. *J. Intell. Manuf.* **2019**, *30*, 1483–1495. [\[CrossRef\]](#)
52. Chen, C.C.; Hung, M.H.; Suryajaya, B.; Lin, Y.C.; Yang, H.C.; Huang, H.C.; Cheng, F.T. A Novel Efficient Big Data Processing Scheme for Feature Extraction in Electrical Discharge Machining. *IEEE Robot. Autom. Lett.* **2019**, *4*, 910–917. [\[CrossRef\]](#)
53. Fattahi, S.; Okamoto, T.; Ura, S. Preparing Datasets of Surface Roughness for Constructing Big Data from the Context of Smart Manufacturing and Cognitive Computing. *Big Data Cogn. Comput.* **2021**, *5*, 58. [\[CrossRef\]](#)
54. Li, C.; Chen, Y.; Shang, Y. A review of industrial big data for decision making in intelligent manufacturing. *Eng. Sci. Technol. Int. J.* **2022**, *29*, 101021. [\[CrossRef\]](#)
55. Nasir, V.; Sassani, F. A review on deep learning in machining and tool monitoring: Methods, opportunities, and challenges. *Int. J. Adv. Manuf. Technol.* **2021**, *115*, 2683–2709. [\[CrossRef\]](#)
56. Hildebrandt, M.; Koops, B.-J. The Challenges of Ambient Law and Legal Protection in the Profiling Era. *Mod. Law Rev.* **2010**, *73*, 428–460. Available online: <http://www.jstor.org/stable/40660735> (accessed on 15 July 2023). [\[CrossRef\]](#)
57. Emmert-Streib, F.; Yli-Harja, O.; Dehmer, M. Artificial Intelligence: A Clarification of Misconceptions, Myths and Desired Status. *Front. Artif. Intell.* **2020**, *3*, 524339. [\[CrossRef\]](#)
58. Bzdok, D.; Altman, N.; Krzywinski, M. Statistics versus machine learning. *Nat. Methods* **2018**, *15*, 233–234. [\[CrossRef\]](#) [\[PubMed\]](#)

59. Citron, D.K.; Pasquale, F.A. The Scored Society: Due Process for Automated Predictions. *Washingt. Law Rev.* **2014**, *89*, 1.
60. d'Alessandro, B.; O'Neil, C.; LaGatta, T. Conscientious Classification: A Data Scientist's Guide to Discrimination-Aware Classification. *Big Data* **2017**, *5*, 120–134. [[CrossRef](#)]
61. Leese, M. The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union. *Secur. Dialogue* **2014**, *45*, 494–511. [[CrossRef](#)]
62. Ullah, A.S.; Harib, K.H. A human-assisted knowledge extraction method for machining operations. *Adv. Eng. Inform.* **2006**, *20*, 335–350. [[CrossRef](#)]
63. Ullah, A.M.M.S.; Shamsuzzaman, M. Fuzzy Monte Carlo Simulation using point-cloud-based probability–possibility transformation. *Simulation* **2013**, *89*, 860–875. [[CrossRef](#)]
64. Puthumana, G.; Joshi, S.S. Investigations into performance of dry EDM using slotted electrodes. *Int. J. Precis. Eng. Manuf.* **2011**, *12*, 957–963. [[CrossRef](#)]
65. Govindan, P.; Joshi, S.S. Experimental characterization of material removal in dry electrical discharge drilling. *Int. J. Mach. Tools Manuf.* **2010**, *50*, 431–443. [[CrossRef](#)]
66. Murickan, R.T.; Jakkamputi, L.P.; Kuppan, P. Experimental Investigation of Dry Electrical Discharge Machining on SS 316L. *Int. J. Latest Trends Eng. Technol.* **2013**, *2*, 100–107. Available online: <http://www.ijltet.org/wp-content/uploads/2013/06/14.pdf> (accessed on 15 July 2023).
67. Puthumana, G.; Agarwal, R.; Joshi, S.S. 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), Visakhapatnam, India, 13–15 December 2010. Available online: <https://backend.orbit.dtu.dk/ws/portalfiles/portal/123939189/AIMTDR2010.pdf> (accessed on 15 July 2023).
68. Fattahi, S.; Baseri, H. Analysis of dry electrical discharge machining in different dielectric mediums. *Proc. Inst. Mech. Eng. Part E J. Process Mech. Eng.* **2017**, *231*, 497–512. [[CrossRef](#)]
69. Fattahi, S.; Shyha, I.; Baseri, H. Optimisation of Dry Electrical Discharge Machining of High Speed Steel using Grey-Relational Analysis. *Int. J. Robot. Mechatron.* **2015**, *2*, 132–139. [[CrossRef](#)]
70. Ghosh, A.K.; Ullah, A.M.S.; Teti, R.; Kubo, A. Developing sensor signal-based digital twins for intelligent machine tools. *J. Ind. Inf. Integr.* **2021**, *24*, 100242. [[CrossRef](#)]
71. Leng, J.; Ruan, G.; Jiang, P.; Xu, K.; Liu, Q.; Zhou, X.; Liu, C. Blockchain-empowered sustainable manufacturing and product lifecycle management in industry 4.0: A survey. *Renew. Sustain. Energy Rev.* **2020**, *132*, 110112. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.