International Journal of Pressure Vessels and Piping 206 (2023) 105061

Contents lists available at ScienceDirect

International Journal of Pressure Vessels and Piping

journal homepage: www.elsevier.com/locate/ijpvp

Adoption of big data analytics for energy pipeline condition assessment - A systematic review

Muhammad Hussain [a,*], Tieling Zhang [a], Minnat Seema [b]

[a] University of Wollongong, Northfields Ave Wollongong, NSW, 2522, Australia
[b] Griffith University, Nathan, QLD, 4111, Australia

ARTICLE INFO

Keywords:
Big data
Predictive analytics
Energy pipeline
Machine learning (ML)
Oil and gas
Asset management
Industry 4.0
Internet of things (IoT)
Pipeline condition assessment
Pipeline integrity management
Artificial intelligence (AI)

ABSTRACT

Due to complexity, the oil and gas industry use various sensors to collect data for analysis to maintain the safety and integrity of pipelines and associated infrastructures. There is an enormous amount of data available to conceal crucial information, including precursor data on failure modes and knowledge that may be analyzed. The availability of large amounts of data has enabled the development of analytical tools that integrate methods like predictive analytics using different decision-making models, artificial intelligence (AI), and machine learning. These tools are crucial for managing pipeline conditions, preventing unwarranted failures, enhancing asset performance, availability, and decision-making. Big data analytics enables energy companies to implement a proactive approach to pipeline condition assessment. By integrating real-time data from sensors embedded in pipelines, weather conditions, and maintenance records, it becomes possible to detect potential issues and predict anomalies. The application of big data analytics in the energy pipeline industry is still at its early stage, although literature review discusses big data in oil and gas, however, the application's specific relevance to energy pipeline integrity and condition assessment has been largely unexplored. Therefore, this study addresses the applications of big data analytics in energy pipeline condition assessment by investigating the challenges and benefits. Collaboration among pipeline operators, data scientists, and technology providers were emphasized for successful adoption. The study envisions a future where big data analytics will be crucial in enhancing pipeline safety, efficiency, availability, integrity, and reliability. Recommendations for further research were also provided, culminating in a proposed conceptual framework for adopting big data analytics in the oil and gas pipeline industry.

1. Introduction

Integrity and Reliability assessment of energy pipelines are critical for ensuring the reliable and safe transportation of petroleum products from one place to another. Pipelines are subjected to various environmental factors and operational conditions that can lead to deterioration and potential hazards. Ensuring these pipelines' integrity and reliability is paramount to prevent leaks, disruptions, and potential environmental hazards. Traditional approaches to pipeline condition assessment often rely on periodic inspections and conventional analysis, which is time-consuming, costly, and prone to human error. However, with the advancements in data analytics and the increasing availability of large volumes of data, adopting big data analytics for energy pipeline condition assessment has emerged as a promising solution. The term 'Big data', used in today's world that refers to high-volume data with high velocity and variety, trails back by almost eighty years. The popularization of the term 'big data' started in the 1990's in some way the credit is given to John Mashey [1], the chief scientist at Silicon Graphics in the 1990's. During that decade, the emergence of the big data application started in its true meaning [2]. Recent research topics in big data analytics include energy asset management and fault detection, predictive maintenance and monitoring of assets [3], and cloud data mining [4,5]. The use of unexpected event monitoring techniques based on real-time data (such as pressure, flow, temperature, and other monitoring data) for pipeline leak event recognition has increased gradually with the rapid development of big data analysis, artificial intelligence, machine learning, and other predictive analytics technologies [6–10].

The Oil and Gas (O&G) industry participants anticipate changes in the O&G value chain due to technological adoption [11]. The worldwide big data market's growth rate from 2012 to 2027 is illustrated in Fig. 1

* Corresponding author.
E-mail address: engr.mohdhussain@yahoo.com (M. Hussain).

https://doi.org/10.1016/j.ijpvp.2023.105061
Received 18 July 2023; Received in revised form 31 August 2023; Accepted 15 September 2023
Available online 16 September 2023

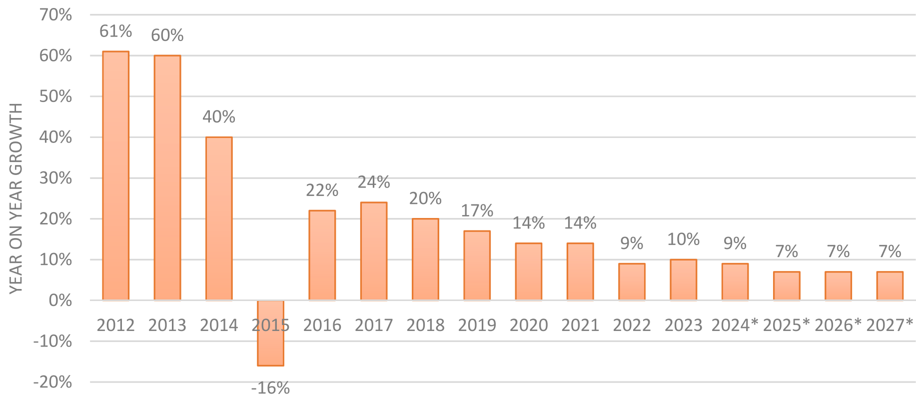

Fig. 1. Global big data market year-on-year growth from 2012 to 2027.

[12]. In 2018, the big data market is anticipated to expand by 20%, reaching 42 billion dollars. According to Wikibon [12], the global big data industry is expected to grow at a Compound Annual Growth Rate (CAGR) of 10.48%, from $42 billion in 2018 to $103 billion in 2027. Global software sales are expected to total $628 billion, with $302 billion coming from applications.

Big data analytics has emerged as a powerful tool for the condition assessment and real-time monitoring of energy pipelines. By leveraging the massive volumes of data generated by sensors, meters, and other monitoring devices installed on these pipelines. Operators can gain valuable insights into the pipeline's health and performance. Predictive maintenance is one of the primary applications of big data in pipeline condition assessment. Traditional maintenance approaches are often based on predetermined schedules or reactive responses to failures. Big Data Analytics is all about managing and analysing large amount of data in order to improve asset performance, predict anomalies, understand the current condition of the asset, reduce maintenance costs, better decision-making, and reduce workplace risks [13]. Implementing big data analytics enables proactive maintenance interventions, reduces downtime, enhances safety, and optimizes maintenance costs. This helps pipeline owners and operators to take proactive actions on time, preventing failures, and ensuring the safe and efficient operation of the system. In such instances, big data analytics help in better ways to analyse the large data effectively that collect by the system from different sources such as weather conditions, geospatial information, operating parameters, and historical maintenance record. By analysing diverse datasets, pipeline companies will get a holistic overview of the pipeline's current condition and identify correlations between different variables, enabling more accurate assessments and informed decision-making process.

Big data in energy pipelines' condition assessment plays a major role in enhancing its safety, enabling proactive maintenance, reducing downtime by predicting failures, optimizing maintenance costs by prioritizing interventions, and improving operational efficiency by providing real-time insights. As this was noticed over the past few years, more technologies have been introduced to the oil and gas sector. Big data analytics will become more crucial as this development progresses in maintaining the dependability and integrity of energy infrastructure. Therefore, big data analytics has received the most research attention trying to better tune data acquisition and processing, to get a clearer outcome. The precise prediction of corrosion and other such defects in oil and gas pipelines holds significant importance when it comes to making decisions about pipe material, predicting remaining useful life, planning maintenance, and more. Currently, various techniques for monitoring corrosion are employed in oil and gas pipeline systems. However, the data being observed often possess behaviour like nonlinearity, multidimensional quantities, and noise interference. To overcome the limitations of relying solely on mathematical models, machine learning can be effectively utilized to achieve intelligent corrosion prediction and improve the effectiveness of corrosion control [14].

Mohammad and Torabi [15] gives an overall overview of the big data analytics in oil and gas sector. Their paper provides valuable insights, however, the authors of this paper noted that there is a lack of real-world examples or case studies that demonstrate the successful implementation of big data analytics in the oil and gas industry. Although the paper mentions the need for data integration, however, it does not extensively discuss the specific challenges faced in integrating diverse data sources within the oil and gas industry. The paper briefly touches on data security concerns but does not explore into the specific privacy and security challenges associated with handling sensitive data in the oil and gas industry. Given the critical nature of the industry and the potential implications of data breaches, a more thorough exploration of data privacy and security issues would have been beneficial.

Ockree et al. [16], employed big data analytics to create AI-based production-type curves using a complex data processing. Authors of this work found few gaps in Ockree's paper, for instance, there is lack of case studies that illustrate the benefits of using big data analytics in development planning optimization or understanding pipeline current condition or calculate remaining life. Also, the paper does not discuss the future of big data analytics in development planning optimization. As big data analytics continues to evolve, it is important to consider how these advances will impact the way that development planning optimization is conducted.

Big Data was used by Cadei et al. [17] to create prediction software that predicted operational disruptions and hazard incidents during oil and gas operations. The methodology Cadei used was tested on a real-world dataset and they were able to predict operational upsets with great accuracy; however, they did not discuss the limitations of the methodology. For example, the methodology they used, would not be able to predict new types of operational upsets that have not been seen before or tested for the model. The study does not address the methodology's ability to improve upstream production systems' decision-making procedures. The methodology, for instance, can be used to identify the major sources of operational disruptions so that mitigation strategies can be developed. According to Pettinger [18], safety predictive analytics can be created using the information obtained from safety inspections. The paper briefly reviewed the use of leading indicators but lacks a comprehensive explanation of what these indicators are and how they can be measured or utilized to predict safety outcomes. Although the paper emphasizes the use of big data, it lacks an in-depth discussion of data analytics techniques and methodologies that can be employed to extract meaningful insights from large datasets. Big Data analytics was reviewed by Tarrahi and Shadravan [19], using Bureau of Labour Statistics (BLS) statistics, to manage risk and increase safety in the oil and gas industry. They reviewed the potential benefits of leveraging big data analytics for improved risk assessment, incident prevention, and overall, Health, Safety, and Environment (HSE) performance. The paper lacks specific technical details about the advanced

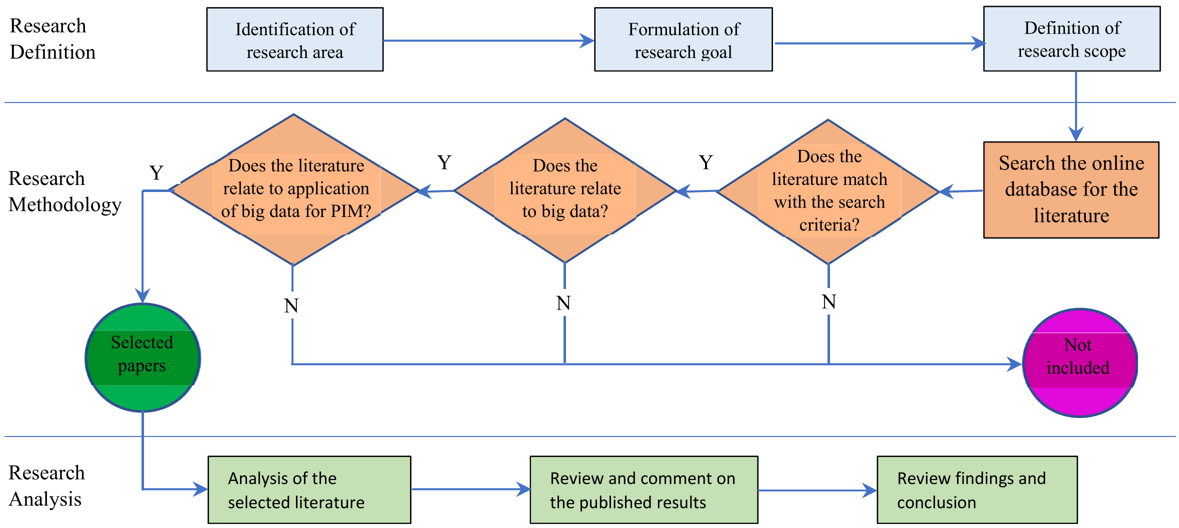

Fig. 2. Methodological framework for research adapted from Ngai et al. and A. Rachman et al.

big data analytics techniques used in HSE management. It would benefit readers to have more information on the specific algorithms, methodologies, and tools employed to analyse and extract insights from large datasets. While the paper emphasizes the use of big data analytics; however, it lacks a comprehensive discussion on data integration challenges and strategies. Integration of heterogeneous data sources is a crucial aspect of big data analytics, and addressing the complexities involved would enhance the applicability and reliability of the proposed solutions. The paper does not provide quantitative results or metrics to support the claim that advanced big data analytics improves HSE management.

Looking at above few examples, and as per the best knowledge of this paper authors, there is no paper published which is discussing big data analytics and its application in oil and gas pipeline condition assessment. The adoption of big data analytics in energy pipeline to understand pipeline current condition has the potential to revolutionize the way pipeline operators monitor, evaluate, and maintain their pipelines and associated infrastructures. By employing big data, pipeline operators can move away from reactive maintenance practices and transition towards a proactive and predictive approach, including improved safety, reduced downtime, optimized maintenance schedules, increased asset performance, asset availability, reduce maintenance cost, and increase production. A comprehensive understanding of big data and big data analytics, its definition, and its classification has not yet been fully established and are still developing. The current advancements in big data and big data analytics are not only showed a dearth of management research in the area but also a clear dearth of theoretical frameworks and academic rigor, which may be a result of a methodological rather than academic difficulty. In general, there have not been many research papers that thoroughly discuss the major big data difficulties or look at potential for novel ideas or developing methods [20].

The objective of this systematic literature review paper is to explore the adoption of big data analytics techniques to understand the current condition of energy pipeline and predict any deviation in the process or prediction of anomaly in the system due to internal or external resources. This review paper aim to investigate how advanced analytics, machine learning algorithms, and predictive modelling can enhance the accuracy, efficiency, and effectiveness of pipeline assessment compared to traditional approaches. By leveraging the large amounts of data available, there is a need to develop a data-driven framework that can provide real-time insights into pipeline health and enable proactive decision-making.

The challenges with the application of big data analytics will be identified and future research and development directions will be pointed out. This paper provides a thorough analysis of recent research work on the use of big data analytics in the oil and gas sector especially in pipeline sector. The findings of this systematic literature review will help both academics and practitioners to develop new solutions based on the challenges identified in this paper using big data analytics. With the purpose described above, this systematic literature review paper is organized as follows:

Section 2 introduces the research methodology framework that was adopted for this research work, the literature search strategy, and a brief analysis of the found and selected literature. This systematic review designed by policies and guidelines from PRISMA and Cochrane Handbook for Systematic Reviews for Interventions. Different databases like Springer, IEEE, Science Direct, and Scopus were used to collect the data using the keywords and eligible criteria (i.e. inclusion and exclusion criteria).

Section 3 provides a detail literature review, and which covered big data analytics, available tools, classification of the data in oil and gas, big data techniques, data cleaning methods and tools, application of big data in energy pipeline anomaly prediction, presented taxonomy of analytics in pipeline and discussed gaps and challenges in developing and implementing big data analytics in oil and gas pipeline sector.

Section 4 discusses the future research directions and potential of big data analytics in energy sector. A conceptual framework is presented to develop and implement big data analytics system in oil and gas pipeline industry.

Section 5 summarizes the review of overall findings and concludes the paper.

2. Methodological framework for research

2.1. Methodology

The methodological framework used in this research is adapted from Ngai et al. and Rachman et al. [21,22], which consists of three essential stages: research definition, research methodology, and research analysis (Fig. 2). Each stage is elaborated in the following sub-sections.

2.2. Protocol development

Researchers have conducted several studies on Big Data analytics in various sectors, including the energy industry, their applications, challenges, and benefits. This section presents a systematic literature review method to study Big Data analytics for energy pipeline asset management comprehensively [23]. The main benefit of systematic review over

Table 1
List of selected research resources.

	Resources	Hyperlinks
1	ACM	https://dl.acm.org/
2	Springer	https://springerlink.com/
3	IEEE	https://ieeexplore.ieee.org/
4	Google Scholar	https://scholar.google.com/
5	ScienceDirect/Elsevier	https://www.sciencedirect.com/
6	Scopus	https://www.scopus.com/
7	Web of Science	https://webofknowledge.com/
8	ASCE	https://ascelibrary.org/

Table 2
Inclusion and exclusion criteria for systematic literature review.

	Criteria	Justification
Inclusion	• Studies that focus on big data analytics that are relevant to pipeline condition assessment and asset integrity management. • Paper published online from 2000 to 2023.	• This allows to have clear picture of big data analytics approaches in pipeline asset management. • Recent and old papers can be useful in finding fundamental and classical literature as well as provides comparative analysis on the timeline of big data analytic use within energy pipeline sector.
Exclusion	• Any paper written other than English paper was excluded. • Book chapters and surveys without relevant information for this review paper were excluded. • Any grey literature published online was excluded.	• Any paper written other than English was excluded. • Book chapters and surveys that do not offer any significant information were excluded. • Grey literature is not considered academically and includes a range of documents that are not controlled by publishing organizations.

other types of studies is that it is a process that presents a taxonomical review to find the answer to the problems identified related to a specific research topic [24]. Although systematic literature review is famous among researchers of medical fields, it can be conducted in any other area of study to understand accurately the subject domain, reducing bias and identifying future directions [25]. Since most of the systematic review articles provide an unstructured procedure, the purpose of this paper is to follow a rigorous process of the literature in this scope. Hence, this systematic review paper was designed by policies and guidelines from PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) and Cochrane Handbook for Systematic Reviews for Interventions [26].

The protocol of this systematic review paper addresses the main theme, which is identifying and synthesizing previous studies related to Big Data analytics and its relation to energy pipeline condition assessment. Defining the review question is an important and essential first step in a systematic review [24]. Therefore, a focus review question needs to be developed to identify the fundamental concepts of the research topic. The review aims to answer the following research questions (RQ):

RQ 1. What are the challenges in energy pipeline condition assessment?

RQ 2. What are the applications of Big Data analytics to energy pipeline condition assessment?

RQ 3. How the challenges in energy pipeline integrity management are handled using big data analytics to predict current conditions?

RQ 4. What are the key challenges of adopting big data analytics technologies for assessing energy pipeline conditions?

RQ 5. What is an optimal energy pipeline integrity management process?

The research scope is a literature review of the application of big data analytics in Pipeline Condition Assessment. The motivation of this research study is to discover the application of big data analytics to understand the pipeline's current condition. The main contributions made in this research study can be summarized as follows:

- To study the state-of-the-art big data technologies that have been applied to the development of the oil and gas industries.
- To investigate existing research gaps in the oil and gas pipeline sector.
- To propose a conceptual framework to adopt big data analytics in oil and gas pipelines to predict pipeline current condition and develop an integrity management system.
- To recommend some potential future research directions to eliminate the gaps and promote a big data-driven industry.

2.3. Identify relevant studies databases

After formulating the review question, the next essential step of the systematic literature review is to identify relevant databases that need to be identified [27]. A detailed literature review was conducted to identify the application of big data analytics to understand the current condition of the energy pipeline and predict anomaly. Research analyses include reading and evaluating the selected papers and publications to synthesize the current body of knowledge, answer the research questions, and suggest directions for further research study in the future.

The literature review covered publications between the years 2000 and 2023 published by scholars and practitioners. The literature search was identified in different databases to collect references shown in Table 1. Later all the references were merged into a single file using MS Excel, for the analysis discussed in section 2.7.

These are electronic databases that consist of a significantly large number of articles, journals, research articles, books, and peer-reviewed papers, covering multidisciplinary studies across social science and scientific domain. Due to the multidisciplinary nature of the research on the use of pipeline integrity management, numerous publications publish pertinent articles. Consequently, the search was not restricted to a particular category of journals. The numerous publications were managed using the reference management program EndNote (endnote.com).

2.4. Inclusion and exclusion criteria

The third most important stage of systematic literature review is the identification of inclusion and exclusion criteria. Inclusion and exclusion criteria aim to represent the searching and selection process as well as filter unnecessary data sources that do not essentially require [24]. Table 2 shows the inclusion and exclusion criteria for this systematic literature review along with justification.

2.5. Search strategy and keywords

A search strategy is the key stage of a systematic literature review where keywords need to be identified for searching academic papers using electronic databases [25]. Seers [28] stated that the search strategy should be specific, clear, and based on the set inclusion and exclusion criteria to find the most appropriate academic papers from selected electronic databases. For this purpose, different mechanisms can be used to ensure a clear and specific search strategy design such as using Boolean operations.

Boolean operators allow one to find specific papers using a combination of the keywords. For example, using "AND", "OR" and "NOT" in conjunction with keywords provides a wide range of results. The search process was searching through electronic databases based on titles and

Table 3
Search keywords.

Big Data	Energy Industry	Pipeline Integrity	Other Terms
Big data OR Big Data Analytics OR Big Data Application OR Big Data Tool OR Big data process	Oil and Gas OR Energy Industry OR Energy Oil OR Gas OR Energy Sector	Pipeline Integrity OR Pipeline Asset OR Pipeline Asset Management OR Asset Management OR Integrity Management OR Reaming Life Assessment OR Condition Assessment	Cloud Computing OR Distributed Computing OR Disruptive Technology OR Technology Framework OR Data Cloud
AND	**AND**	**AND**	**AND**

keywords that were identified within search keywords which are listed in Table 3 below:

2.6. Quality assessment of included reviews

Quality assessment is essential for a systematic literature review process to reduce the risk of bias associated with the searched peer-reviewed articles [25]. For this purpose, a CASP (Critical Appraisal Skills Programme) tool [23] was used to ensure the credibility of papers included in this study.

2.7. Distribution of publications

Based on the keyword search a total number of 33,547 publications, shown in Fig. 3, were identified. Based on the identification of publications between the years 2000 and 2023 and the removal of duplicates, 23,178 publications were identified and screened using the five research questions, further screening, 7804 publications were selected using keywords. The 508 publications were further filtered by reading the publication titles for content-based inclusion.

Further filtered by reading the paper abstracts for content-based inclusion 142 publications were selected for synthesis. 03 items were inaccessible, and 139 items were judged to be related to the research's goal. Eligibility of 139 publications were checked by using the Scopus database and divided them into book chapters, journal papers, editorial materials, review papers, and conference papers. Peer-reviewed journals, Systematic Literature Review (SLR) review papers, and conference papers with subjects that were not sufficiently covered in the journal papers received preference throughout the final selection process. The documents from the Google Scholar and Google websites underwent the same procedure.

A total of 139 publications in all were considered for thorough reading and review. Fig. 4 shows the flow diagram for the preferred reporting items for systematic reviews and meta-analyses (PRISMA).

2.8. Descriptive analysis of the selected studies

A total of 139 papers were selected after meeting eligibility criteria such as inclusion and exclusion criteria. Fig. 5 illustrates the total number of papers finally selected and included in this systematic review. The figures show that the number of selected articles published each year from 2000 related to big data applications in energy sectors. It shows an increasing trend after 2013. This is mainly because the use of advanced sensors and technology development and implementation within the oil and gas industry has rapidly increased, especially after the COVID pandemic, people started finding new ways to predict the remaining life of critical, understand asset current condition, enhance asset efficiency, increase safety, and maintain its integrity.

3. Literature review and discussion

3.1. Energy pipeline condition assessment

A huge investment is involved in the infrastructure of energy pipelines; hence its integrity is a must for reliable operations and free from the risk of degradation or deterioration, which could cause expensive downtime, environmental hazards, and potential threats to life. Failures cause the pipeline to be shut down, terminating product distribution. In recent years, the increase in failure incidents has raised attention to the integrity and condition assessment of gas pipeline [29–32]. A few factors affecting pipeline condition are shown in Table 4 and some of the causes of gas pipeline failure in Europe is shown in below Fig. 6.

There are 22 reasons regarded by the American Society of Mechanical Engineers (ASME) B31.8S as posing a "Threat" to pipeline integrity, and these threats have been divided into nine categories. ASME B31.8S, Canadian Standards Association (CSA) Z662, and the American Petroleum Institute (API) 1160 have integrity threat types that are listed in Table 6 [33].

Conducting regular condition assessments of these pipelines is crucial to ensure its integrity, prevent leaks or failures, and maintain uninterrupted supply. The importance of pipeline condition is shown in Table 5.

3.2. Pipeline condition assessment methods

3.2.1. Non-destructive testing (NDT)

NDT techniques, including magnetic flux leakage (MFL), ultrasonic testing (UT), radiographic testing (RT), and eddy current testing (ECT), are commonly employed to detect defects, corrosion, and wall thickness

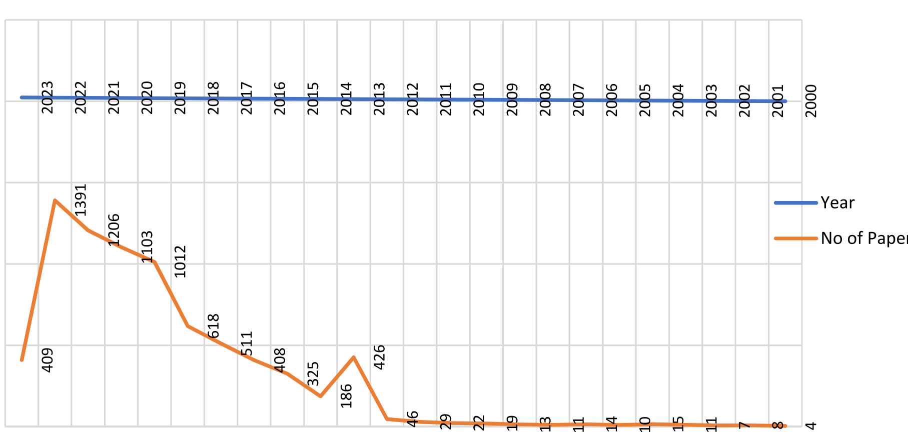

Fig. 3. The number of papers searched - 2000 to 2023.

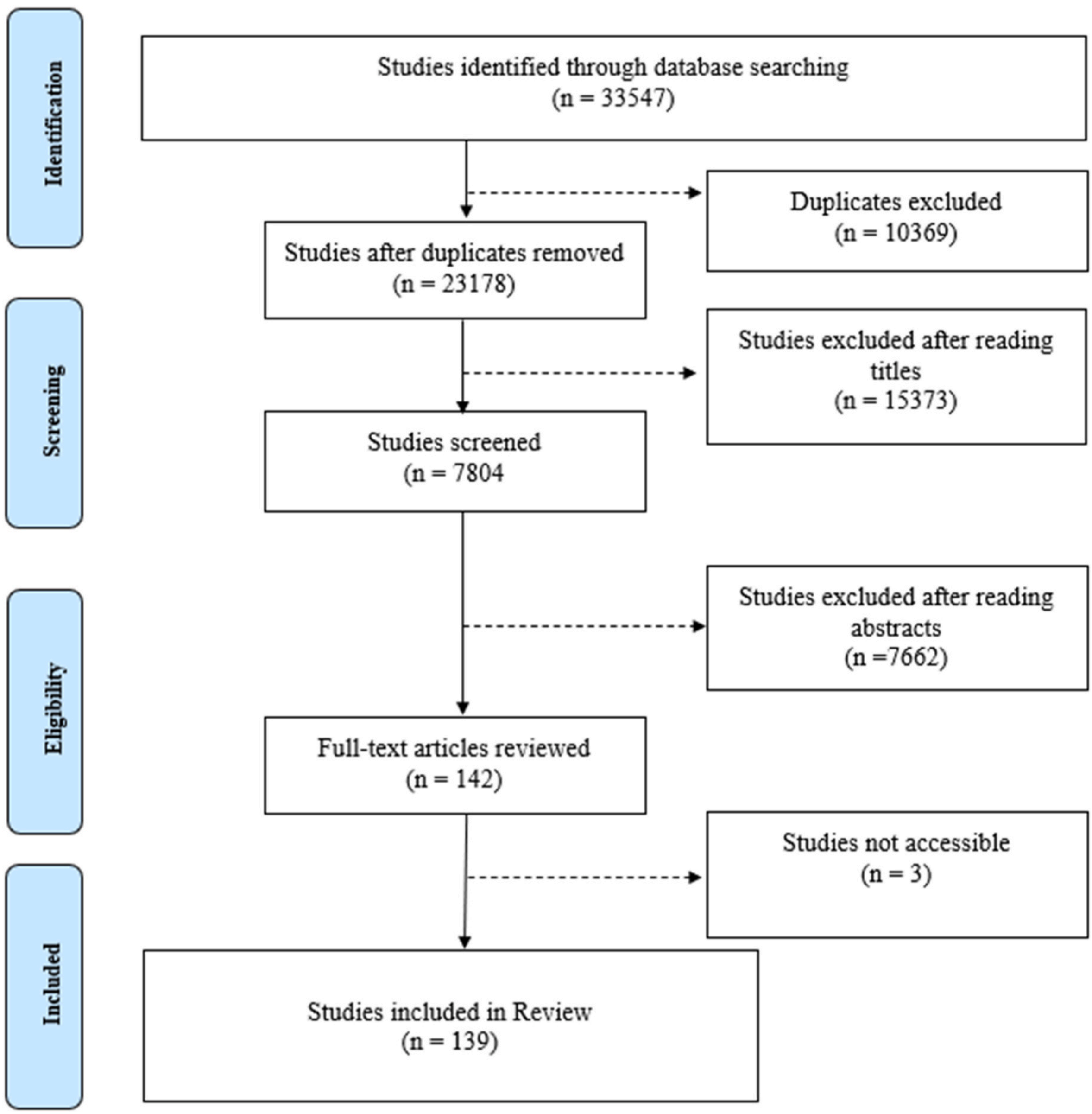

Fig. 4. Prisma flow diagram.

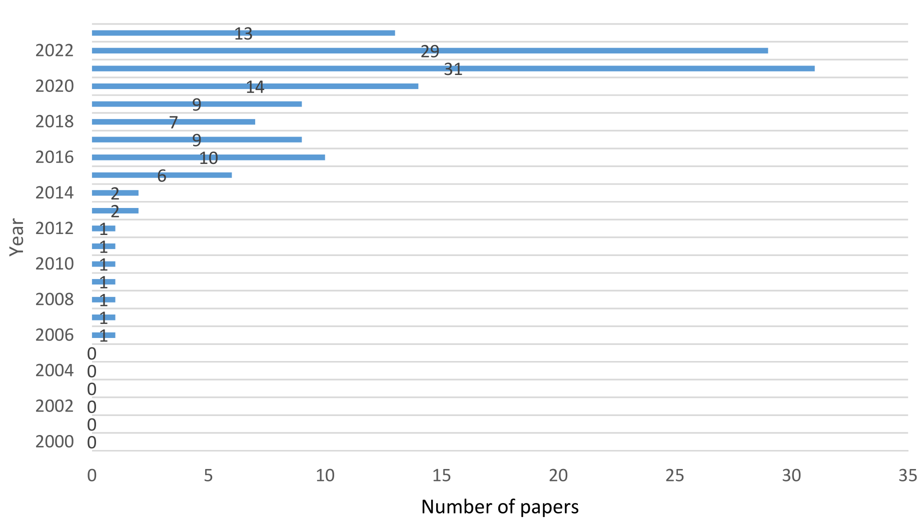

Fig. 5. The number of selected papers in this review - 2000 to 2023.

Table 4
Factors affecting pipeline condition.

Factor	Details
Corrosion	One of the primary concerns for oil and gas pipelines is corrosion, which can lead to metal degradation and structural deterioration.
Mechanical Damage	External factors such as excavation activities, natural disasters, or accidental impacts can cause mechanical damage to pipelines, compromising their integrity.
Operational Factors	Parameters like pressure fluctuations, temperature variations, and fluid characteristics can affect the condition of pipelines over time.

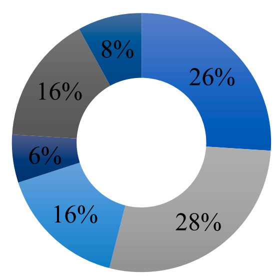

Fig. 6. Causes of gas pipeline failures in Europe [33].

Table 5
Importance of pipeline condition assessment.

Important Aspect	Details
Safety	Assessing pipeline conditions helps identify potential integrity issues that could lead to accidents, spills, or explosions, thereby safeguarding human lives and the environment.
Reliability	Regular assessments help identify areas of concern, allowing for timely repairs or replacements to minimize disruptions in the energy supply chain.
Cost Optimization	Effective condition assessment aids in prioritizing maintenance activities, reducing unnecessary expenses, and optimizing the lifespan of pipeline assets.

losses in pipelines.

3.2.2. Inline inspection (ILI)

Utilizing smart pigs or inspection tools, ILI allows for internal inspection of pipelines, providing detailed data on defects, corrosion, cracks, and other abnormalities.

3.2.3. External monitoring systems

Various external monitoring systems, such as cathodic protection, strain gauges, and leak detection systems, are deployed to monitor pipeline conditions continuously and detect potential issues.

3.2.4. Predictive modelling

Advanced data analytics and predictive modelling techniques leverage historical data, inspection results, and real-time monitoring to forecast potential pipeline deterioration and prioritize maintenance activities.

Table 6
Correlation of pipeline integrity threat names from industry standards.

ASME B31.8S Gas Transmission Pipelines	CSA-Z662 (Canada) Gas, Liquid Hydrocarbons, Oilfield Water and Steam, Liquid or Dense Phase Carbon Dioxide Pipeline Systems	API 1160 Hazardous Liquid Pipelines	UKOPA (United Kingdom) Major Accident Hazard Pipelines
External corrosion	Metal loss	External corrosion	External corrosion
Internal corrosion	Metal loss	Internal corrosion	Internal corrosion
Stress corrosion cracking—SCC (i.e., environmentally assisted cracking)	Cracking	Crack and crack-like	Other
Manufacturing-related (e.g., seam weld and pipe body)	Material or manufacturing	Design and material	Seam weld defect
Welding/fabrication related (e.g., girth weld, wrinkle bend, threads)	Construction	Construction	Girth weld and pipe defect, construction/ material
Mechanical damage (e.g., immediate/ delayed failure, vandalism/theft)	External interference	Third party damage	External interference
Incorrect operations (e.g., procedure)		Operation errors	Operator error
Weather-related and outside force (e.g., hydro-geotechnical, cold weather, lightning, heavy rains or floods, Earth movements)	Weather Geotechnical failures	Ground movement	Ground movement
Pipeline equipment (e.g., valves, seals)	Ancillary equipment	System operations	N/A

3.3. Pipeline integrity management (PIM)

To evaluate the oil and gas pipeline integrity, the integrity management program is one of the tools to predict the pipeline remaining life and assess the current condition. In other words, the pipeline integrity management system, referred to in Fig. 7, is a program using activity models, process models, and knowledge structures to maintain the pipeline and associated infrastructure integrity. The term pipeline integrity helps to understand the themes of failure prevention, inspection activities, and repair procedures, and it also involves products, practices, and services that help stakeholders improve their assets [34, 35].

Integrity Management requires considering pipeline construction, operation, maintenance records, and detection of leaks history, emergency response, and training. The integrity management strategy that considers both the likelihood or probability of failure and the consequences of a failure is focused on a known risk [36]. A pipeline integrity management system (PIMS) is a program that controls risk and cost control techniques, instruments, and activities to determine the health conditions of pipelines and plan inspection and maintenance activities [37]. The integrity management aspect only discusses how to handle the properties i.e. the pipeline system practically and efficiently [38].

3.4. Big data in the oil and gas sector

The oil and gas sector are one of the largest and most complex industries in the world, involving the exploration, extraction, refining, and distribution of hydrocarbon resources. With the advent of new

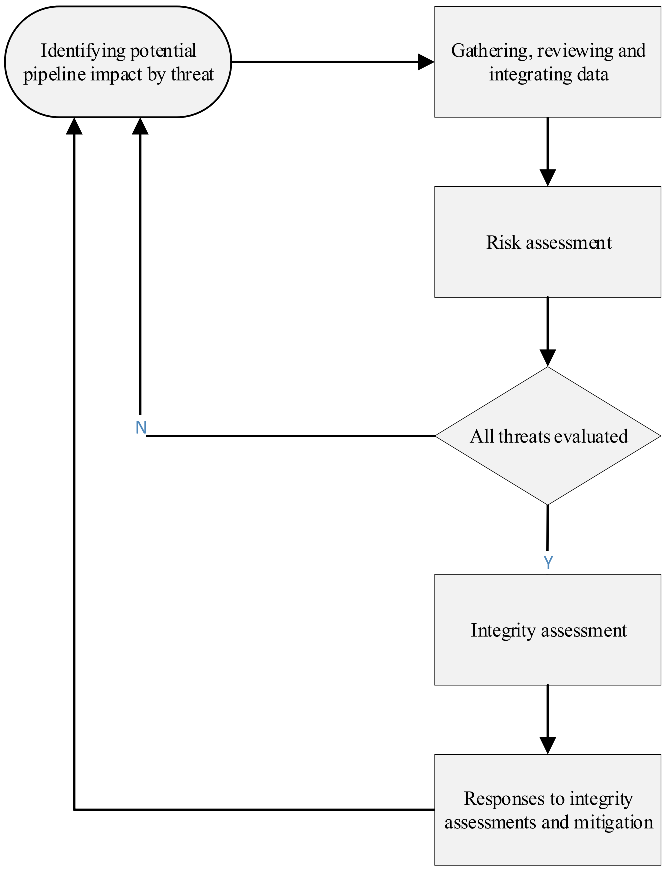

Fig. 7. ASME B31.8S-2014 integrity management plan process flow diagram.

Table 7
The amount of data captured by various sectors of the oil and gas industry.

Data Collection Section	Amount of Data
Pipeline inspection	1.5 TB (600 km)
Plant operational	8 GB (annually)
Vibration	7.5 GB (per customer annually)
Plant process	4–6 GB (daily)
Submersible pump monitoring	0.4 GB/well (daily)
Drilling	0.3 GB/well (daily)
Wireline	5 GB/well (daily)
Seismic	100 GB–2 TB/survey

technologies, there has been a massive increase in the amount of data generated within this sector, giving rise to the concept of "big data." Big data refers to the large volumes of structured and unstructured data that are generated at high velocity and variety, Fig. 11. Big Data Analytics is now one of the most important aspects of the oil and gas industry's digital transformation [39]. With the aid of big data technology, the energy sector can control the massive volumes of data, for instance, Table 7 shows data that was reviewed by Subramaniyan, A. in 2017 [40, 41].

In the oil and gas sector, big data plays a crucial role in improving operational efficiency, optimizing exploration and production activities, reducing costs, and enhancing safety. Table 8 shows some key areas where big data is making a significant impact.

In conclusion, big data has emerged as a transformative force in the oil and gas sector, enabling companies to unlock valuable insights, improve operational efficiency, reduce costs, and mitigate risks. As technology continues to evolve, the effective utilization of big data will become increasingly essential for staying competitive in the dynamic and data-intensive oil and gas industry.

Table 8
Impact of big data in oil and gas.

Key Areas	Big Data Impact
Exploration and Reservoir Analysis	Big data analytics allows oil and gas companies to process and analyse seismic data, well logs, and other geospatial information to identify potential drilling locations and assess the viability of reservoirs. This helps in minimizing risks, improving success rates, and maximizing the recovery of hydrocarbon resources.
Predictive Maintenance	By analysing large volumes of operational data, including sensor readings and maintenance records, big data analytics can predict equipment failures before they occur. This allows oil and gas companies to schedule maintenance activities proactively, minimize unplanned downtime, and reduce maintenance costs.
Asset Performance Management	Big data analytics allows oil and gas companies to analyse historical and real-time data on asset performance, including drilling rigs, pipelines, and refineries. By identifying patterns and trends, companies can optimize maintenance schedules, improve asset reliability, and extend the lifespan of critical infrastructure.
Predictive Analytics and Decision Support	Big data analytics provides valuable insights for decision-making in the oil and gas sector. By leveraging advanced analytics techniques, such as machine learning and artificial intelligence, companies can predict reservoir behaviour, optimize drilling techniques, assess investment opportunities, and make informed business decisions.
Supply Chain Management	Big data analytics helps optimize the complex supply chains in the oil and gas sector. By analysing data related to inventory levels, transportation logistics, demand forecasts, and market trends, companies can improve procurement processes, minimize inventory holding costs, and ensure timely delivery of products.
Safety and Environmental Monitoring	Big data analytics can analyse data from various sources, such as sensors, drones, and satellite imagery, to monitor and detect potential safety hazards and environmental risks. This includes identifying leaks, assessing air quality, monitoring pipeline integrity, and ensuring compliance with regulatory standards.
Production Optimization	Big data analytics enables real-time monitoring and analysis of production processes, equipment performance, and sensor data from oil wells and refineries. This helps in identifying bottlenecks, predicting equipment failures, and optimizing production rates to enhance efficiency and reduce downtime.

3.5. *Big data analytics for energy pipeline*

Big Data analytics is expected to play a crucial role in energy pipeline asset management by providing valuable insights and enabling proactive decision-making. Fig. 8 shows that the big data for the energy pipeline is extracted from single or multiple components of data namely, design data, operational data, attribute data, inspection data, and maintenance data. These categories are depicted in Table 9.

Some key applications and benefits of Big Data analytics in energy pipeline condition and integrity assessment are listed in Table 10.

Overall, Big Data analytics empowers energy pipeline asset managers with actionable insights, improved decision-making, and efficient resource allocation, ultimately enhancing the safety, reliability, and performance of pipeline networks and associated infrastructure.

3.6. *Classification of big data in the oil and gas sector*

In the oil and gas sector, big data can be classified into various categories based on different aspects and applications, Fig. 9 shows a

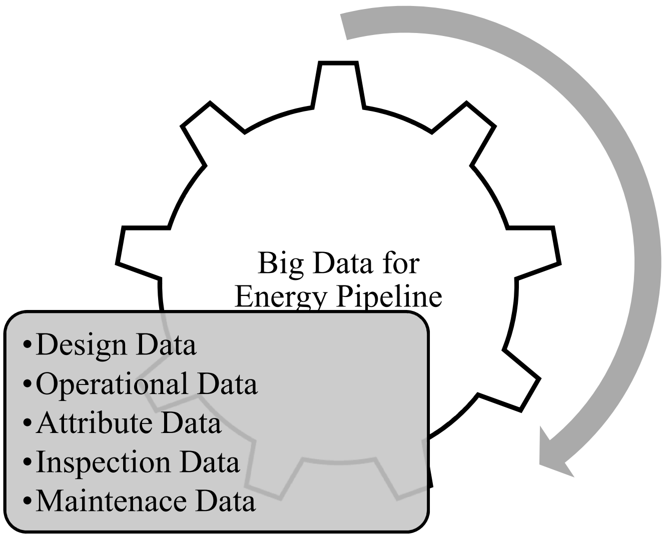

Fig. 8. Data utilized in Pipeline Asset Management.

classification of big data in the oil and gas sector.

3.7. Big data analytics techniques

Many latest technologies are in use to detect defects and enhance the integrity of pipelines. Over the years, from these inspection and maintenance activities, huge amounts of data have been gathered. These vast volumes of data provide us with great value, but it's also a great challenge to understand and deal with this data. The conventional way of managing data, including data storage, processing, and management, is unable to meet the current needs of industrial activities [42–44]. Hence, well-developed systems that can deal with large data volumes require the appearance of a new 'Big Data Analytics' concept. According to a market research report in 2018 by Research and Markets, the pipeline integrity management market is expected to exceed $ 10.4 billion by 2023 [45]. This means with the advancement in technologies and increased sensors, higher volume, velocity, and variety of, valued data generation or big data generation. Energy companies use thousands of sensors mounted on its critical assets including energy pipelines to collect data and track properties and environmental conditions in real-time [45]. Energy companies need tools that incorporate and synthesize disparate data sources into a single whole to support real-time decision-making. The ability to process Big Data allows for the extraction of information from the relationships that emerge when all these sources are processed together. However, to realize this benefit, these businesses must have access to the necessary technology, resources, and expertise. Table 11 showcases various Big Data Analytics techniques used in pipeline monitoring and condition assessment.

3.8. Data clean-up methods and tools

To develop well advanced analytical system, data cleansing is one of the crucial steps. Data cleaning is basically a process to ensure that the data used for analysis and modelling is accurate, consistent, and reliable. Few common methods for data cleaning is listed in Table 12 and some common tools for cleaning are listed in Table 13.

3.9. Big data analytics tools

It is critical to select tools and software that will assure the greatest results when performing high-quality data analysis [46]. Table 14depicted some of the most common types of data analysis tools.

Table 9
Co-related data.

Category	Data
Design Data	• Year of Construction • Joining Method • Hydro Test • Materials Type • Materials Specification • Field Coating Methods • Coating Type • Bending Method • Soil, backfill • Cathodic protection (CP) installed • Quality Control Document • Quality Assurance Document • Final delivery inspection reports
Operational Data	• Gas quality • Flow rate • Normal maximum and minimum operating pressure • Leak/failure history • Coating condition • CP system performance • Pipe wall temperature • Pipe inspection reports • OD/ID corrosion monitoring • Pressure fluctuations • Regulator/relief performance • Encroachments • Repairs • Vandalism • External forces
Attribute Data	• Pipe wall thickness • Diameter • Seam type and joint factor • Manufacturer • Manufacturing date • Material properties • Equipment properties
Inspection Data	• Pressure tests • In-line inspections • Geometry tool inspections • Bell hole inspections • CP inspection (CIS) • Coating condition inspections (DCVG) • Condition Monitoring Data • Non-Destructive Testing (NDT) Data • Audits and reviews
Maintenance	• Maintenance Repair Cost • Downtime • Production Losses • Repair Cost • Man- days loss

3.10. Evaluation criteria

Brule [45] stated that the oil and gas sector spends over half of its time managing datasets. For this purpose, big data is rapidly becoming one of the most important aspects of the oil and gas industry's digital transformation [60]. Anand [61] discussed the benefits of big data and how it can now uncover a lot of hidden information from the large amount of data that is available in the oil and gas sector. Table 15 provides an evaluation criterion for big data analytics techniques in pipeline condition assessment.

3.11. Taxonomy of analytics in pipeline asset management

The pipeline is considered the heart-line of oil and gas transportation which triggered an increasing demand for advanced and sophisticated mechanisms to handle the implementation of pipeline projects, asset management, operation, and maintenance [62]. Because of the existence of the atmosphere and high temperatures, metallic pipelines are prone to corrosion, which is a leading cause of pipeline defects. As a result, pipeline operators must use reliable and efficient intelligent

Table 10
Applications and benefits of Big Data analytics in energy pipeline condition.

Application/Benefits	Brief Details
Predictive Maintenance	By analysing historical and real-time data from sensors and other sources, Big Data analytics can identify patterns and anomalies that indicate potential equipment failures or maintenance needs. This enables the implementation of predictive maintenance strategies, where maintenance activities are scheduled based on the actual condition of the assets rather than fixed intervals. This approach reduces downtime, improves asset reliability, and minimizes maintenance costs.
Condition Monitoring	Big Data analytics allows for continuous monitoring of asset conditions by processing large volumes of sensor data. By analysing trends and patterns, operators can detect early signs of asset deterioration or performance degradation. This enables timely intervention and prevents potential failures, ensuring the overall health and longevity of the pipeline infrastructure.
Real-time Monitoring and Alerts	Real-time analytics on streaming data from sensors and internet of things (IoT) devices can provide immediate insights into pipeline operations. By setting up automated alerts and notifications, operators can quickly respond to abnormal conditions, such as pressure drops, leaks, or equipment malfunctions. Real-time monitoring improves operational efficiency, enhances safety, and minimizes the risk of incidents.
Risk Assessment and Mitigation	Big Data analytics enables comprehensive risk assessments by analysing a wide range of data sources, including historical pipeline data, environmental conditions, geological information, and asset metadata. By identifying potential risks, such as corrosion, ground movement, or external threats, operators can implement targeted mitigation measures to minimize the probability and impact of incidents.
Optimization of Asset Performance	By integrating various data sources, including operational data, maintenance records, and external factors like market conditions or demand forecasts, Big Data analytics facilitates the optimization of asset performance. Advanced algorithms can identify opportunities for operational improvements, such as optimizing flow rates, reducing energy consumption, or maximizing throughput, leading to cost savings and increased efficiency.
Regulatory Compliance and Reporting	Big Data analytics can help streamline regulatory compliance by capturing, analysing, and reporting relevant data to regulatory authorities. This ensures adherence to safety, environmental, and operational standards, reducing the risk of penalties or legal issues.

methods to identify and locate pipeline defects. A well-developed system that can deal with large data volume requires the appearance of a new 'Big Data Analytics' concept [63]. The availability of new technology based on commodity hardware that is Big Data and analytics, has evolved in recent years. Big Data Analytics is rapidly becoming one of the most important chapters in the oil and gas industry's digital transformation. The application of big data and analytics in the oil and gas industry is in the experimental stage but it can boost pipeline integrity and asset management. It can help develop analytical resources and skills for incorporating and aligning a range of datasets for enhanced defect characterization and evaluation of coincident anomalies for better decision-making [64,65]. Big data allows more sophisticated integrity assessments that require fewer repairs, as well as more precise estimation of corrosion growth rates to extend remaining life and re-inspection intervals [66]. Furthermore, it improves transparency by monitoring and visualizing operational and integrity KPIs through digital dashboards [37,66]. Thus, big data analytics aids in timely identification, accompanied by classification, and remains the best means of prevention for pipeline companies Worldwide.

In today's world, it is relatively simple to collect all the data in one location and then use it for the analysis. This ensures that all the assets can be viewed together. Instead of a risk-based inspection approach, a predictive maintenance strategy can be implemented in a pipeline's asset management by incorporating machine learning-based predictive analytics, real-time pipeline tracking, and cloud computing. This will save time as well as money that would otherwise be spent on costly repairs and inspections. Big data analytics also support the maintenance of inspection schedules, allowing pipeline operators to reduce operating costs in the long run [67].

The taxonomy is shown in Fig. 10 below. There are six most discussed and used big data analytics tools in previous literature regarding the energy sector. Further, these six analytics tools are divided into five major purposes. This big data analytics taxonomy for managing the assets of energy pipelines is based on two main components, one based on the technologies and the other on the goal. The tools as discussed in previous sections include business intelligence tools, statistics analysis, SQL consoles, data visualization, Apache Hadoop, MongoDB, Cassandra, and Spark. The purposes include Predictive analytics, Descriptive analytics, prescriptive analytics, real-time analytics, and diagnostic analytics. Predictive analytics is to find patterns and relationships in data using statistical approaches. A machine learning big data framework was presented by Hasan, Shamsuddin, and Lopes [68], which envisioned the big picture of machine learning in relation to big data issues. Forecasting and statistical modelling are key components of predictive analytics, which use supervised, unsupervised, and semi-supervised learning models to predict potential outcomes [69–71]. Descriptive analytics provides a detailed plan description of how to perceive the asset management strategy, predictive analytics entails evaluating data to anticipate the future outcome for pipeline asset management [57]. Prescriptive analytics is data processing that recommends the best course of action for managing energy pipeline assets while considering all pertinent criteria [58]. Therefore, this is often known as a valuable tool for making data-driven decision-making process [60]. In real-time analytics for energy pipeline asset management, a large volume of incoming data can be stored and at the same time used for real-time feeds [61]. Finally, diagnostic analytics help engineers and other experts in this domain to determine the correlations between different variables for the root cause and establish what happens to monitor and control the process [63].

3.12. Analysis of big data analytics in energy pipeline

The outcomes of the systematic review of this research endeavor are discussed in this section. The primary goal of this paper was to investigate big data analytics for the future of pipeline condition assessment and monitoring and provide stakeholders with a better overview of operations, predict current condition, and more control and flexibility for managing their assets [72]. As shown in Fig. 3, over the past decade, there has been a significant increase in research and publications exploring the utilization of big data analytics in the pipeline industry. Numerous academic papers, industry reports, and conference proceedings discussed the applications of big data, highlighting the growing interest and recognition of the potential benefits of big data analytics in the oil and gas sector.

The literature covers a wide range of aspects related to big data applications in oil and gas pipelines, including pipeline monitoring, predictive maintenance, safety enhancement, operational efficiency, risk assessment, data integration, maintenance cost, and visualization. Researchers have investigated various analytics techniques, such as machine learning algorithms, statistical models, data fusion methods, and optimization approaches, to extract insights and optimize pipeline operations. The published literature provides experimental studies, and theoretical frameworks to demonstrate the effectiveness of big data analytics in addressing challenges to energy pipelines. These challenges include pipeline integrity management, condition assessment, performance, leak detection, security and intrusion detection, environmental impact assessment, and emergency response planning. Additionally, the literature discusses the importance of data integration, data quality

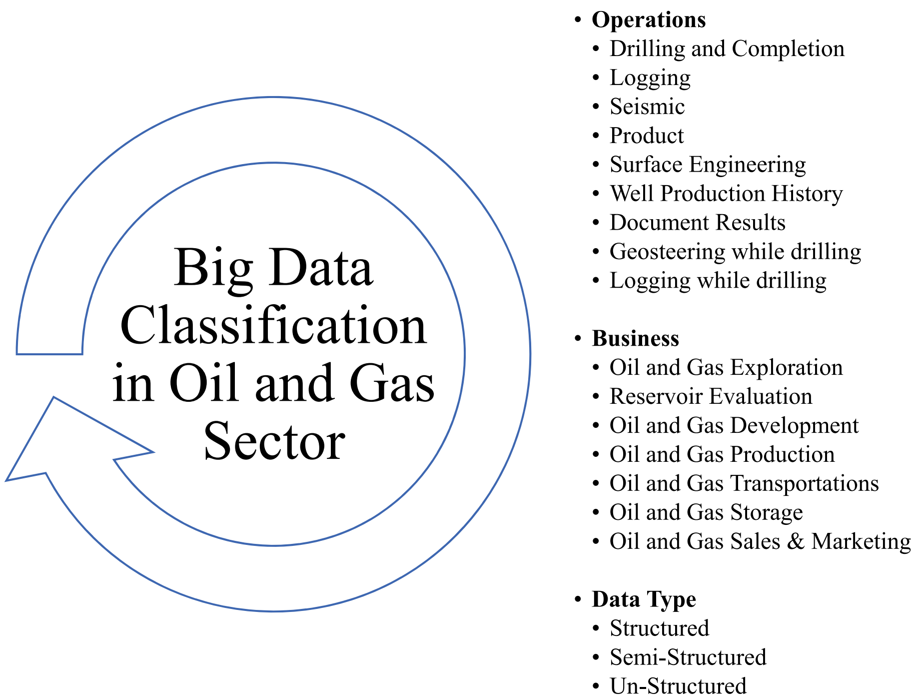

Fig. 9. Classification of big data in the oil and gas sector.

Table 11
Big data analytics techniques.

Technique	Description
Data Integration	Integrating diverse data sources, such as sensor data, operational data, and external data.
Data Fusion	Combining multiple data sources to create a unified view of the pipeline system.
Data Harmonization	Standardizing and aligning data from different sources to ensure consistency and compatibility.
Real-time Monitoring	Monitoring pipeline parameters in real time, such as pressure, temperature, and flow rates.
Anomaly Detection	Identifying abnormal patterns or deviations from normal operating conditions.
Predictive Maintenance	Using historical and real-time data to forecast potential failures and optimize maintenance.
Machine Learning	Employing algorithms that learn from data to make predictions or take automated actions.
Statistical Analysis	Analysing data using statistical methods to uncover trends, patterns, and correlations.
Visualization	Presenting data in a visual format, such as graphs, charts, and dashboards.
Data Mining	Extracting meaningful insights and patterns from large datasets.
Text Mining	Analysing and extracting valuable information from unstructured text data.
Pattern Recognition	Identifying recurring patterns or sequences in data.
Predictive Analytics	Applying statistical models and algorithms to make predictions and forecast future outcomes.
Clustering	Grouping similar data points together based on their characteristics or attributes.
Association Rule Mining	Discovering relationships and dependencies between variables or events in the data.
Sentiment Analysis	Evaluating and determining the sentiment or opinion expressed in textual data.
Deep Learning	Utilizing neural networks with multiple layers to learn complex patterns and make predictions.
Natural Language Processing	Analysing and understanding human language to extract information and meaning.

assurance, data governance, and cybersecurity in the context of implementing big data analytics in oil and gas pipelines. Researchers have reviewed different data integration approaches, such as data lakes, data warehouses, and real-time data streaming, to consolidate and harmonize data from various sources, enabling comprehensive engineering assessment analysis and decision-making. As new technologies emerge and industry practices advance, there is ongoing research to explore emerging topics, such as the integration of the IoT, cloud computing, and edge computing, and AI in pipeline engineering analytics. These advancements aim to further enhance the capabilities of big data analytics and address existing challenges in the industry [73].

Most of the oil and gas storage and transportation companies currently use a three-tiered approach to monitor the security of their storage systems and pipelines. This approach involves gathering temperature, pressure, and other data from long-distance pipes and transmitting it in real-time to a data monitoring centre via the Internet, where it is then analyzed and processed in real-time for potential early warning situations [74]. In recent years, the increase in failure incidents has raised attention to the integrity and condition assessment of gas pipeline [62,63,75].

Big data is a broad range of structured, semi-structured, and unstructured datasets, as the phrase "big data" suggests. In other words, big data analytics [76,77] is the process of extracting important conclusions, patterns, and information from enormous and intricate databases. It entails applying cutting-edge methodologies and tools to analyse enormous amounts of structured, semi-structured, and unstructured data to find patterns, correlations, trends, and other important data that can help with decision-making, enhance business operations, or spur innovation. Unstructured data are those datasets that do not have proper classifications or labels, whereas structured data does. Most of the data in the oil and gas sector is unstructured, which makes analysis quite challenging. The technology for unstructured data analysis is still in its infancy, which makes it more challenging to analyse the data and predict any issues. Even though that 80% of data are unstructured, Garter claims that the amount of data generated will increase by 800% over the next five years [78]. Big data analytics projects can produce intangible and tangible business value that refer to as the benefits [79]. The oil and gas sector may reduce operating costs and have a beneficial influence on all phases of the oil and gas lifecycle by leveraging the capabilities of big data and analytics technologies [80,81].

Detecting anomalies and accurately predicting future behavior during operations will enable more effective decision-making that can help

Table 12
Data cleaning methods.

Data Cleaning Methods		Brief Details
Outlier Detection and Handling	*Statistical Methods*	Statistical methods for data cleaning involve using various statistical techniques to identify and handle anomalies, errors, and inconsistencies in a dataset. These methods rely on the underlying distribution and statistical properties of the data to detect unusual or erroneous observations.
	Visualization	Using scatter plots, box plots, and other visualizations to identify data points that deviate from the norm or defined operating parameters.
	Transformations	Applying transformations like logarithmic scaling to mitigate the impact of outliers.
Missing Data Handling	*Imputation*	Replacing missing values with estimated or calculated values.
	Deletion	Removing rows or columns with missing data, if they are not critical for analysis.
	Predictive Modelling	Using other variables to predict missing values based on patterns in the data.
Inconsistent Data Handling	*Standardization*	Converting data to a consistent format, such as converting dates to a common format.
	Normalization	Data normalization is a data cleaning technique used to transform numerical data into a standard scale, making it easier to compare and analyse across different variables.
	Regular Expressions	Cleaning text data using pattern matching and substitution.
Duplicate Data Removal	*Exact Matching*	Identifying and removing exact duplicates.
	Fuzzy Matching	Detecting approximate duplicates using similarity measures.
Data Validation	*Cross-Field Validation*	Checking relationships between fields to identify inconsistent or incorrect data.
	Domain-Specific Validation	Ensuring that data adheres to the constraints and rules of the specific domain.

Table 13
Data cleaning tools.

Cleaning Tools	Brief Details
Pandas	A popular Python library for data manipulation and analysis, offering tools for cleaning, pre-processing, and transforming data.
Microsoft Excel	Widely used for basic data cleaning tasks, such as removing duplicates, filtering data, and basic transformations.
OpenRefine (formerly Google Refine)	A powerful open-source tool for data cleaning and transformation, particularly useful for cleaning messy, large datasets.
Trifacta Wrangler	A user-friendly data preparation tool that assists in cleaning, transforming, and structuring data for analysis.
R	A programming language commonly used for statistical analysis and data manipulation, with packages for data cleaning.
SQL	Used for data cleaning tasks involving database querying and manipulation.
DataRobot	An automated machine learning platform that also provides data preparation tools for cleaning and transforming data.
Open Data Kit (ODK)	Primarily for mobile data collection, it includes features for data validation and cleaning.

Above methods and tools for data cleaning is basically depends on the type and nature of the collected data, the extent of cleaning required, and the goal of the analysis.

Table 14
Big data analytics tools.

Analytical Tools	Tools Brief Description
Business Intelligence (BI) Tools	BI tools allow you to process large amounts of data from a variety of sources in any format. It provides a full-service solution, including cutting-edge data analysis, key performance indicators (KPI) visualization, live dashboards, and reporting, as well as artificial intelligence technology to predict trends and reduce risk [46,47].
Statistical Analysis	These tools are often created for data scientists, statisticians, market researchers, and mathematicians to do complex statistical analyses using methods such as regression analysis, predictive analysis, and statistical modelling [47].
SQL Consoles	Structured Query Language (SQL) is a programming language that is commonly used in relational databases to manage structured data. These technologies include a visual tool for database modelling and monitoring, comprehensive SQL optimization, administrative tools, and visual performance dashboards to track key performance indicators [47].
Data Visualization	Data visualization tools are used to depict data in the form of charts, graphs, and maps, allowing analysts to spot patterns and trends [47].
Apache Hadoop	Originally developed in Java [48], Hadoop makes use of a sizable computer cluster for distributed processing [49]. Hadoop tracks unstructured and sentiment data from a variety of sources, including social media, geolocation, and sensor and machine data. It is intended specifically for the storing of large amounts of data and the execution of concurrent processes. Hadoop can also be utilized to improve security and compliance [49,50].
MongoDB	MongoDB is a technology that has the ability to regulate and handle unstructured and chaotic data such as documents, multi-media, and social media. MongoDB is incompatible with applications that demand durability, isolation, consistency, and atomicity, such as database-level transactions and the design of a bank's core financial system [51]. Additionally, MangoDB offers a dynamic and adaptable structure that may be tailored to meet the needs of different users [52–55].
Cassandra	It is more effective when more time can be spent studying a complex system that provides a lot of power and flexibility. Cassandra is utilized in real-time analytics, streaming data, and fraud and criminal detection [56].
Spark	An open-source framework for handling enormous amounts of data, Apache Spark is made for speed and sophisticated analyses. You may simply construct apps in Scala, Java, or Python with Spark [57]. Large-scale datasets that are structured, semi-structured, or unstructured can all be handled using Spark, an incredibly powerful technology. In contrast, Spark saves data in memory, reducing the number of read and write cycles [58]. Due to its broad range of capabilities, many research fields, including pattern mining and machine learning, have adopted Spark as a processing engine to handle Big Data difficulties [59].

best focus operational spending on risk reduction. To predict pipeline failures and conditions, multiple models have been developed over the years. Most of these models, however, used corrosion characteristics to determine the condition of pipelines as the single factor [82]. Frequent checks are mandatory to keep the pipeline in a reliable and safe condition. In the last decade, numerous inspection techniques, like Ultrasound and Magnetic Flux leakage (MFL), have been developed. These methods provide precise and efficient information for the energy pipeline operators to identify pipeline defects that could lead to catastrophic failure. In-line inspection (ILI) is another very useful and proficient tool, that detects and/or predicts pipeline anomalies [83].

Gu, J. et al. [84] presented the pipeline lifetime state monitoring framework which provides lifetime analysis based on machine learning and understanding the optimal solution in terms of making advance decisions to avoid risk that will result in serious consequences and

Table 15
Evaluation criteria for big data analytics techniques.

Evaluation Criteria	Description
Accuracy	The degree of precision and correctness in predicting pipeline conditions or identifying anomalies.
Reliability	The consistency and stability of the results produced by the analytics techniques over multiple evaluations.
Scalability	The ability of the technique to handle large volumes of pipeline data and scale with increasing data sizes.
Computational Efficiency	The speed and efficiency of the technique in processing and analyzing data within acceptable time frames.
Interpretability	The ease of understanding and interpreting the results generated by the analytics technique.
Flexibility	The adaptability and versatility of the technique to different types of pipeline data and varying monitoring requirements.
Robustness	The ability of the technique to handle noisy or incomplete data and still provide reliable insights and predictions.
Resource Requirements	The hardware, software, and infrastructure resources needed for implementing and running the analytics technique.
Integration Capability	The extent to which the technique can be integrated with existing pipeline monitoring systems or data management platforms.
Cost-effectiveness	The balance between the benefits provided by the technique and the associated costs in terms of implementation and usage.

endangering personal safety. Sarker S. et al. [85] reviewed Big Data opportunities and challenges in different industries. They discuss issues related to data collection, storage, processing, and analysis, including data quality, data integration, data privacy, and security concerns. They give an overall benefit of big data to different industries including energy pipelines; however, they did not discuss the application of big data in the pipeline industry. Swetapadma et al. [86] proposed a system for defect detection based on decision trees. By analysing the magnetic flux leakage data gathered from the scanning of oil and gas pipelines, Layouni et al. [87] proposed a method to detect metal loss by applying pattern-adapted wavelets and two ML algorithms, i.e., artificial neural network and linear regression. They also discussed the numerical and non-numerical techniques to determine the defect lengths, depths, and location of Metal-loss in the oil and gas pipeline. Lv et al. [88] discussed the data types, storage models, data security, and privacy along with the benefits of big data networks. Nguyen et al. [72] reviewed the idea of big data in the oil and gas sector, the opportunity, and related difficulties. However, they did not discuss the advantages and disadvantages of each big data analytics paradigm for Industry 4.0. Layouni M. et al. [87] identifies metal-loss problems from the MFL scans of oil and gas pipelines and estimates their size using machine-learning approaches. The suggested solution, which achieves high levels of accuracy and computing efficiency, is based on pattern-adapted wavelets and artificial neural networks.

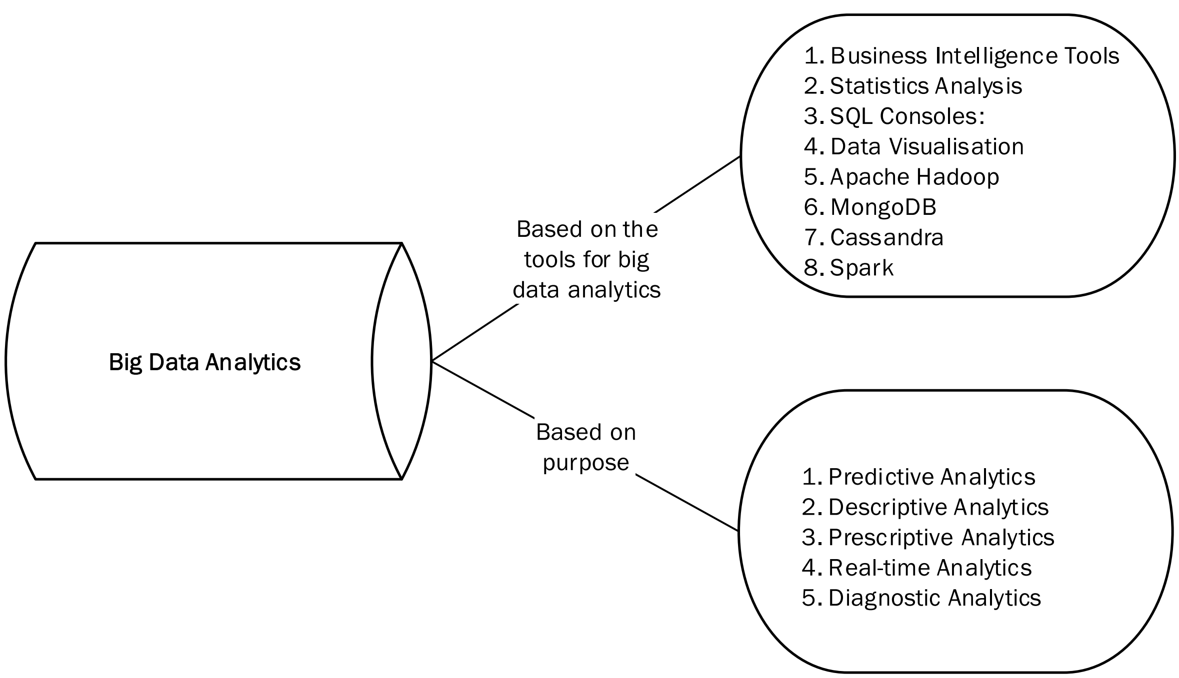

Fig. 10. Taxonomy in big data analytics for energy pipeline asset management.

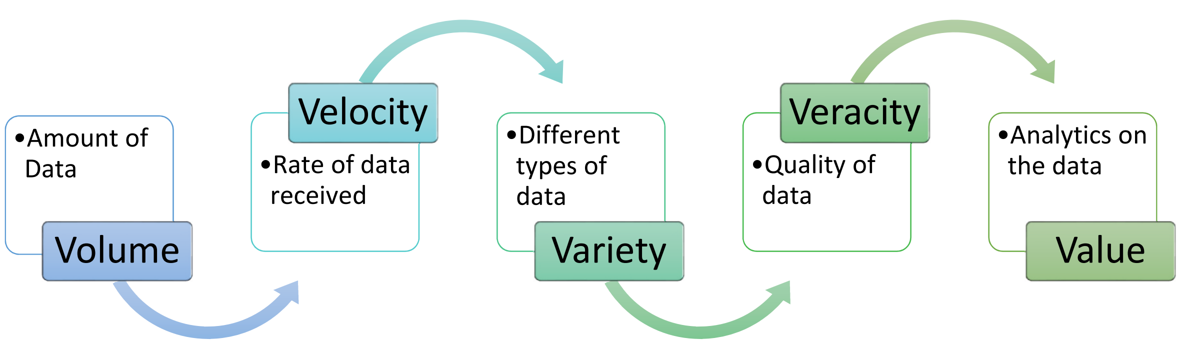

Fig. 11. The 5V's of big data [27].

Table 16
Big data 5Vs of big data in the energy sector.

Industry Type	5V's	Data Type
Big Data's 5VS in Production and Transportation	Volume	• Production operations' sensor data • Transportation operations' sensor data • Automated SCADA • Archived company data and processing
	Velocity	• Real-time data streaming of data from operation-specific sensors • Location-based varied data stream for big data analytics • Evaluation-based varied data stream (flow, pressure, temperature, etc.) for big data analytics
	Variety	• Structured (sensory data) • Unstructured (images, maps, audio, video) • Semi-structured (operational data analysis reports)
	Veracity	• Operation-specific data cleaning and segregation • Amalgamation of all data sources for a quality-rich dataset • Combination of selected geological locations' data for specific analysis
	Value	• Reduced safety hazards & environmental impacts • Optimum and enhanced recovery of hydrocarbons • Economic viability determination • Operation-based optimization
Big Data's 5VS in Energy (Natural Gas)	Volume	• Extraction operations' sensor data • Treatment operations' sensor data • Storage and transportation operations' sensor data • Distribution operations' sensor data • Automated SCADA • Archived company data and processing
	Velocity	• Real-time data streaming of data from activity-specific sensors • Equipment-based varied data stream for big data analytics • Evaluation-based varied data stream (flow, pressure, temperature, etc.) for big data analytics
	Variety	• Structured (sensory data) • Unstructured (images, maps, audio and video) • Semi-structured (operational data analysis reports)
	Veracity	• Activity-specific data cleaning and segregation • Amalgamation of all data sources for a quality-rich dataset
	Value	• Reduced safety hazards & environmental impacts • Optimum and enhanced natural gas generation operations • Economic viability determination and operational optimization

Using big data analytics in the energy sector reduces costs and implementations through different new solutions to improve the efficiency of pipeline asset management through the use of fewer resources and more delivery. As more hardware-based tools require more engineering work, and in this case, spending more time on the analysis is not a cost-effective approach for many companies [72,89,90]. The monitoring of energy pipelines serves a variety of applications beyond production measurement, including security, preventative and predictive pipeline maintenance, pipe leak detection, equipment control, and automation systems. For pipeline surveillance, location detection and knowledge of the pipeline route are crucial. Using geographic information systems and global positioning systems makes it easier to locate pipeline incidents and trace reported occurrences [91,92]. Corrosion, environmental conditions, human error, installation and erection, and manufacturing are a few of them. Today, pipeline monitoring frequently makes use of wireless sensor networks (WSN) [93–95]. WSN, IoT, big data, and AI work together to improve smart monitoring and enable remote access to data received from the pipeline [96–101]. To reduce the danger of corrosion, erosion, wear, and tear, the data collected via IoT must be processed using the proper framework for decision-making [102]. AI has been used to forecast pipeline defects [103], simulate two-phase flow in pipes [104], and predict the rate of degradation in pipe fittings [105].

Many studies have shown the advantages of big data implementation in pipeline asset management. There was a correlation found between the lifetime of ground pipelines and the big data of pipelines through various model-fitting methods [106]. The method of analysing pipeline performance based on collected data statistics is generally accepted [29,32]. The implementation of the safety factor parameter in well-established evaluation criteria is limited because the parameters are static, while environmental elements, infrastructure components, and material makeup vary from pipeline to pipeline [42]. Detecting anomalies and accurately predicting future behaviour during operations will enable more effective decision-making that helps best focus operational spending on risk reduction. Using big data analytics, they can spend fewer resources and improve the decision-making process [66].

Rouhanizadeh, B., & Kermanshachi, S [107]. work on a predictive model using the multiple regression method to forecast the performance of sewer pipes. The model considers historical data, including past failures and hydraulic information, as input. The results of the model indicate that it is particularly sensitive to the age of the pipeline system and the type of pipes used. The model was able to assist pipeline owners in identifying pipe segments that are more prone to failure, enabling them to develop a comprehensive risk management system for sewer pipeline networks.

Cobanoglu et al. [85] conducted a systematic review of big data and provided a comprehensive overview of the current state of research in big data, examining various research approaches and highlighting future prospects. They discuss various methods and techniques used for data collection, storage, processing, and analysis, including machine learning, data mining, natural language processing, and visualization in different industries, such as healthcare, finance, marketing, transportation, and social media. It explores the impact of big data analytics on decision-making processes and highlights the potential benefits and risks associated with its utilization. Therefore, using big data provide support in term of promoting "digital energy" to enhance operational and business performance. Hassani and Silva [75] presented the 5 V's of big data such as volume, velocity, variety, veracity, and value as illustrated in Fig. 11 and big data 5Vs of big data in the energy sector in Table 16 [60]. This is a process of big data analytics which extract a significantly large volume of a wide variety of data. Any big data system is expected to deliver one or more of them for energy pipeline asset management through a cost-effective mechanism and its architecture to extract information from a very large volume of data [15]. However, the successful exploitation of big data requires the alignment between business services and IT services so that the necessary infrastructure can be established [89].

Today oil and gas industry including pipelines and processing facilities all make use of various sensing devices, creating the heaviest and largest IoT networks on the globe. Big data analytics has emerged as a powerful tool in the field of pipeline monitoring, enabling the processing and analysis of large volumes of data generated by various sensors and monitoring systems. By leveraging advanced analytics techniques, researchers and industry professionals can extract valuable insights from this data to enhance the safety, efficiency, and reliability of pipeline infrastructure. The use of improved monitoring systems, more sensor data collection, larger storage volumes, quick data processing, and improved decision-making processes are a few examples of how technology is being applied to understanding the current state of a gas pipeline in order to develop strategies to extend asset life and prevent

Table 17
Application of big data analytics in pipeline monitoring.

Application	Application Details
Data Collection and Integration	- Big data analytics in pipeline monitoring starts with the collection and integration of diverse data sources, including sensor data, operational data, maintenance records, and external data such as weather conditions and geographical information. - Integration techniques, such as data fusion and data harmonization, enable a unified view of the pipeline system, facilitating comprehensive analysis.
Real-time Monitoring and Anomaly Detection	- Big data analytics allows for real-time monitoring of pipeline parameters, such as pressure, temperature, flow rates, and vibration levels, providing early detection of abnormal conditions. - Advanced anomaly detection algorithms analyse streaming data to identify deviations from normal operating patterns, enabling proactive maintenance and mitigating potential risks.
Predictive Maintenance and Asset Management	- By analysing historical and real-time data, big data analytics can facilitate predictive maintenance models that forecast potential failures or degradation of pipeline components. - Predictive analytics algorithms and machine learning techniques help optimize maintenance schedules, reduce downtime, and extend the lifespan of pipeline assets.
Data-driven Decision Making	- Big data analytics enables data-driven decision making by providing actionable insights and visualizations to support risk assessment, operational planning, and asset integrity management. - Visualization tools, dashboards, and interactive interfaces allow stakeholders to easily interpret complex data and make informed decisions.
Continuous Improvement and Optimization	- Big data analytics facilitates the continuous improvement of pipeline monitoring systems through the analysis of historical data, identifying patterns, trends, and areas for optimization. - By applying machine learning algorithms and statistical models, pipeline operators can optimize operational parameters, enhance energy efficiency, and reduce costs.

the pipeline from corrosion or other defects [84].

Pipeline owners and operators began start making plans to keep checking and maintaining their critical assets for corrosion or other damages [15]. Many companies expanded their current workforces and came up with plans to make up for the lost revenue. The year 2023 is bringing even more fast change and digital transformation after a year of enormous upheaval. In order to forecast the current status of assets including pipelines, and develop more flexible and robust asset management systems, oil and gas companies are accelerating the implementation of artificial intelligence with machine learning for data analytics [108]. Below Table 17 provides a brief overview of the application of big data analytics in pipeline monitoring and assessment.

3.13. Gaps and challenges

Leveraging big data in the oil and gas sector also poses challenges, such as data integration from diverse sources, data quality assurance, data privacy, and security concerns, and the need for skilled team to understand and analyse the data. Addressing these challenges requires robust data governance frameworks, advanced analytics capabilities, and collaboration between domain experts and data professionals. A comprehensive understanding of big data analytics in energy pipelines, its definition, applications, architecture requirements, and classification has yet to be fully established because big data as a research topic is still in the developing stage. The current advancements in big data analytics showed an absence of management research and an apparent shortage of theoretical frameworks and academic consistency, which may result from methodological rather than academic difficulty. Few research studies have addressed the significance of big data difficulties and discussed the potential for novel ideas or emerging approaches. Some limitations and gaps were identified after reviewing the literature on big data applications in oil and gas pipelines. These limitations are classified and discussed as shown in Table 18.

4. Future research opportunities

The viability of using big data platforms and technologies and a rise in the efficacy and efficiency of operations are a few long-term possibilities for big data. It's fascinating to observe that the big data generated by the energy sector including pipelines is not yet genuinely big data because only 3–5% of the infrastructure in this sector is currently connected [171], making big data-based jobs (including financial analysis, transportation, and drilling) unprofitable. Additionally, given that simply realizing the full potential of the available big data can increase operational efficiency by 20% [75], increasing connection by 3–5% can have significant advantages in the long run [60]. Over the past ten years, cloud services have become more popular, viable, and cost-effective. The use of cloud services is currently being heavily engaged in by platforms like Amazon Web Services, Google Cloud Platform, and Azure (Microsoft). A major benefit of such platforms is that complicated, and computationally intensive investigations may be carried out quickly and easily, as was already indicated. These developments enable the entire sector to adopt cloud-based big data implementation (storage, maintenance, and analysis). The oil and gas industry now starting to adopt digital technologies mostly from the bottom up and with asymmetrical implementation [172]. As a result, businesses are not utilizing digitalization and data analytics (DT) to their full potential. A thorough grasp of DT technology, the present state of oil and gas-related data analytics research activities, and the opportunities and challenges connected with adopting DT in the oil and gas sector are required to realize the full potential of DT and related technical adoption [173].

Even though technology has been employed in the oil and gas business for decades, it has typically come with high installation costs, making it uneconomical to use in real-world situations. This is no longer the case, though, as sensors are now lightweight, affordable, and connected to secure wireless networks, making it possible to use them in conjunction with big data analytics to acquire useful insights for better decision-making across all oil and gas activities. According to a survey by International Data Corporation (IDC) Energy, the main barrier to the adoption of big data in the sector is the lack of knowledge about the advantages of the subject outside of its proper comprehension in the sector, which has a negligible impact on business support [109]. This is currently changing, however, as a subsequent Mehta survey [141] conducted five years later in 2018 reveals that 81% of volunteered executives viewed big data to be one of the top three objectives for their energy sector in the next years. This obvious increase in big data understanding and acceptance is undoubtedly something to watch for in the future. It should be noted that, despite the clear advantages of using big data in the oil and gas industry, such as "model-based beforehand" efficiency improvements and real-time data analysis, other advantages, such as near-real-time visualizations, data storage (and management), and near-real-time anomaly detection (for improved safety), actually rank as the biggest advantages of doing so [174]. These elements make the application of big data's future possibilities exciting and perhaps probable.

Furthermore, given that the pipeline industry is infamous for lagging behind other sectors in terms of the adoption and integration of technology [175], the situation is similar in terms of the full deployment of big data. However, given the numerous benefits listed above and the extensive study done on it, there is potential for its acceptance in the coming decade. In fact, it has a lot of room to expand practically in the energy pipelines [60]. Despite the claim made in one of the challenges

Table 18
Challenges and issues with the application of big data use.

Challenges/Issues	Author's Name	Reference	Year of Publication	Journal/Conference Indication	Outcome/Findings
Corporate support	J Feblowitz	[109]	2013	SPE Digital Energy Conference	The largest obstacle to utilise big data in the oil and gas industry is a lack of corporate support.
Finance/Budget	R Beckwith	[110]	2011	Journal of Petroleum Technology (JPT)	Budget is a big issue for most of the companies, as most of the companies in oil and gas industry are unaware of big data analytics and its application to their assets, hence they are reluctant to allocate any budget.
	N Mounir et al.	[111]	2018	Offshore Technology Conference	Oil prices are highly volatile these days, especially after Covid19. Such fluctuations in the oil price can significantly impact on the profitability and long-term planning of oil and gas companies.
	U Sivarajah et al.	[39]	2017	Journal of business	Even with the availability of cloud computing solutions, hardware is still expensive, which is another issue that oil and gas companies need to come with a solution to address this problem.
Skills Set	J Feblowitz	[109]	2013	SPE Digital Energy Conference	An inadequate level of business understanding and support, as well as a limitation of personnel expertise in the wide range of oil and gas sector related to big data and petroleum
	Phillips-Wren & Hoskisson	[112]	2015	Journal of Decision Systems	Shortage of qualified individuals who can evaluate data analytically is another problem.
	GH Kim, S Trimi, JH Chung	[113]	2014	Communications of the ACM	Lack of resources and understanding of big data is a big challenge on oil and gas industry.
	S Priyadarshy	[114]	2017	Book Chapter (John Willey & Sons)	Due to the advent of new technology and, in some cases, the retirement of experienced labour, the oil and gas sector is experiencing a scarcity of skilled field specialists and workers.
	MS Sumbal et al.	[115]	2019	Journal of Knowledge Management	Oil and gas companies have to rely on consultants because they lacked people with the necessary technical expertise for big data analysis.
	A Alhosani, and S. M. Zabri	[116]	2018	International Journal of Advanced and Applied Sciences	The industry is further challenged by the lack of suitable creative solutions to organise and manage knowledge because of the dispersed nature of information.
Development Issues	J Thiyagalingam et al.	[117]	2022	Nature Reviews Physics	One of the major problems with development of predictive analytics system is to develop required architecture and system.
	S Kaisler et al.	[118]	2023	ScholarSpace	How to scale back current algorithms and systems or need to develop new ones while maintaining a certain level of accuracy and precision in the outputs is a crucial issue.
	E Barbierato, M Gribaudo, M Iacono	[119]	2014	Future Generation Computer Systems	Inadequate infrastructure and insignificant data warehouse architecture is a big issue to develop advanced big data analytics system in energy pipeline sector.
	Al Nuaimi et al.	[120]	2015	Journal of Internet Services and Applications	Data processing is expansive that could have an impact on how businesses adopt and use technological solutions.
Understanding the algorithms					Expertise in this field and an in-depth understanding of science, technology, and mathematics are few required for a data scientist to develop and understand ML modelling. Given the complexity of machine learning, data scientists must configure and optimize the algorithms to achieve the best results. Many businesses need more internal expertise to comprehend algorithms and how they operate, which can cause them to lose out on crucial insights.
Technological difficulties	Maidla et al.	[121]	2018	IADC/SPE Drilling Conference and Exhibition	The technological problems were mostly caused by the constraints of the data-recording sensors from the field. The frequency of data recording and data quality that was collected were the other problems.
	H Jiang et al.	[122]	2015	Cluster Computing	Synchronising large data from different sources is not an easy task.
	J House	[123]	2014	Supply Chain Strategy	More effective predictive analytics system and routing with visualization and real-time tracking during shipments, and highly optimized distribution network management are all made possible by big data analytics.
	Preveral et al.	[124]	2014	SPE Intelligent Energy International Conference and Exhibition	Each business should create its own unique Big Data tools, including data recording and storage facilities and data analytic tools.
	R Aliguliyev et al.	[125]	2016	Problems of information technology	Each data cluster contains a sizable amount of data that is problem-oriented and necessitates extra pre-processing.
Software/Tools	DP Acharjya and K Ahmed	[126]	2016	International Journal of Advanced Computer Science and Applications	Technical challenges arise in the deployment of big data technologies employing readily accessible software tools and hardware computing platforms. There is not a perfect big data model available right now that guarantees a high-profit improvement while taking time and financial limits into account. The performance of current large data

(continued on next page)

Table 18 (*continued*)

Challenges/Issues	Author's Name	Reference	Year of Publication	Journal/Conference Indication	Outcome/Findings
					analysis tools is subpar when dealing with computational complexity, ambiguity, and inconsistency. The development of methods and tools that can effectively handle computational complexity, ambiguity, and inconsistencies presents a significant challenge.
Regulatory requirements	J Gohil and M Shah	[60]	2022	Book (CRC Press)	Every country has its own set of governmental regulations governing the execution of geoscience, production, transportation, drilling, and refining activities. These regulatory restrictions were primarily put in place to protect the environment, increase safety and to make sure that geoscientists don't exceed a certain level of carbon emissions.
Third-party attacks	J Schmidt	[127]	2019	CET Journal-Chemical Engineering Transactions	Given that the oil and gas industry is seeing an increase in cyber-attacks, protecting the pipeline system from outside threats like third-party damages is essential. In order to develop and implement a successful protection system, like PipeSecure2020 are being started to specify new levels of protection for gas pipelines.
Interoperability	S Sun et al.	[128]	2020	Sensors Journal	The integration between assets in the field and software systems, across energy pipeline system and a collaboration between these systems require the timely exchange of accurate information. In managing a complex information technology environment for the integration of information technology (IT) and operational technology (OT), this is likely to present difficulties. The handling of the exchange of real-time and non-real-time data, as well as the lack of a uniform data standard that enables information processing, is not an easy task.
	A Gandomi, M Haider	[129]	2015	International journal of information management	Data integration is one of the significant issues in big data analysis successful development and implementation.
	M Igamberdiev et al.	[130]	2018	Software & Systems Modeling – Springer	A multi-level paradigm for data interoperability in the energy pipeline business is a crucial part due to complexity of the oil and gas system and complex heterogeneous data structures.
Decision Making Model					Despite having high-quality data, using the right analysis tool and model can make or break a report or insight. This presents a significant problem. There are hundreds, if not thousands, of available analysis models and tools; selecting the right one can be a time-consuming and laborious effort; as a result, the team needs experts who can make wise decisions on model usage based on their knowledge and experience.
Type of dataset	MT Muhammed, WJ Obidallah, R Bijan	[131]	2018	Archives of Information Science and Technology	To enhance the application of deep learning algorithms on different types of data that is collected from various sources, these algorithms are mainly applied to large image datasets, showcasing their effectiveness. They also highlight that the current trends in research on deep learning algorithms for big data, providing valuable insights for researchers aiming to comprehend the cutting-edge utilization of these algorithms.
	R Toshniwal, KG Dastidar, A Nath	[132]	2015	International Journal of Innovative Research in Advanced Engineering	To start using big data, the big issue for oil and gas companies is to pick and prioritize the most pertinent and crucial data. Organizations must be able to distinguish the pertinent data from the irrelevant data when there are such large amounts of data.
	A Divyakant, B Philip, B Elisa	[133]	2012	Proc. VLDB Endowment	The big issue to fully utilizing big data analytics include heterogeneity, incompleteness, scale, and timeliness.
	A Oguntimilehin, EO Ademola	[134]	2014	Journal of Emerging Trends in Computing and Information Sciences	
Lack of Training Data	U Sivarajah et al.	[39]	2017	Journal of business	Cleaning data is another challenge for the data that is collected from different sources, and most of the data collected in oil and gas is unstructured data, and cleaning of unstructured is not an easy task.
	R Du Mars	[135]	2012		Categorizing, modeling, and mapping the data as it is recorded and stored is a big challenge in the process of managing big data, mostly because of the complex and unstructured nature of the collected data.
	Shah et al.	[136]	2015	Cluster Computing	Outdated data analysis and modeling focus to identify complex relationships between data and the outcome. To get the best results from this relationship, data cleaning is the first step to adopt and get best from the collected data.
Data Transfer	M Xie, Z Tian	[37]	2018	Engineering Failure Analysis – Elsevier	They identified that more reliable and effective real-time data processing and analysis techniques needs to be developed for noise removal in ILI data that is collected from energy pipeline. Advanced pipeline integrity

(*continued on next page*)

Table 18 (*continued*)

Challenges/Issues	Author's Name	Reference	Year of Publication	Journal/Conference Indication	Outcome/Findings
					management frameworks based on machine learning and advanced technologies also need to be established.
	Chen et al.	[137]	2013	Frontiers of computer Science	The requirement to manage high influx data rate of non-homogenous is a complex process.
	MF Uddin, N Gupta	[138]	2014	Proceedings of the 2014 Zone 1 Conference of the American Society for Engineering Education - IEEE	The data transfer between system for big data analytics, privacy rights is an issue to be addressed by the oil and gas companies.
	Al Nuaimi et al.	[120]	2015	Journal of Internet Services and Applications	To transfer and sharing the information between distant organizations or departments is not an easy task.
	MT Muhammed, WJ Obidallah, R Bijan	[131]	2018	Archives of Information Science and Technology	Securing big data presents challenges due to its vast volume, diverse variety, rapid information flow, and susceptibility to potential threats.
	AA Tole	[139]	2013	Database Systems Journal	Access to and sharing of information is a problem that is indirectly related to the preceding one. There are also institutional and technical problems when data is stored in places and formats that make it hard to get to it and move it.
Data Volume	N Khan et al. A Mehta	[140] [141]	2014 2016	The scientific world journal, Journal of Petroleum Technology	Managing large and rapidly growing volumes of data has been a difficult problem for decades. In the past, this issue was mitigated by faster processors, which provide the resources necessary to deal with growing data volumes.
	D Che et al.	[142]	2013	International Conference on Database Systems for Advanced Applications	Fast access to and mining of big data are not merely desirable qualities; they are also a requirement, particularly for data streams (a popular big data format), as the value of the results of processing/mine decreases or even vanishes after a certain amount of time has passed.
Data Storage	Y Gidh et al.	[143]	2016	SPE Intelligent Energy International Conference and Exhibition	Digitalization of energy industry uses many sensors which generate huge amounts of data every day. Storage and protocol of the data is one of the issues that needs to be addressed to get best from the collected data.
	P Neri	[144]	2018	Offshore Technology Conference	
Data Quality Issue	O Kwon et al.	[145]	2014	International Journal of Information Management	They considered factors like managing data quality, and benefits of using big data analytics to enhance decision-making capabilities, affirms the vital influence of internal factors in driving big data initiatives.
	K Al-Barznji and A Atanassov	[146]	2016	Proceedings of 24th International Symposium, "Control of Energy, Industrial and Ecological Systems"	Data qualities include elements such as completeness, accessibility, and context, in addition to data correctness. Data can be both structured and unstructured. 80% of the data produced by organizations is unstructured. They might not be saved in row-column format as structured data. The upcoming big data research will face a lot of difficulties because of this high level of heterogeneity.
	A Cuzzocrea	[147]	2021	IEEE International Conference on Big Data and Smart Computing	The significant diversity, along with the substantial volume and rapid pace of contemporary big data, creates a crucial challenge for future big data research: designing, implementing, and operating an effective and efficient infrastructure for handling large datasets.
	J Merino et al.	[148]	2016	Future Generation Computer Systems	They emphasize the importance of recognizing data quality to determine its suitability for anomaly detection or prediction. Managing Data Quality is impacted not solely by data characteristics but also by analytical procedures, particularly the emerging technologies that facilitate novel means of bolstering these analytical processes.
	D Loshin	[149]	2013	Book	There is a necessity to identify crucial dimensions of Data Quality that hold significance for data handled within Big Data projects.
	M Hussain et al.	[150]	2021	11th International Conference on Modelling in Industrial Maintenance and Reliability (MIMAR)	Data quality is one of the significant factors to be considered in evaluation of the condition of pipelines using big data analytics.
	RY Wang, DM Strong	[151]	1996	Journal of management information systems	This study establishes a hierarchical framework designed to encompass key aspects of data quality essential for predictive analytics assessment.
	N Abdullah et al.	[152]	2015	International Journal of Advances in Soft Computing and its Applications	Data quality is a vital attribute governing data's dependability for decision-making.
	I Lee	[153]	2017	Business horizons -Elsevier	To create a good quality metrics, analyse data quality, correct inaccurate data, and weigh the costs and benefits of quality assurance, a data quality control process must be established by oil and gas companies.
	RK Mobley	[154]	2002	Book	The goal is to use diverse procedures and technologies on growing larger and more complicated data sets to gain insights into the data's health
Data Privacy	GA Gow	[155]	2005	Network Societies, International Telecommunication	The most delicate issue is privacy, which has conceptual, legal, and technological implications. This concern elevates its significance within the context of big data.
	S Kaisler et al.	[156]	2013	Hawaii International Conference on System Sciences	

(*continued on next page*)

Table 18 (*continued*)

Challenges/Issues	Author's Name	Reference	Year of Publication	Journal/Conference Indication	Outcome/Findings
	R Toshniwal,	[132]	2015	International Journal of Innovative Research in Advanced Engineering	
	K Al-Barznji, A Atanassov	[146]	2016	Proceedings of 24th International Symposium Control of Energy, Industrial and Ecological Systems	Access to data from various systems is a significant benefit of big data, and security and privacy will be crucial to big data research and technology.
	N Khan	[140]	2014	The Scientific World Journal	An eminent concern within the realm of Big Data revolves around data leakage, posing a significant threat to privacy.
	Lu et al.	[157]	2014	IEEE Network	Security is a significant problem and contend that if security issues are not properly handled, the phenomena of big data analytics in the oil and gas will not be widely accepted.
	Krishnamurthy & Desouza	[158]	2014	Information Polity	Organizations are having trouble to handle data privacy, which is one of the concerns that oil and gas companies are behind to use big data.
	J Moreno et al.	[159]	2016	Future Internet	The issue of data privacy is perhaps the one the issue that a business needs to consider on priority to avoid any data breach.
Data Security	S Phuyal	[160]	2020	Sustainable Futures	The security of data and information from the sending source, communication connection, network, and receiving node with global identification and end-to-end data encryption is required for the sharing of information.
	A Oguntimilehin and EO Ademola	[134]	2014	Journal of Emerging Trends in Computing and Information Sciences	
	I Lee	[153]	2017	Business Horizons -Elsevier	Without a proper security measure in place, private data may unintentionally reach unintended recipients.
	RT Kouzes et al.	[161]	2009	IEEE	Data inconsistency and incompleteness, scalability, timeliness, and security are difficulties in big data analysis.
	D Agrawal et al.	[162]	2011	Technical Report	
	P Barnaghi, A Sheth, C Henson	[163]	2013	IEEE Intelligent Systems	Data privacy and security is a big challenge in big data analytics implementation.
	S Wang et al.	[164]	2016	International Journal of Distributed Sensor Networks	Cyberattacks have the potential to cause accidents and environmental degradation. Significant financial loss, harm to the company's reputation, and potential public outcry could result from this. Therefore, every energy firm will continue to place a high importance on making ongoing efforts to safeguard each node of the network and adopt cybersecurity standards against outside assaults and data misuse. The Energy Pipelines companies' cyber environment's expansion to make use of advanced technologies could put businesses and their assets at danger of hacking. Attacks on business confidential information may result in significant financial losses or legal issues.
Data Ownership	D Cameron	[165]	2014	SPE Intelligent Energy International Conference and Exhibition	Among the challenges in integrating big data for oilfield service companies are staff comprehension and data ownership concerns.
Data Visualization	H Chen, RHL Chiang, VC Storey	[166]	2012	MIS Quarterly	To select a proper tool for data visualization is a crucial step due to data complexity, type and condition of available data.
	A Graves	[167]	2015	Future Generation Computer Systems	
	Chen and Zhang	[168]	2014	Information sciences	Numerous big data algorithms are in use to analyse the data, scalability, and response time. This is particularly problematic when undertaking data visualization.
	J Taheri et al.	[169]	2014	Future Generation Computer Systems	Data visualization involves employing diverse visual representations, such as a pictorial or graphical arrangement, to more instinctively and efficiently represent important information and knowledge. Due to complexity of the asset and process, this needs a good review to visualize the data.
	Wadhwani, K. and D. Y. Wang	[170]	2014	Technical Report	Data visualization helps us identify outliers and auxiliary data by showing us where they are located. To ensure that the data is clean, businesses should be using a data control, surveillance, or information management procedure.

raised earlier that low petroleum prices [176] prevent the industry from investing in data science personnel, some [75,171,177,178] contend that the low prices may encourage the investment in big data (and present a significant future opportunity) to make better and more positive operational decisions. These are different perspectives on the same thing, yet this positive aspect and the combination of all the aforementioned possibilities of an incredible future potential is unmatched. Fig. 12 compiles all the elements relating to the potential application of big data.

By putting various parts of vast sensory datasets to use with the aid of big data analysis, big data has the potential to increase the lifespan of expensive equipment used by asset owners. With big data support, businesses may proactively decide when to maintain the appropriate equipment to increase both the equipment's performance and lifespan and to ensure that as few operational hiccups or halts as possible are caused by maintenance. Using IoT data analysis, Shell company was able to save $1 million in Nigeria [179].

The prediction capabilities that big data carries with it because of technology like AI and accompanying predictive models are one of the main advantages of employing it on energy pipelines. This is very

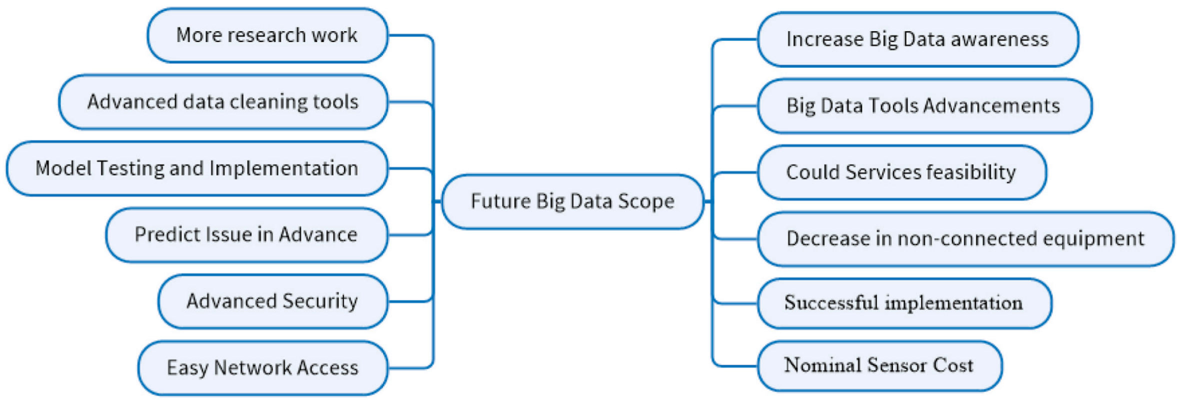

Fig. 12. Future scope of Big Data's implementation in energy pipeline.

beneficial for almost every aspect of pipelines because it allows for the most precise understanding possible by feeding enormous amounts of live operations data to the AI model(s), which then provides an accurate picture of the current condition of the pipeline for each operation. Using such patterns, if there are any variations in pipeline operating parameters, a model can detect the operating window and will calculate the impact on the overall pipeline's remaining life or materials deterioration. Additionally, big data offers direct advantages due to its predictive capabilities, including the ability to forecast operational performance (using big data-based AI analysis) and increase production output (via big data tools and technologies). These advantages help increase the product-to-profit ratio, high-frequency time series prediction, malfunction forecasting, optimal operating condition prediction of issues or damages, enhance safety, predict remaining life, and understand current condition while having a direct impact on industry outcomes including timeline, environmental impact, and profitability.

Such applications offer optimization (by increasing efficiency), risk reduction, performance improvement(s), avoidance of protracted legal disputes, resource waste (that could be used elsewhere, like on high-quality equipment), and a significant amount of time [180–182]. Beyond financial gains, the optimization (by increasing efficiency), risk reduction, and performance improvement(s) that big data delivers in operations are particularly beneficial to oil and gas organizations. Additionally, it facilitates the efficient application of predictive maintenance software, which examines the sensory big data from operation-based gear sensors to identify abnormalities (such as corrosion, fatigue cracks, fault detection, safety hazards identification, and stress corrosion cracking) before they reach a critical level. Through careful observation and proactive equipment maintenance, this all contributes to better supply chain management and lower energy use. Keeping a careful eye on things and doing routine maintenance on the underlying pipeline or its components, assists in preventing circumstances that could significantly damage the viability of operations in advance [183]. This has great financial benefits. The use of big data also aids in the continuous and continuing maintenance and monitoring of all ongoing activities in real-time (including environmental monitoring, infrastructure maintenance, and working environment [184], while making sure that everything runs without a problem. It should be noted that routine maintenance and monitoring also aid in danger prevention through the use of automated alarms, fault/breakage detection, leak detection, and malfunction alarms, among other features [184,185].

Big data's inherent optimization in all processes where it is used is a significant application in addition to its frequently discussed attribute. Geologic, operational, and performance data are collected in vast quantities, and analysis of this data enables several major and minor changes to be made in each aspect of energy pipelines' activities, ensuring that each aspect's efficiency is increased while lowering risks and reducing maintenance costs. This is especially beneficial because ongoing equipment monitoring, optimization, and maintenance can result in significant performance and financial benefits. These advantages are easily discernible by remembering that an O&G company can lose $180,000 in annual income for every 0.5% decrease in efficiency in a natural gas compression operation [186]. Fig. 13 illustrates the summary of big data applications in the oil and gas sector [60].

4.1. Future research potential using big data in oil and gas pipelines

The future research potential and development using big data in oil and gas pipelines offer several exciting possibilities, Table 19 shows some examples, where detailed research is required to get the best from big data analytics to understand the current condition of energy pipelines and predict any issues or remaining life for safe pipeline operations.

4.2. Proposing a conceptual framework for the adoption of big data analytics in the pipeline

The future research potential and development using big data in oil and gas pipelines offer several exciting possibilities. It typically requires a conceptual framework as proposed by the author of this paper in Fig. 14, for Adoption of the Big Data Analytics.

Fig. 14 presents a proposed conceptual framework with layers implemented. This framework represents the big data value chain (horizontal axis) and business value chain (vertical axis). In the value chain, the value can be achieved by accessing relevant data streams through data collection, analysis, processing, and visualization which lead to business insights for energy pipeline asset management. In the business value chain dimension, the big data value can be achieved by providing infrastructure, platforms, and relevant application tools. This provides energy pipeline asset management with a layered process to develop a holistic approach to policy development through managing resources, processes, competencies, and technologies. Most importantly, this framework supports managing the most important part and right assets in the first place, by knowing when and how to get rid of them. Therefore, to reap the advantages of asset management, big data analytics can be used to align business direction with the best whole-life cycle through the combination of acquiring, utilizing, maintaining, and renewing assets which turn reorganization into rewards, and assets deliver the best possible value. The process will be used to further manage assets more effectively, and better coordinate and performance over the whole life cycle.

In summary, this systematic literature review serves as a detail and classified source for pipeline integrity management and condition assessment using big data analytics.

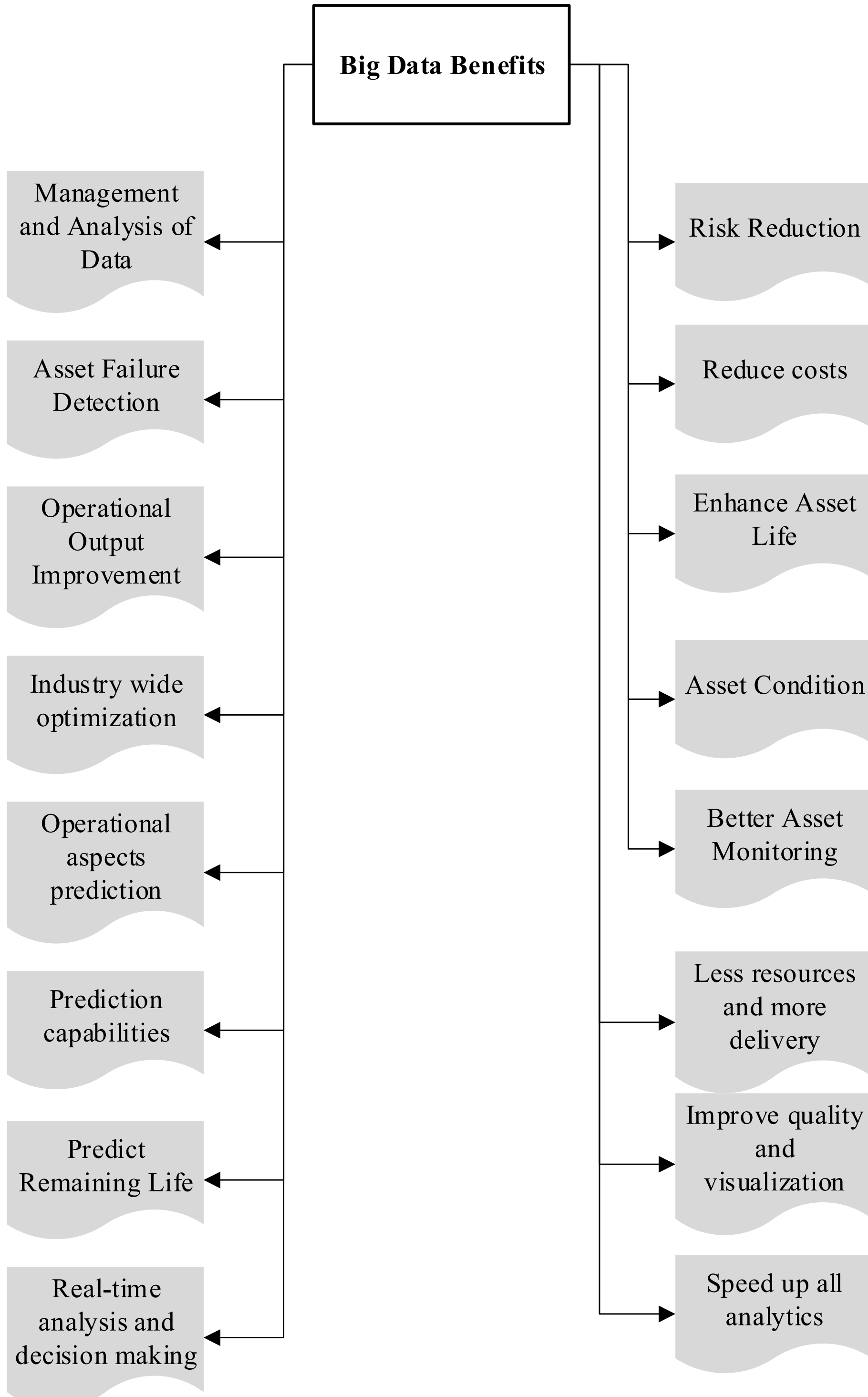

Fig. 13. Industry-wide benefits of big data.

5. Findings and discussion

This systematic literature review presents a detailed analysis of big data analytics in energy pipeline condition assessment. These types of reviews usually come with few drawbacks, however, the results of systematic literature reviews are mainly considered reliable and credible. Technological advancements in the energy sector in recent years have had an enormous impact and also generated a significantly large number of datasets in different sectors of the energy industry including energy pipeline condition assessment. Big data is an analytical tool that needs to be studied from the perspective of the energy sector and help to process these huge datasets and their applications [90]. Big data emerging technology for handling and analysing large datasets. These datasets are reported in a variety of formats and produced in large quantities in various operations of the upstream and downstream energy industries [187–195].

There is a frequent use of big data analytics in other sectors as discussed in previous sections such as electric, automobile, hospital, education, supply chain, etc. The use of the big data analytics is used in renewable energy by predicting the energy generation sources such as

Table 19
Area of focus for future research.

Area of focus	Brief details
Real-time Monitoring and Alert Systems	As per the author best knowledge, there is no research available that discuss real-time monitoring system and alert systems that can quickly detect anomalies, leaks, or security breaches in pipelines. Research can be conducted to develop robust algorithms and techniques for real-time data processing, anomaly detection, and automated alert generation to enable proactive response and prevent incidents.
Advanced Predictive Analytics	Few models have been developed to detect corrosion and other anomalies however there is no real example available where pipeline owners implemented predictive analytics to understand pipeline current condition assessment. Researchers need to get field data and test these models for real site application. Also, most of the models developed one or two parameters, especially corrosion. There is a need to focus on developing more advanced predictive analytics models that can accurately forecast pipeline failures, identify maintenance needs, and optimize operational performance using multiple parameters. This includes exploring advanced machine learning algorithms, incorporating real-time data, and leveraging historical patterns to improve accuracy and reliability.
Integrated Data Analytics Platforms	Looking at the literature, there is no details available on data integration. Developing comprehensive and integrated data analytics platforms specific to the oil and gas pipeline industry can streamline the implementation of big data analytics. Such platforms would combine various data sources, provide data visualization capabilities, and offer a wide range of analytics tools and algorithms tailored to pipeline operations.
Integration with IoT and Sensor Networks	The integration of big data analytics with the IoT and sensor networks presents significant research opportunities. Exploring how sensor data can be effectively collected, processed, and integrated with other data sources can enhance real-time monitoring, condition-based maintenance, and overall pipeline performance
Data Fusion and Integration Techniques	Research can focus on developing advanced data fusion and integration techniques to combine data from multiple sources, including legacy systems, SCADA systems, satellite imagery, and external data providers. This research can lead to more accurate and comprehensive insights for decision-making.

These research areas can contribute to the development of innovative solutions and technologies that enhance the safety, efficiency, and sustainability of oil and gas pipeline operations through the utilization of big data analytics.

wind, solar, and hydropower [196]. Big data analytics can be used far beyond insight and hindsight for future assumptions which help in developing contingency strategies for effective and efficient energy pipeline asset management [197], understanding materials deterioration, and predicting failure in advance to overcome safety and production loss. This further help in various areas of energy sectors such as repair, finance, and life cycle management. Big data analytics can be also important in making advanced decisions through long-term and short-term scheduling of maintenance, staff planning, and allocation strategies [72]. In the operation section of the energy pipeline asset management, big data analytics assist in the risk assessment and develop strategies on the future availability profile for the projects [15]. In addition to lowering maintenance expenses and covering insurance policy-related expenditures, big data analytics can also aid in the financial management of energy pipeline assets. Big data analytics can be also useful and most important in the life-cycle management for energy pipeline asset management by providing covers to the remaining useful life, replacement, retrofit planning, and optimal exploitation [196]. As discussed, earlier pipelines are considered the heart-line of the energy industry, especially within transportation and there is enormous demand for sophisticated and advanced technology and mechanisms to cater to pipeline project implementation, asset management, and operation and maintenance. Big data analytics is designed for handling a large amount of data corresponding to the pipeline which makes it a powerful tool for asset management. The core application of big data analytics is a user-friendly environment that acts as an on-the-fly management of right-of-the-use data and is efficient in the retrieval of design and quality documents [75]. Meaningful insights extraction such as patterns and themes as well as professional preferences, market trends, and unknown correlations make it more important to use in energy

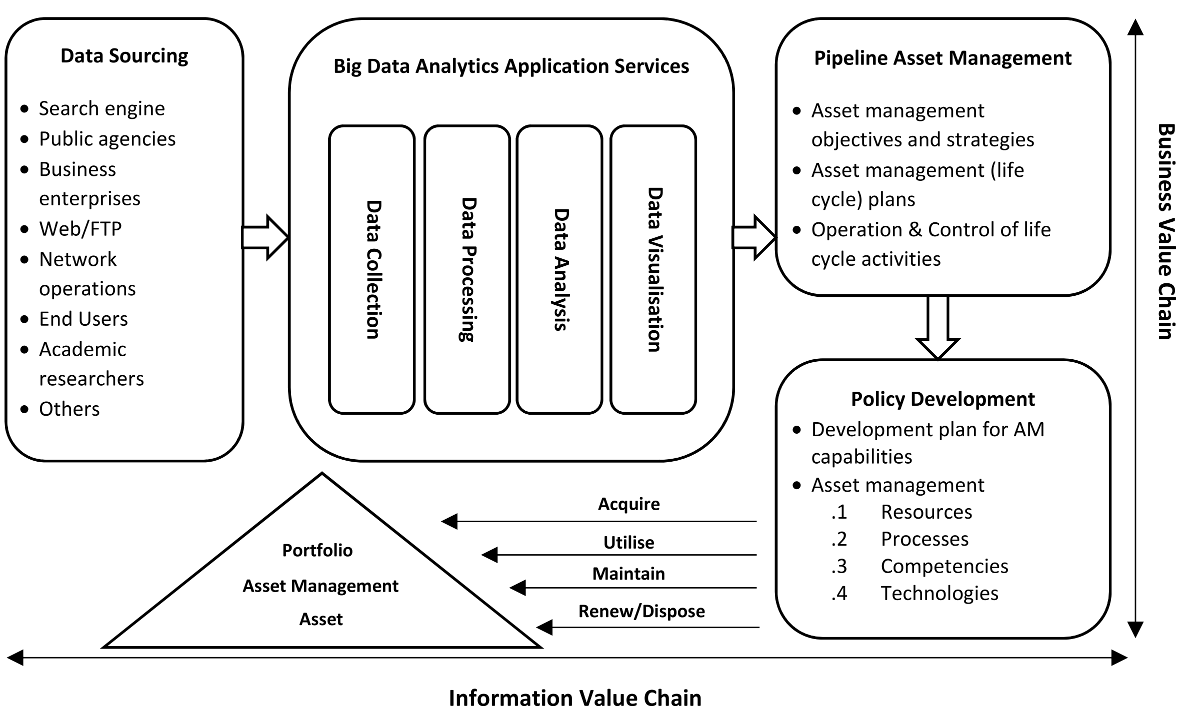

Fig. 14. Conceptual framework for adoption of the big data analytics for energy pipeline asset management.

pipeline asset management. It can be used in making better decision-making processes and preventing flaws in the pipeline asset management [66]. Big data helps in energy industry to predict required maintenance work to avoid any risk of failure and address issues that can improve infrastructure. There are a few drawbacks to the use of big data systems in energy sectors as well such as privacy issues, cybersecurity related challenges, government regulations, and others [198]. However, with effective measures data privacy and cybersecurity related issues can be overcome such as using advanced data encryption solutions [106].

This paper also revealed a few challenges in finding appropriate literature during the search. Although the research question focused on the application of big data analytics for energy pipeline asset management, studies assessing big data analytics within pipeline asset management are still scarce. Most of the reviews assessed the performance, benefits, and drawbacks of big data analytics in energy sectors as large and most of the reviews were assorted with applications in energy sectors. This is the major pitfall identified in this paper which needs to be further investigated through empirical studies. This paper recommended a conceptual framework based on the limited papers for energy pipeline asset management to adopt big data analytics techniques.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

[1] J.R. Mashey, Big data... and the next wave of infrastress, in: Computer Science Division Seminar, University of California, Berkeley, 1997.
[2] S. Bryson, et al., Visually exploring gigabyte data sets in real time, Commun. ACM 42 (8) (1999) 82–90.
[3] Y. Zhang, T. Huang, E.F. Bompard, Big data analytics in smart grids: a review, Energy informatics 1 (1) (2018) 1–24.
[4] P.D. Diamantoulakis, V.M. Kapinas, G.K. Karagiannidis, Big data analytics for dynamic energy management in smart grids, Big Data Research 2 (3) (2015) 94–101.
[5] J. Li, et al., Methods and applications for artificial intelligence, big data, internet-of-things, and blockchain in smart energy management, Energy and AI (2022), 100208.
[6] L. Chen, et al., Research on an oil pipeline anomaly identification method for distinguishing true and false anomalies, Mobile Inf. Syst. (2022) 2022.
[7] D.H. Wang, Y.N. Shi, W.B. Chen, Design of coal mining roof pressure monitoring system based on Labview, in: Applied Mechanics and Materials, Trans Tech Publ, 2014.
[8] J. Sun, et al., Study on nanoscale obstructed flow with Molecular Dynamics Simulation method, Progress in Computational Fluid Dynamics, an International Journal 10 (1) (2010) 51–61.
[9] L. Wang, et al., The processing method of temperature drift data for prediction based on wavelet theory, in: 2015 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), IEEE, 2015.
[10] H. Zhang, S. Zhang, M.M. Hanafiah, Localization and recognition algorithm for fuzzy anomaly data in big data networks, Open Phys. 16 (1) (2018) 1076–1084.
[11] T.R. Wanasinghe, et al., Human centric digital transformation and operator 4.0 for the oil and gas industry, IEEE Access 9 (2021) 113270–113291.
[12] Wikibon, Global Big Data Market Year-On-Year Growth from 2012 to 2027, 2018 [cited 2023 25 June].
[13] J. Van Dijck, Datafication, dataism and dataveillance: big Data between scientific paradigm and ideology, Surveill. Soc. 12 (2) (2014) 197–208.
[14] L. Xu, et al., The Research progress and prospect of data mining methods on corrosion prediction of oil and gas pipelines, Eng. Fail. Anal. (2022), 106951.
[15] M. Mohammadpoor, F. Torabi, Big Data analytics in oil and gas industry: an emerging trend, Petroleum 6 (4) (2020) 321–328.
[16] M. Ockree, et al., Integrating big data analytics into development planning optimization, in: SPE/AAPG Eastern Regional Meeting, OnePetro, 2018.
[17] L. Cadei, et al., Big data advanced analytics to forecast operational upsets in upstream production system, in: Abu Dhabi International Petroleum Exhibition & Conference, OnePetro, 2018.
[18] C.B. Pettinger, Leading indicators, culture and big data: using your data to eliminate death, in: ASSE Professional Development Conference and Exposition, OnePetro, 2014.
[19] M. Tarrahi, A. Shadravan, Advanced big data analytics improves HSE management, in: SPE Bergen One Day Seminar, OnePetro, 2016.
[20] G. George, M.R. Haas, A. Pentland, Big Data and Management, Academy of Management Briarcliff Manor, NY, 2014, pp. 321–326.
[21] E.W. Ngai, et al., The application of data mining techniques in financial fraud detection: a classification framework and an academic review of literature, Decis. Support Syst. 50 (3) (2011) 559–569.
[22] A. Rachman, T. Zhang, R.C. Ratnayake, Applications of machine learning in pipeline integrity management: a state-of-the-art review, Int. J. Pres. Ves. Pip. 193 (2021), 104471.
[23] B.J. Shea, et al., Amstar 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both, bmj (2017) 358.
[24] K.S. Khan, et al., Five steps to conducting a systematic review, Journal of the royal society of medicine 96 (3) (2003) 118–121.
[25] A.M. Drucker, P. Fleming, A.-W. Chan, Research techniques made simple: assessing risk of bias in systematic reviews, J. Invest. Dermatol. 136 (11) (2016) e109–e114.
[26] D. Moher, A. Liberati, J. Tetzlaff, D.G. Altman, T.P. Group, Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement, PLoS Med. 6 (7) (2009), e1000097-6.
[27] G.M. Tawfik, et al., A step by step guide for conducting a systematic review and meta-analysis with simulation data, Trop. Med. Health 47 (1) (2019) 1–9.
[28] K. Seers, Qualitative systematic reviews: their importance for our understanding of research relevant to pain, Br. J. Pain 9 (1) (2015) 36–40.
[29] W. Ahmad, et al., Formal reliability analysis of oil and gas pipelines, Proc. Inst. Mech. Eng. O J. Risk Reliab. 232 (3) (2018) 320–334.
[30] O.K. Aronu, Integrity Management in the Energy Sector-An Investigation of Oil & Gas Assets, NTNU, 2017.
[31] J. Ramasamy, M.Y. Sha'ri, A literature review of subsea asset integrity framework for project execution phase, Procedia Manuf. 4 (2015) 79–88.
[32] M. Hussain, et al., Application of big data analytics to energy pipeline corrosion management, Corrosion Manag. 2021 (2021) 28–29.
[33] R.G. Mora, et al., Pipeline Integrity Management Systems: A Practical Approach, vol. 374, ASME Press, New York, 2016.
[34] H.A. Kishawy, H.A. Gabbar, Review of pipeline integrity management practices, Int. J. Pres. Ves. Pip. 87 (7) (2010) 373–380.
[35] T. Jiang, et al., Application of FBG based sensor in pipeline safety monitoring, Appl. Sci. 7 (6) (2017) 540.
[36] R. Palmer-Jones, S. Turner, P. Hopkins, A new approach to risk based pipeline integrity management, in: International Pipeline Conference, 2006.
[37] M. Xie, Z. Tian, A review on pipeline integrity management utilizing in-line inspection data, Eng. Fail. Anal. 92 (2018) 222–239.
[38] M.N.b.M.A. Napiah, Total pipeline integrity management system–PETRONAS gas berhad's practices, in: 3rd Pipeline Technology Conference 2008, 2008.
[39] U. Sivarajah, et al., Critical analysis of Big Data challenges and analytical methods, J. Bus. Res. 70 (2017) 263–286.
[40] H. Lu, et al., Oil and Gas 4.0 era: a systematic review and outlook, Comput. Ind. 111 (2019) 68–90.
[41] A. Subramaniyan, Putting Analytics to Work for You: Data and Models, 2017.
[42] P. McMahon, T. Zhang, R. Dwight, Requirements for big data adoption for railway asset management, IEEE Access 8 (2020) 15543–15564.
[43] E. Baccarelli, et al., Fog of everything: energy-efficient networked computing architectures, research challenges, and a case study, IEEE Access 5 (2017) 9882–9910.
[44] H. Wang, et al., Towards felicitous decision making: an overview on challenges and trends of Big Data, Inf. Sci. 367 (2016) 747–765.
[45] M. Brulé, The data reservoir: how big data technologies advance data management and analytics in E&P, in: SPE Digital Energy Conference and Exhibition, OnePetro, 2015.
[46] S. Brandt, S. Brandt, Data Analysis, Springer, 1998.
[47] P. Zikopoulos, C. Eaton, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, McGraw-Hill Osborne Media, 2011.
[48] H.E. Pence, What is big data and why is it important? J. Educ. Technol. Syst. 43 (2) (2014) 159–171.
[49] J. Anuradha, A brief introduction on Big Data 5Vs characteristics and Hadoop technology, Procedia Comput. Sci. 48 (2015) 319–324.
[50] M.R. Ghazi, D. Gangodkar, Hadoop, MapReduce and HDFS: a developers perspective, Proc. Comput. Sci. 48 (2015) 45–50.
[51] M.R. Trifu, M.L. Ivan, Big Data: present and future, Database Systems Journal 5 (1) (2014) 32–41.
[52] C. Győrödi, et al., A Comparative Study: MongoDB vs. MySQL. In 2015 13th International Conference on Engineering of Modern Electric Systems (EMES), IEEE, 2015.
[53] N. Mounir, et al., Integrating Big Data: simulation, predictive analytics, real time monitoring, and data warehousing in a single cloud application, in: Offshore Technology Conference, OnePetro, 2018.
[54] P. Warden, Big Data Glossary, O'Reilly Media, Inc, 2011.
[55] T. Kudo, et al., A proposal of transaction processing method for MongoDB, Proc. Comput. Sci. 96 (2016) 801–810.
[56] V. Manoj, Comparative study of nosql document, column store databases and evaluation of cassandra, Int. J. Database Manag. Syst. 6 (4) (2014) 11.
[57] A. Spark, Cluster Mode Overview, 2022.

[58] N. Ahmed, et al., A comprehensive performance analysis of Apache Hadoop and Apache Spark for large scale data sets using HiBench, Journal of Big Data 7 (1) (2020) 1–18.
[59] R. Hicham, B.M. Anis, Processes meet big data: scaling process discovery algorithms in big data environment, J. King Saud Univ.-Comput. Inf. Sci., 34 (10, Part A) (2022) 8478–8489.
[60] J. Gohil, M. Shah, Application of Big Data in Petroleum Streams, CRC Press, 2022.
[61] P. Anand, Big Data is a big deal, J. Petrol. Technol. 65 (4) (2013) 18–21.
[62] H. Hamzeh, Application of Big Data in Petroleum Industry, Department of Electronics and Computer Engineering Istanbul Sehir University hamedhamzeh@ std. sehir. edu. tr, 2016.
[63] G. He, et al., A framework of smart pipeline system and its application on multiproduct pipeline leakage handling, Energy 188 (2019), 116031.
[64] Teisman, G. and F. Boons, International Public Management and Policy Shivant Jhagroe-280055..
[65] S. Morrow, M. Coplen, Safety Culture: a Significant Influence on Safety in Transportation, United States. Federal Railroad Administration. Office of Research, 2017.
[66] H. Zhang, et al., A modified method for the safety factor parameter: the use of big data to improve petroleum pipeline reliability assessment, Reliab. Eng. Syst. Saf. 198 (2020), 106892.
[67] M. Mohammadpoor, F. Torabi, Big Data analytics in oil and gas industry: an emerging trend, Petroleum 5 (4) (2018) 321–328.
[68] S. Hasan, S.M. Shamsuddin, N. Lopes, Machine learning big data framework and analytics for big data problems, Int. J. Advance Soft Compu. Appl 6 (2) (2014).
[69] R.C. Joseph, N.A. Johnson, Big data and transformational government, It Professional 15 (6) (2013) 43–48.
[70] M.H. ur Rehman, et al., Big data reduction framework for value creation in sustainable enterprises, Int. J. Inf. Manag. 36 (6) (2016) 917–928.
[71] M.A. Waller, S.E. Fawcett, Data Science, Predictive Analytics, and Big Data: a Revolution that Will Transform Supply Chain Design and Management, Wiley Online Library, 2013, pp. 77–84.
[72] T. Nguyen, R.G. Gosine, P. Warrian, A systematic review of big data analytics for oil and gas industry 4.0, IEEE Access 8 (2020) 61183–61201.
[73] T.A. Tu, The relationship between big data and IoT, Journal of Computing and Electronic Information Management 10 (3) (2023) 150–154.
[74] W. Kong, et al., Model building and simulation for intelligent early warning of long-distance oil & gas storage and transportation pipelines based on the probabilistic neural network, in: IOP Conference Series: Earth and Environmental Science, IOP Publishing, 2020.
[75] H. Hassani, E.S. Silva, Big Data: a big opportunity for the petroleum and petrochemical industry, OPEC Energy Review 42 (1) (2018) 74–89.
[76] M. Nazmul Alam, et al., Big Data: An Overview with Legal Aspects and Future Prospects 10 (2023) 476–485.
[77] B. Singh, et al., A survey on big data: challenges, tools and technique, Int. J. Adv. Res. Comput. Sci. 7 (6) (2016).
[78] Shah, V., et al., Big data analytics in oil and gas industry, in Emerging Technologies for Sustainable and Smart Energy. CRC Press Boca Raton and London. p. 37-55..
[79] M.H. Jensen, P.A. Nielsen, J.S. Persson, From big data technologies to big data benefits, Computer 56 (6) (2023) 52–61.
[80] K.R. Holdaway, Harness Oil and Gas Big Data with Analytics: Optimize Exploration and Production with Data-Driven Models, John Wiley & Sons, 2014.
[81] S.A. Muhammad Shaheed Abdullahi, A. Isa, Development of hybrid software architectural model for big data analytics in the oil and gas industry, in: 4th YUMSCIAIC, July 2019.
[82] M.S. El-Abbasy, et al., A condition assessment model for oil and gas pipelines using integrated simulation and analytic network process, Structure and Infrastructure Engineering 11 (3) (2015) 263–281.
[83] W.R. Byrd, R.G. McCoy, D.D. Wint, A success guide for pipeline integrity management, Pipeline Gas J. 231 (2004) 249–254.
[84] J. Gu, et al., The application of the big data algorithm for pipeline lifetime analysis, in: 2019 Chinese Automation Congress (CAC), IEEE, 2019.
[85] C. Cobanoglu, et al., A systematic review of big data: research approaches and future prospects, Journal of Smart Tourism 2 (1) (2022) 21–31.
[86] A. Swetapadma, A. Yadav, Data-mining-based fault during power swing identification in power transmission system, IET Sci. Meas. Technol. 10 (2) (2016) 130–139.
[87] M. Layouni, M.S. Hamdi, S. Tahar, Detection and sizing of metal-loss defects in oil and gas pipelines using pattern-adapted wavelets and machine learning, Appl. Soft Comput. 52 (2017) 247–261.
[88] Z. Lv, et al., Next-generation big data analytics: state of the art, challenges, and future research topics, IEEE Trans. Ind. Inf. 13 (4) (2017) 1891–1899.
[89] N. Rawat, Big Data analytics in oil and gas industry, Int. J. Sci. Eng. Res. (2014) 2601–2611.
[90] R. Skrynkovskyy, et al., Big Data Approach Application for Steel Pipelines in the Conditions of Corrosion Fatigue, 2018.
[91] O. Elijah, et al., A survey on industry 4.0 for the oil and gas industry: upstream sector, IEEE Access 9 (2021) 144438–144468.
[92] J. Eze, C. Nwagboso, P. Georgakis, Framework for integrated oil pipeline monitoring and incident mitigation systems, Robot. Comput. Integrated Manuf. 47 (2017) 44–52.
[93] M.A. Adegboye, W.-K. Fung, A. Karnik, Recent advances in pipeline monitoring and oil leakage detection technologies: principles and approaches, Sensors 19 (11) (2019) 2548.
[94] N.G. Franconi, et al., Wireless communication in oil and gas wells, Energy Technol. 2 (12) (2014) 996–1005.
[95] J. Wan, et al., Hierarchical leak detection and localization method in natural gas pipeline monitoring sensor networks, Sensors 12 (1) (2011) 189–214.
[96] H. Wu, et al., One-dimensional CNN-based intelligent recognition of vibrations in pipeline monitoring with DAS, J. Lightwave Technol. 37 (17) (2019) 4359–4366.
[97] P. Yi, X. Lizhi, Z. Yuanzhong, Remote real-time monitoring system for oil and gas well based on wireless sensor networks, in: 2010 International Conference on Mechanic Automation and Control Engineering, IEEE, 2010.
[98] R. Barani, V.J. Lakshmi, Oil well monitoring and control based on wireless sensor networks using Atmega 2560 controller, International Journal of Computer Science & Communication Networks 3 (6) (2013) 341.
[99] Y. Seo, et al., Development of AI-based diagnostic model for the prediction of hydrate in gas pipeline, Energies 14 (8) (2021) 2313.
[100] M.E.A.B. Seghier, et al., Advanced intelligence frameworks for predicting maximum pitting corrosion depth in oil and gas pipelines, Process Saf. Environ. Protect. 147 (2021) 818–833.
[101] H. Liu, et al., Matching pipeline in-line inspection data for corrosion characterization, NDT E Int. 101 (2019) 44–52.
[102] M. Singh, S. Hetlevik, Data–information–knowledge hierarchy based decision support system for risk based inspection analysis, International Journal of System Assurance Engineering and Management 8 (2) (2017) 1588–1595.
[103] N.A. Akram, et al., Active incremental Support Vector Machine for oil and gas pipeline defects prediction system using long range ultrasonic transducers, Ultrasonics 54 (6) (2014) 1534–1544.
[104] M. Chaari, et al., An integrated genetic-algorithm/artificial-neural-network approach for steady-state modeling of two-phase pressure drop in pipes, SPE Prod. Oper. 35 (3) (2020) 628–640.
[105] P. Zahedi, et al., Random forest regression prediction of solid particle Erosion in elbows, Powder Technol. 338 (2018) 983–992.
[106] M.-w. Fan, C.-c. Ao, X.-r. Wang, Comprehensive method of natural gas pipeline efficiency evaluation based on energy and big data analysis, Energy 188 (2019), 116069.
[107] B. Rouhanizadeh, S. Kermanshachi, Predictive model development to perform condition assessment on pipeline networks, in: Computing in Civil Engineering 2019: Smart Cities, Sustainability, and Resilience, American Society of Civil Engineers Reston, VA, 2019, pp. 24–32.
[108] M. Hussain, T. Zhang, M.S. Naseer, Impact of Covid-19 and Needs of Digital Transformation to Protect Assets from Corrosion, 2022.
[109] J. Feblowitz, Analytics in oil and gas: the big deal about big data, in: SPE Digital Energy Conference, OnePetro, 2013.
[110] R. Beckwith, Managing big data: cloud computing and co-location centers, J. Petrol. Technol. 63 (10) (2011) 42–45.
[111] N. Mounir, et al., Integrating Big Data: simulation, predictive analytics, real time monitoring, and data warehousing in a single cloud application, in: Offshore Technology Conference, OTC, 2018.
[112] G. Phillips-Wren, A. Hoskisson, An analytical journey towards big data, J. Decis. Syst. 24 (1) (2015) 87–102.
[113] G.-H. Kim, S. Trimi, J.-H. Chung, Big-data applications in the government sector, Commun. ACM 57 (3) (2014) 78–85.
[114] S. Priyadarshy, IoT Revolution in Oil and Gas Industry. Internet of Things and Data Analytics Handbook, 2017, pp. 513–520.
[115] M.S. Sumbal, et al., Value creation through big data application process management: the case of the oil and gas industry, J. Knowl. Manag. 23 (8) (2019) 1566–1585.
[116] A. Alhosani, S. Mohamed Zabri, A uniform supply chain management framework for oil and gas sector: a preliminary review, International Journal of Advanced and Applied Sciences 5 (2) (2018) 19–24.
[117] J. Thiyagalingam, et al., Scientific machine learning benchmarks, Nature Reviews Physics 4 (6) (2022) 413–420.
[118] S. Kaisler, et al., Big Data and Analytics: Issues and Challenges for the Past and Next Ten Years, 2023.
[119] E. Barbierato, M. Gribaudo, M. Iacono, Performance evaluation of NoSQL big-data applications using multi-formalism models, Future Generat. Comput. Syst. 37 (2014) 345–353.
[120] E. Al Nuaimi, et al., Applications of big data to smart cities, Journal of Internet Services and Applications 6 (1) (2015) 1–15.
[121] E. Maidla, et al., Drilling analysis using big data has been misused and abused, in: IADC/SPE Drilling Conference and Exhibition, OnePetro, 2018.
[122] H. Jiang, et al., Scaling up MapReduce-based big data processing on multi-GPU systems, Cluster Comput. 18 (2015) 369–383.
[123] J. House, Big data analytics= Key to successful 2015, Supply Chain Strategy (2014).
[124] A. Preveral, A. Trihoreau, N. Petit, Geographically-distributed databases: a big data technology for production analysis in the oil & gas industry, in: SPE Intelligent Energy International Conference and Exhibition, SPE, 2014.
[125] R. Aliguliyev, Y. Imamverdiyev, F. Abdullayeva, The investigation of opportunities of big data analytics as analytics-as-a-service in cloud computing for oil and gas industry, Problems of information technology 7 (1) (2016) 9–22.
[126] D.P. Acharjya, K. Ahmed, A survey on big data analytics: challenges, open research issues and tools, Int. J. Adv. Comput. Sci. Appl. 7 (2) (2016) 511–518.
[127] J. Schmidt, Plant security-public awareness and mitigation of third party attacks as a new layer of protection in the safety concept, CET Journal-Chemical Engineering Transactions 77 (2019).
[128] S. Sun, et al., Data handling in industry 4.0: interoperability based on distributed ledger technology, Sensors 20 (11) (2020) 3046.

[129] A. Gandomi, M. Haider, Beyond the hype: big data concepts, methods, and analytics, Int. J. Inf. Manag. 35 (2) (2015) 137–144.
[130] M. Igamberdiev, et al., An integrated multi-level modeling approach for industrial-scale data interoperability, Software Syst. Model 17 (2018) 269–294.
[131] M. Muhammed, W. Obidallah, R. Bijan, Applying deep learning techniques for big data analytics: a systematic literature review, Arch Inf Sci Tech 1 (1) (2018) 20–41.
[132] R. Toshniwal, K.G. Dastidar, A. Nath, Big data security issues and challenges, International Journal of Innovative Research in Advanced Engineering (IJIRAE) 2 (2) (2015).
[133] A. Divyakant, B. Philip, B. Elisa, Challenges and opportunities with big data, Proc. VLDB Endowment 5 (12) (2012) 2032–2033.
[134] A. Oguntimilehin, E.-O. Ademola, A review of big data management, benefits and challenges, A Review of Big Data Management, Benefits and Challenges 5 (6) (2014) 1–7.
[135] R. Du Mars, Mission Impossible? Data Governance Process Takes on "Big Data, 2012.
[136] T. Shah, F. Rabhi, P. Ray, Investigating an ontology-based approach for Big Data analysis of inter-dependent medical and oral health conditions, Cluster Comput. 18 (2015) 351–367.
[137] J. Chen, et al., Big data challenge: a data management perspective, Front. Comput. Sci. 7 (2013) 157–164.
[138] M.F. Uddin, N. Gupta, Seven V's of big data understanding big data to extract value, in: Proceedings of the 2014 Zone 1 Conference of the American Society for Engineering Education, IEEE, 2014.
[139] A.A. Tole, Big data challenges, Database systems journal 4 (3) (2013).
[140] N. Khan, et al., Big data: survey, technologies, opportunities, and challenges, Sci. World J. 2014 (2014).
[141] A. Mehta, Tapping the value from big data analytics, J. Petrol. Technol. 68 (12) (2016) 40–41.
[142] D. Che, M. Safran, Z. Peng, From big data to big data mining: challenges, issues, and opportunities, in: Database Systems for Advanced Applications: 18th International Conference, DASFAA 2013, International Workshops: BDMA, SNSM, SeCoP, Wuhan, China, April 22-25, 2013. Proceedings 18, Springer, 2013.
[143] Y. Gidh, et al., WITSML v2. 0: paving the way for big data analytics through improved data assurance and data organization, in: SPE Intelligent Energy International Conference and Exhibition, OnePetro, 2016.
[144] P. Neri, Big data in the digital oilfield requires data transfer standards to perform, in: Offshore Technology Conference, OTC, 2018.
[145] O. Kwon, N. Lee, B. Shin, Data quality management, data usage experience and acquisition intention of big data analytics, Int. J. Inf. Manag. 34 (3) (2014) 387–394.
[146] K. Al-Barznji, A. Atanassov, A survey of Big Data Mining: challenges and techniques, in: Proceedings of 24th International Symposium" Control of Energy, Industrial and Ecological Systems, 2016. Bankia, Bulgaria.
[147] A. Cuzzocrea, Big data lakes: models, frameworks, and techniques, in: 2021 IEEE International Conference on Big Data and Smart Computing (BigComp), IEEE, 2021.
[148] J. Merino, et al., A data quality in use model for big data, Future Generat. Comput. Syst. 63 (2016) 123–130.
[149] D. Loshin, Big Data Analytics: from Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph, Elsevier, 2013.
[150] M. Hussain, T. Zhang, M.S. Nasser, The importance of data quality in energy pipelines condition assessment, in: 11th IMA International Conference on Modelling in Industrial Maintenance and Reliability (MIMAR), Institute of Mathematics & its Applications, Online, 2021.
[151] R.Y. Wang, D.M. Strong, Beyond accuracy: what data quality means to data consumers, J. Manag. Inf. Syst. 12 (4) (1996) 5–33.
[152] N. Abdullah, et al., Data quality in big data: a review, Int. J. Advance Soft Compu. Appl 7 (3) (2015) 17–27.
[153] I. Lee, Big data: dimensions, evolution, impacts, and challenges, Bus. Horiz. 60 (3) (2017) 293–303.
[154] R.K. Mobley, An Introduction to Predictive Maintenance, Elsevier, 2002.
[155] G.A. Gow, Privacy and ubiquitous network societies: background paper, in: Proceedings of the ITU Workshop on Ubiquitous Network Societies, International Telecommunication Union, 2005. *Retrieved August*.
[156] S. Kaisler, et al., Big data: issues and challenges moving forward, in: 2013 46th Hawaii International Conference on System Sciences, IEEE, 2013.
[157] X. Yi, et al., Building a network highway for big data: architecture and challenges, Ieee Network 28 (4) (2014) 5–13.
[158] R. Krishnamurthy, K.C. Desouza, Big data analytics: the case of the social security administration, Inf. Polity 19 (3–4) (2014) 165–178.
[159] J. Moreno, M.A. Serrano, E. Fernández-Medina, Main issues in big data security, Future Internet 8 (3) (2016) 44.
[160] S. Phuyal, D. Bista, R. Bista, Challenges, opportunities and future directions of smart manufacturing: a state of art review, Sustainable Futures 2 (2020), 100023.
[161] R.T. Kouzes, et al., The changing paradigm of data-intensive computing, Computer 42 (1) (2009) 26–34.
[162] D. Agrawal, et al., Challenges and Opportunities with Big Data 2011-1, 2011.
[163] P. Barnaghi, A. Sheth, C. Henson, From data to actionable knowledge: big data challenges in the web of things [Guest Editors' Introduction], IEEE Intell. Syst. 28 (6) (2013) 6–11.
[164] S. Wang, et al., Implementing smart factory of industrie 4.0: an outlook, Int. J. Distributed Sens. Netw. 12 (1) (2016), 3159805.
[165] D. Cameron, Big data in exploration and production: Silicon snake-oil, magic bullet, or useful tool?, in: SPE Intelligent Energy International Conference and Exhibition SPE, 2014.
[166] H. Chen, R.H. Chiang, V.C. Storey, Business intelligence and analytics: from big data to big impact, MIS Q. (2012) 1165–1188.
[167] A. Graves, Techniques to reduce cluttering of rdf visualizations, Future Generat. Comput. Syst. 53 (2015) 152–156.
[168] C.P. Chen, C.-Y. Zhang, Data-intensive applications, challenges, techniques and technologies: a survey on Big Data, Inf. Sci. 275 (2014) 314–347.
[169] J. Taheri, et al., Pareto frontier for job execution and data transfer time in hybrid clouds, Future Generat. Comput. Syst. 37 (2014) 321–334.
[170] K. Wadhwani, D.Y. Wang, Big Data Challenges and Solutions, Technical Report, 2017, https://doi.org/10.13140/RG, 2.2. 16548.88961.
[171] B. Mathew, How Big Data Is Reducing Costs and Improving Performance in the Upstream Industry, World Oil, 2016.
[172] A. Føllesdal Tjønn, Digital twin through the life of a field, in: Abu Dhabi International Petroleum Exhibition & Conference, OnePetro, 2018.
[173] T.R. Wanasinghe, et al., Digital twin for the oil and gas industry: overview, research trends, opportunities, and challenges, IEEE Access 8 (2020) 104175–104197.
[174] D. Zaidi, Role of Data Analytics in the Oil Industry, 2018.
[175] A. Endress, Big Data Analytics Burst onto Upstream Industry as Next Frontier, 2017.
[176] L. Dodgson, Cutting Costs: How Big Data Can Help, 2016.
[177] K. Boman, Big Data, Internet of Things Transforming Oil and Gas Operations, 2015.
[178] N. Dhunay, Big Data's Next Big Impact, 2016. *Oil & Gas. Available via:*, http://dataconomy.%20com/big-datas-next-big-impact-oil-gas/ [Accessed January 2023].
[179] S. Nigeria, Shell saves $1 million using IoT, Internet Bus.: IoT, AI, Data and Edge Computing in the Connected World 2016 (2016).
[180] Q. Yin, et al., Improve the drilling operations efficiency by the big data mining of Real-Time logging, in: SPE/IADC Middle East Drilling Technology Conference and Exhibition, OnePetro, 2018.
[181] A. Farris, How Big Data Is Changing the Oil & Gas Industry, Analytics Mag., Dec, 2012.
[182] Brancaccio, E., Big Data in Oil and Gas Industry. Oil and Gas Portal..
[183] S. Gopinath, L. Hampton, Big Oil Turns to Big Data to Save Big Money on Drilling, 2017.
[184] G. Dixit, Internet of things in the oil and gas industry, J. Petrol. Technol. (2017). https://jpt.spe.org/internet-things-oil-and-gas-industry.
[185] A. Bakker, How to Benefit from Big Data Analytics in the Oil and Gas Industry?, 2020.
[186] B. Gillette, Big Data Analytics Protects O&G Midstream Compressors, 2017.
[187] M. Akoum, A. Mahjoub, A unified framework for implementing business intelligence, real-time operational intelligence and big data analytics for upstream oil industry operators, in: SPE Middle East Intelligent Energy Conference and Exhibition, OnePetro, 2013.
[188] A. Bin Mahfoodh, et al., Introducing a big data system for maintaining well data quality and integrity in a world of heterogeneous environment, in: SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition, Society of Petroleum Engineers, 2017.
[189] B. Crockett, K. Kurrey, Smart decision making needs automated analysis" making sense out of big data in real-time, in: SPE Intelligent Energy Conference & Exhibition, Society of Petroleum Engineers, 2014.
[190] K. Hilgefort, Big data analysis using bayesian network modeling: a case study with WG-ICDA of a gas storage field, in: NACE International Corrosion Conference Proceedings, NACE International, 2018.
[191] R.K. Perrons, J. Jensen, The unfinished revolution: what is missing from the E&P industry's move to" big data, J. Petrol. Technol. 66 (5) (2014) 20–22.
[192] R.K. Perrons, J.W. Jensen, Data as an asset: what the upstream oil & gas industry can learn about" big data" from companies like social media, in: SPE Annual Technical Conference and Exhibition, Society of Petroleum Engineers, 2014.
[193] C. Sousa, et al., Applying big data analytics to logistics processes of oil and gas exploration and production through a hybrid modeling and simulation approach, in: Offshore Technology Conference, OTC Brasil, 2015.
[194] A. Sukapradja, et al., Sisi nubi dashboard: implementation of business intelligence in reservoir modelling & synthesis: managing big data and streamline the decision making process, in: SPE/IATMI Asia Pacific Oil & Gas Conference and Exhibition, Society of Petroleum Engineers, 2017.
[195] W. Wu, et al., Retrieving information and discovering knowledge from unstructured data using big data mining technique: heavy oil fields example, in: International Petroleum Technology Conference, 2014. International Petroleum Technology Conference.
[196] A. Mohamed, M.S. Hamdi, S. Tahar, A machine learning approach for big data in oil and gas pipelines, in: 2015 3rd International Conference on Future Internet of Things and Cloud, IEEE, 2015.
[197] J. Chen, et al., Big data challenge: a data management perspective, Front. Comput. Sci. 7 (2) (2013) 157–164.
[198] H. Patel, et al., Transforming petroleum downstream sector through big data: a holistic review, J. Pet. Explor. Prod. Technol. 10 (6) (2020) 2601–2611.