

Review

A Survey of Key Technologies for Multi-source Heterogeneous Data in Intelligent Manufacturing

Minghao Zhu¹, Pengfei Yang², Bo Gao³, Xuehan Li⁴, Letian Wang^{1*}

¹ School of Economics and Management, Beijing Jiaotong University, Beijing 100044, China

² Jiangsu Yunmanman Information Technology Co., Ltd., Nanjing 210012, China

³ School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China

⁴ School of Automation and Intelligence, Beijing Jiaotong University, Beijing 100044, China

* Corresponding author email: letianw@bjtu.edu.cn

Abstract: Currently, most enterprises have adopted information software and digital equipment and gradually established digital factories. They conduct enterprise data collection and decision-support activities, generating large volumes of multi-source heterogeneous data across all stages of the product life cycle. However, current data utilization methods remain simplistic, and the goal of leveraging multi-source heterogeneous data to drive manufacturing value has yet to be fully realized. To address this issue, this study first defines the concept and characteristics of multi-source heterogeneous data in intelligent manufacturing, based on an analysis of its relationship with industrial big data. Then, integrating principles from data science, a technological framework for multi-source heterogeneous data is proposed. The key technologies involved in each stage of data processing are investigated, and typical applications of such data in intelligent manufacturing are discussed. Finally, this paper analyzes the challenges and future development directions of multi-source heterogeneous data processing in intelligent manufacturing. The goal is to provide theoretical and technical support for integrating intelligent manufacturing with data science.

Keywords: intelligent manufacturing; multi-source heterogeneous data; feature fusion; data system; technological framework



Copyright: © 2026 by the authors.
This article is licensed under a
Creative Commons Attribution 4.0
International License (CC BY) license
(<https://creativecommons.org/licenses/by/4.0/>).

Citation: Minghao Zhu, Pengfei Yang, Bo Gao, Xuehan Li, Letian Wang. "A Survey of Key Technologies for Multi-source Heterogeneous Data in Intelligent Manufacturing." *Instrumentation* 13, no.1 (March 2026). <https://doi.org/10.15878/j.instr.202600306>

1 Introduction

With the continuous advancement of Industry 4.0, the manufacturing industry is experiencing a critical window for transformation and upgrading. Emerging information technologies—such as artificial intelligence, big data, the Internet of Things (IoT), and blockchain—are rapidly evolving, offering strong technical support for the digitalization, networking, and intelligentization of the manufacturing sector. Intelligent manufacturing is a representative paradigm of data-driven value enhancement in production^[1,2], where data has become the core element and a key enabler of manufacturing

intelligence^[3].

Currently, manufacturing enterprises widely adopt information technologies, generating large-scale and complex data. For instance, a typical textile manufacturing enterprise produces 84 GB of data per workshop per day, including image data, textual data, sensor readings, and data from information management systems. These datasets originate from diverse sources, exhibit varied structures, and differ significantly in quality. Therefore, focusing on multi-source heterogeneous data in intelligent manufacturing and realizing its deep integration is essential to supporting full-lifecycle and end-to-end resource scheduling and

decision-making in smart factories—an inevitable path for the industry's development^[4].

Notable contributions have been made to multi-source heterogeneous data fusion in intelligent manufacturing. Traditional statistical methods, such as Bayesian networks and Kalman filtering, were applied to structured sensor data including temperature and pressure readings. These methods enabled real-time production monitoring, basic fault diagnosis, and process optimization in discrete manufacturing scenarios^[5,6]. Machine learning-based approaches, particularly deep learning models such as convolutional neural networks and recurrent neural networks, were employed to fuse unstructured data such as product surface images and maintenance text logs with structured data. Higher accuracy was achieved in tasks of defect detection and predictive maintenance^[7,8]. Hybrid fusion frameworks were developed by combining rule-based reasoning with data-driven models. Partial heterogeneity issues in supply chain data integration were addressed, and semi-automated inventory management and logistics scheduling were supported^[9].

As the integration of manufacturing with next-generation information technologies deepens—coupled with the emergence of concepts like the Industrial Internet and cloud manufacturing—the types of data involved in intelligent manufacturing have expanded far beyond sensor signals. These data show strong interdependencies and increasing complexity in both types and sources. It is thus crucial to comprehensively address the multi-source heterogeneous data across the entire life cycle and manufacturing process. Through effective data fusion and mining, significant improvements in production efficiency and product quality can be achieved.

However, distinct limitations have also been observed. Statistical methods were unable to process high-dimensional unstructured data such as high-resolution production line videos. Non-linear correlations between data from different manufacturing links, such as design Bill of Materials (BOM) data and workshop production data, could not be captured. Deep learning models, although powerful in feature extraction, required large-scale labeled datasets. This requirement became an obstacle in scenarios where labeled data, such as rare equipment failure cases, was scarce. The "black-box" nature of these models reduced interpretability and made industrial decision-making traceability difficult^[10]. Hybrid frameworks lacked unified standards for data preprocessing and feature alignment. As a result, scalability was poor when new data sources, such as real-time data from edge devices in smart workshops, needed to be integrated.

Redundancy and high correlation among diverse data features further exacerbate the difficulty of integration, becoming a bottleneck that limits the progress of intelligent manufacturing. Hence, there is an

urgent need to establish a comprehensive theoretical framework for multi-source heterogeneous data in intelligent manufacturing. This includes clarifying its definition and scope, identifying key technologies for data processing and mining, addressing the standardization of complex heterogeneous data in industrial settings, and enhancing feature extraction and fusion—ultimately enabling value creation from manufacturing data.

2 Literature Review

In this paper, we take the Web of Science Core Collection and CNKI (China National Knowledge Infrastructure) database as examples. We conduct a statistical analysis of literature related to multi-source heterogeneous data in the manufacturing industry from 2018 to 2024. CiteSpace (version 6.3.R1)^[11] is used to generate a knowledge map. This helps to analyze the current research status of multi-source heterogeneous data in both domestic and international contexts.

In the Web of Science Core Collection, we search journal articles, conference papers, and review articles. The search formula is: ((TS = "Multisource heterogeneous data") OR TS = "multi sensor data") AND TS = "manufacturing") AND DOP = (2018-01-01 to 2024-01-01). This retrieves literature where the subject includes "multisource heterogeneous data" or "multi sensor data" related to manufacturing between 2018 and 2024.

In the CNKI database, we search for literature on "manufacturing" with the theme containing "multi-source heterogeneous data" or "multi sensor data" from 2018 to 2024. A total of 506 papers are retrieved. Another search for "multi-source heterogeneous data" related to "manufacturing" in the topic field yields 1108 papers.

From the statistical analysis of recent English literature, the research mainly focuses on data fusion, industrial robotics, and rotating machinery maintenance, as shown in Fig. 1. Most studies target multi-sensor data. For example, multi-sensor fusion is used for equipment health monitoring^[12] and to improve machinery fault diagnosis accuracy^[13]. These studies have achieved promising results. Sensor data represent the information collected at the perception layer in intelligent manufacturing and inherently display multi-source heterogeneity.

In the realm of intelligent manufacturing, the application of multi-source heterogeneous data has also been explored in broader contexts. A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems was proposed by Lee et al. The seamless flow and integration of data across different manufacturing elements were emphasized, which is closely related to the handling of multi-source heterogeneous data in intelligent manufacturing scenarios^[14]. Cloud manufacturing was introduced by Zhang, Lin, Luo, and Tao as a new manufacturing

paradigm. In this paradigm, data from cloud platforms, manufacturing resources, and customer demands are required to be integrated and processed. The significance of multi-source heterogeneous data fusion is further highlighted^[15].

Recently, large language models (LLMs) have also been introduced into multi-source heterogeneous data research. Semantic features can be extracted from unstructured text and log data by LLMs, and then fused with structured sensor data to enhance cross-modal analysis in intelligent manufacturing^[16,17]. Such approaches have shown potential in digital twin integration and semantic-level fusion for intelligent decision-making.

According to the statistical analysis of recent Chinese literature, research themes include multi-source heterogeneity, multi-source data^[18], metadata, heterogeneous data, machine learning, data sharing, IoT, deep learning, and information fusion, as shown in Fig.2. The research focus has shifted over time. It evolved from

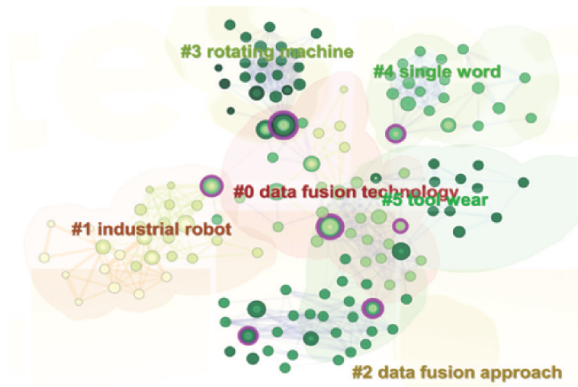


Fig.1 Co-occurrence map of English research distribution directions for multi-source heterogeneous data

algorithm-driven fields like deep learning and IoT to areas such as data governance (e. g., metadata and sharing). It is now gradually moving toward issues like efficient fusion, cross-source collaboration, and consistency assurance, as shown in Fig.3.

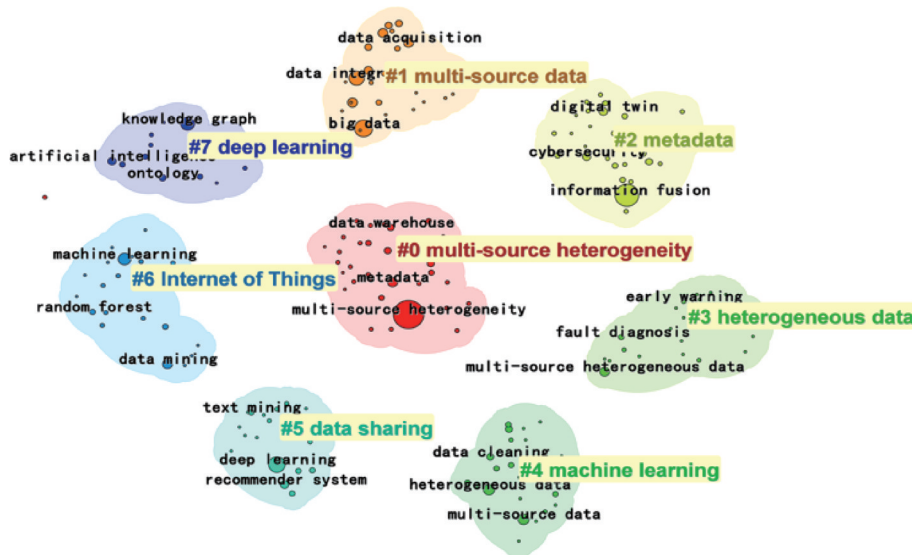


Fig.2 Co-occurrence map of Chinese research distribution directions for multi-source heterogeneous data

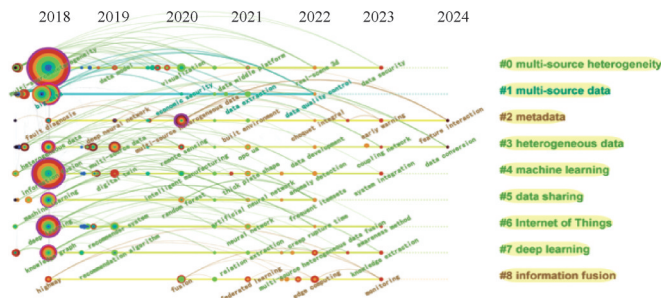


Fig.3 Timeline chart of Chinese research on multi-source heterogeneous data

The rise of *multi-source heterogeneity* as a research hotspot is driven by two main factors. Firstly, with the rapid advancement of information technology, data sources have become increasingly diverse, encompassing sensor data, social media data, and enterprise internal

data. These different types of data vary in structure, format, and semantics, creating an urgent need for effective multi-source and heterogeneous data processing technologies. For example, the integration and processing of massive real-time data generated by Internet of Things (IoT) devices with data from traditional databases rely heavily on multi-source and heterogeneous data management technologies. Secondly, the interdisciplinary integration of computer science, statistics, information science, and other disciplines has provided novel methods and perspectives for the research of multi-source and heterogeneous data. The introduction of methods such as machine learning and deep learning into the processing of multi-source and heterogeneous data has further enhanced the research enthusiasm and popularity of related keywords. For instance, deep learning models are utilized for feature extraction and fusion of multi-source

and heterogeneous data to improve the accuracy and efficiency of data analysis.

Under the joint drive of policy guidance and industrial demand, research and application of multi-source heterogeneous data have been promoted. From the policy perspective, a series of national documents on digital transformation and big data development were issued in China. In 2015, the *Action Plan for Promoting Big Data Development* was released. Cross-domain data integration and deep utilization were guided. In 2016, the Cybersecurity Law was enacted. In 2021, the Data Security Law and the Personal Information Protection Law were issued. Security and privacy requirements in data collection, storage, and processing were clarified. Directions for multi-source heterogeneous data development were provided. Enterprises and research institutions were compelled to explore compliant technologies. Practical adoption in various sectors was thus promoted.

Recent research has further advanced the integration of multi-source heterogeneous data in intelligent manufacturing. An IoT-based data fusion method was proposed to enhance integration within IoT-enabled manufacturing environments^[19]. Digital twin technologies for electromechanical-hydraulic systems were reviewed, with emphasis placed on prognostics and health management, where multi-source data are fused for real-time monitoring and predictive maintenance. A comprehensive review of Industrial Internet of Things (IIoT) intelligence in intelligent manufacturing was provided, outlining a hierarchical architecture and underscoring the role of multi-source data fusion across the value chain^[20]. In addition, an unsupervised multimodal fusion framework based on contrastive learning was introduced, enabling the fusion of heterogeneous modalities such as imagery, audio, and laser signals for process monitoring and anomaly detection^[21].

Compared with international research, more emphasis has been placed in China on algorithmic model construction and fundamental studies of data governance. In recent years, domestic studies have gradually extended toward integrated applications. However, gaps remain in the maturity of industrialization of core technologies and in the unification of cross-scenario collaboration standards.

Several underlying reasons can be identified. First, the intelligent manufacturing industry started earlier abroad. Enterprise practice and research were combined more closely. Technology deployment was therefore accelerated. Second, differences exist between domestic and international legal frameworks and industrial standards for data security. Compliance requirements in China slowed the pace of technological iteration. Third, in areas such as high-end sensors and industrial software, a clear first-mover advantage has been held internationally. More complete hardware and technical

support for multi-source data acquisition and fusion have thus been provided abroad.

To better understand multi-source heterogeneous data and reveal its underlying mechanisms in intelligent manufacturing, this paper explores its concepts and characteristics. We further analyze the key technologies and application scenarios in data processing and mining. Finally, we clarify the data and technology system frameworks and point out future development directions.

3 Concept and Characteristics

3.1 Concept

Industrial big data refers to data, along with related technologies and applications, generated throughout all stages of a product's life cycle, with intelligent manufacturing as its core focus in the industrial domain^[22]. Multi-source heterogeneous data refers to data collected from different sources, exhibiting varied structures or types^[23]. It is a subcategory of industrial big data and represents one of the most valuable data types for research, as illustrated in Fig.4.

Industrial big data emphasizes information collected across the full product life cycle. Its data types include traditional structured and semi-structured data, as well as unstructured engineering data. In contrast, multi-source heterogeneous data in intelligent manufacturing highlights the integration of data from different sources and structures. The extraction and fusion of features from this data is a key research focus. Its technical framework builds upon and extends the capabilities of traditional industrial big data processing^[24,25].

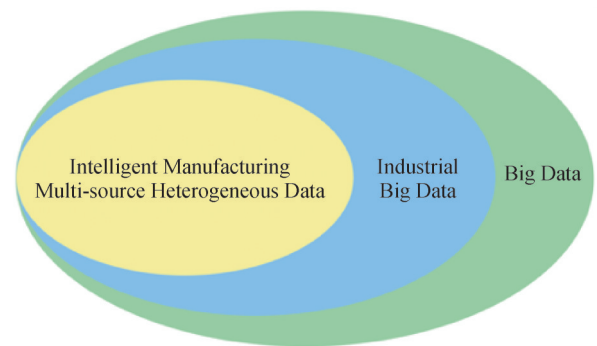


Fig.4 The relationship between multi-source heterogeneous data, industrial big data and intelligent manufacturing big data

Due to differing research priorities, there are notable differences between traditional industrial big data processing and the processing of multi-source heterogeneous data in intelligent manufacturing. Traditional industrial big data processing follows the entire data life cycle—from data acquisition to application. It covers a wide range of tasks including data collection, storage, analysis, decision-making support, and governance.

On the other hand, processing multi-source

heterogeneous data in intelligent manufacturing aims to achieve standardization and normalization. It focuses on uncovering and utilizing the similarities, differences, and correlations among various data sources. Challenges such as feature redundancy and fusion complexity are central. The goal is to eliminate redundant and repetitive information, identify incompatible content, and refine similar features while retaining the unique characteristics of each data type.

Based on this analysis, this paper defines multi-source heterogeneous data in intelligent manufacturing as data collected from various stages of the product life cycle—such as design, research and development, processing, and production. This data varies in source, type, and structure. It often shows strong correlation or redundancy, and has distinctive industrial traits, including high real-time requirements and considerable complexity.

In terms of application boundaries, industrial big data is usually focused on efficiency improvement in single stages, such as equipment performance monitoring or quality enhancement of a single process. Multi-source heterogeneous data, by contrast, is oriented toward global decision-making that supports cross-lifecycle and cross-department collaboration. Typical examples are design-manufacturing-operation integration and supply chain fusion. The value orientation of the two paths is therefore not identical. The latter places greater emphasis on cross-domain integration and high-dimensional value extraction.

3.2 Characteristics

As a special subset of industrial big data, multi-source heterogeneous data in intelligent manufacturing shares characteristics such as high throughput, strong correlations, and low quality with high noise^[26]. However, multi-source heterogeneous data in intelligent manufacturing distinguishes itself from traditional industrial big data by the diversity and heterogeneity of data sources, as well as the similarities, differences, and correlations among data features.

3.2.1 Diversity of Data Sources

Multi-source heterogeneous data in intelligent manufacturing can be classified into four categories based on its origin. The first category comes from internal enterprise information systems, such as Product Data Management (PDM) and Enterprise Resource Planning (ERP) systems. The second category originates from various intelligent devices on the production floor. Examples include Computer Numerical Control (CNC) machines, measuring instruments, forging and pressing equipment, and milling machines equipped with industrial IoT sensors. The third category comes from external feedback to manufacturing enterprises, such as online search engines and market feedback^[27]. The fourth category is derived from the production environment, such as temperature, humidity, air quality, terrain, and

traffic conditions^[28,29].

By integrating these sources, intelligent manufacturing data includes R&D records, equipment attributes, energy consumption logs, production schedules, operation status, environmental parameters, product-related information, monitoring videos and audio, and third-party data. It spans the entire production process, reflecting the high complexity and broad diversity of its sources.

3.2.2 Heterogeneity of Data

The heterogeneity of intelligent manufacturing data mainly stems from differences in data structure and data type. In industrial scenarios, the data can be grouped into three types: structured, semi-structured, and unstructured.

- Structured data is stored in relational databases and is typically two-dimensional. A common example is product order data.

- Unstructured data lacks a predefined model. It exists in varied formats and is harder to mine and integrate. Examples include monitoring videos, product images, Word/PDF documents, and engineering files like BOM drawings.

- Semi-structured data contains tags that separate semantic elements but do not follow relational database formats. This type includes formats like JSON and XML, which blend structured content with a flexible schema.

Even within the same data type, structures may vary across sources. For example, structured data from different databases may have inconsistent naming, fields, formats, or contents. Unstructured data also varies widely in format and structure. Semi-structured data, being self-describing, often lacks uniform standards. This diversity contributes significantly to the overall heterogeneity of multi-source heterogeneous data.

The heterogeneity of industrial big data is also reflected in the fundamental differences in information representation—namely, its multimodal characteristics. Multimodal data refers to datasets obtained through various types of perception devices or sensors, presenting different forms such as images, audio, vibration signals, temperature signals, and more. This type of data overcomes the limitations of single-source perception by collaboratively representing the same object or event^[30]. While multimodal data and multi-source heterogeneous data are categorized from different perspectives—the former focusing on differences in perception modalities, and the latter emphasizing differences in data sources and structures—they are closely related. Deep learning methods based on multimodal data are being increasingly applied in industrial fault diagnosis^[31,32].

3.2.3 Similarity, Correlations, and Differences Among Data Features

- Similarity means that data generated at different stages of the product life cycle describe the same product. Therefore, overlapping or similar features are inevitable.

- Correlation refers to the interdependence among data from different sources. Since the data describes the same product from multiple perspectives, it often contains mutually influential information. Extracting these correlations helps generate more refined and accurate features.

- Difference refers to the variation in features due to data type or perspective. Some data captures product characteristics that are unique to a specific stage of the life cycle.

The existence of similarity and correlation creates complementarity between data sources. For example, in fault diagnosis, infrared images and vibration signals are both useful. Each has its limitations, but combining them enhances the outcome. This leads to a synergistic effect—"1 + 1 > 2."

Based on this analysis, a multi-source heterogeneous data in an intelligent manufacturing system is constructed, as shown in Fig. 5. Data collected by intelligent workshop equipment includes unstructured data like images and videos, and structured time-series sensor data. Data from information systems is mostly structured, but also includes semi-structured log files and JSON, as well as unstructured images and PDF files. External enterprise data is mainly unstructured. Environmental data from the workshop surroundings is typically structured. Among all data, the similarities, correlations, and differences in features are the most meaningful aspects. These elements also serve as the key distinguishing characteristics of intelligent manufacturing data, setting it apart from traditional industrial big data.

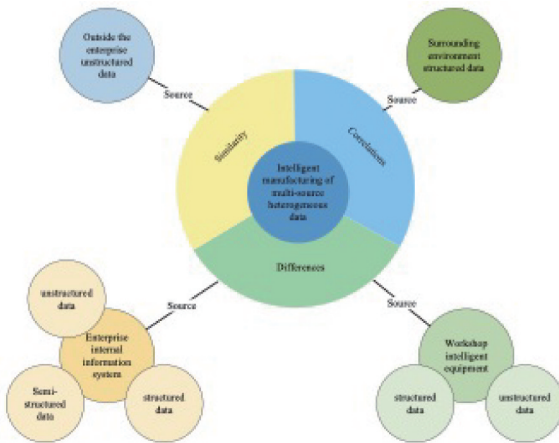


Fig.5 The multi-source heterogeneous data system of intelligent manufacturing

3.3 Architecture

To fully exploit the massive multi-source heterogeneous data generated throughout a product's life cycle, timely and effective data management and utilization are necessary to achieve the goal of enhancing manufacturing value driven by data. Therefore, a system for multi-source heterogeneous data in intelligent manufacturing is constructed^[33] to fully mine the value of

multi-source heterogeneous data in intelligent manufacturing and realize technological and intelligent upgrades for manufacturing enterprises.

Data acquisition is the foundation for processing multi-source heterogeneous data in intelligent manufacturing. Accurate data is a prerequisite for reasonable production scheduling, monitoring, and maintenance of production equipment. For different types of multi-source heterogeneous data in intelligent manufacturing, appropriate data collection methods and transmission approaches should be adopted to ensure that data is stored in a timely and accurate manner.

The diversity of the structures of multi-source heterogeneous data in intelligent manufacturing poses a significant challenge for data storage. Therefore, how to store the collected multi-source heterogeneous data is the foundation for analyzing multi-source heterogeneous data.

Multi-source heterogeneous data in intelligent manufacturing processing mainly includes processes such as data cleaning and dimensionality reduction, data integration, fusion, and data modeling. By leveraging techniques such as data mining, machine learning, and deep learning, the inherent patterns within the data can be extracted, and the complementary advantages of different data sources can be utilized to provide guidance for actual production activities.

At present, the multi-source heterogeneous data of intelligent manufacturing is mainly concentrated in many fields such as research and development design, equipment maintenance and production guidance, such as digital twin model development and equipment remaining life prediction.

4 Key Technologies

Based on the system architecture of multi-source heterogeneous data in intelligent manufacturing, a corresponding key technology framework is constructed, as shown in Fig.6. Core technologies are presented across six stages: data acquisition, transmission, organization, storage, processing, and application. Current methods are summarized, and a technical route for enterprise practice is offered. The novelty of the framework can be identified in three aspects. First, clear boundaries are defined between integration and fusion processes. Second, standardized interfaces for heterogeneous data governance are embedded. Third, each technical stage is explicitly linked to industrial application scenarios, including fault diagnosis, digital twin modeling, and production decision-making.

4.1 Acquisition

The sources of multi-source heterogeneous data are complex. They include equipment data, product data, process data, environmental data, and operational data. Due to different industrial communication protocol standards, the protocols used by various automation

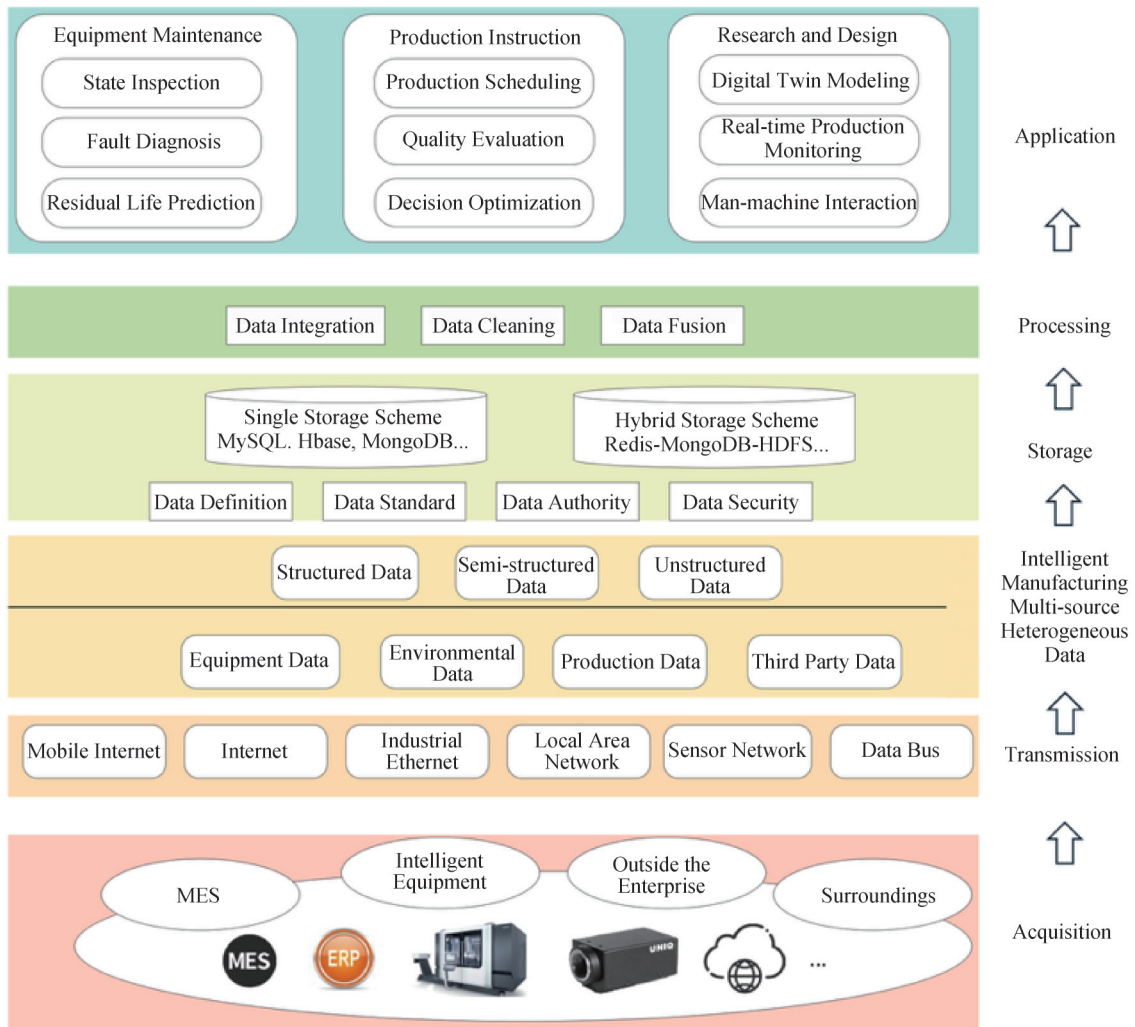


Fig.6 Key technology system for multi-source heterogeneous Data in intelligent manufacturing

devices are often inconsistent and incompatible. This results in differences in data formats, versions, and semantics across sources. Therefore, acquiring multi-source heterogeneous data requires designing specific access devices for each data type, along with protocol parsing and data transmission functions^[34].

The proposed framework has also been compared with the Industry 4.0 Reference Architecture Model. In the Reference Architectural Model Industrie 4.0 (RAMI 4.0), vertical and horizontal integration of business, functional, and technical layers is emphasized. In our framework, cross-stage feature correlations from design, production, and operation are highlighted. This difference shows the novelty of our approach. Redundant features are reduced. Decision support across the full product life cycle is improved.

Radio Frequency Identification (RFID) technology is used to collect real-time information such as product status, timestamps, and quality^[35,36]. Industrial sensors are deployed to obtain key performance indicators, such as current, voltage^[37], and vibration signals^[38]. Tools such as Flume and Kafka^[39] are used to process large volumes of log data. Software development kits provided by hardware vendors are utilized to collect multimedia data,

including audio and video^[40]. In addition, data can be exported from information management systems.

For communication protocols, technologies such as protocol conversion, parsing, and middleware can be used to unify different industrial protocols. In scenarios with high real-time and accuracy requirements, such as industrial robots, industrial Ethernet or fieldbus systems are typically adopted. Highly confidential data should be transmitted via local area networks. For workshop sensors, sensor networks can be constructed. For multimedia data such as images, audio, and video, industrial 5G can be utilized. Different transmission methods should be selected according to specific application requirements.

A modular acquisition scheme is proposed. Each module is designed for one type of data such as sensor signals, enterprise information, or multimedia content. The modules are flexibly combined according to enterprise needs. In a workshop environment, vibration signals and thermal images are acquired simultaneously and aligned by time stamps. A dual-source dataset for fault diagnosis is formed. This modular design ensures scalability and adaptability to different industrial scenarios.

4.2 Storage

Traditional databases such as MySQL, Oracle, and other relational systems, or non-relational databases like HBase and MongoDB, typically support only one or two data types. They struggle to meet the complex storage demands of heterogeneous data. As a result, hybrid storage solutions have emerged. One example is the Redis-MongoDB-HDFS architecture^[41].

Besides database research, some scholars have proposed storage algorithms for multi-source heterogeneous data. For example, they convert design, product, and service data into unified semantic expressions and store them in index structures^[42]. These emerging architectures and algorithms provide new ideas for data storage.

Storing multi-source heterogeneous data is not only about selecting databases and algorithms. It also involves managing massive and complex data. This includes data definitions, standards, permissions, and security. Data is distributed across systems. Data from different sources may have different storage modes, logical structures, and encoding protocols. To enable integration, the data must be standardized. Clear definitions and unified standards should be established to ensure consistency. Permissions must be properly managed. Security measures should be strengthened to prevent data leaks and ensure control.

4.3 Integration

In the actual production process, due to the complexity of the multi-source heterogeneous data sources in intelligent manufacturing and the diversity of collection methods, the collected raw data inevitably contains data inconsistencies, data conflicts, incomplete data, and data errors. Therefore, the first step is to clean the raw data, detect and process the "dirty data" in the raw data. Techniques such as missing value imputation^[43], noise removal, outlier identification, and data dimensionality reduction^[44] can be used to improve data quality, and the multi-source heterogeneous data can be integrated.

The integration of multi-source heterogeneous data primarily aims to overcome the structural differences among data from various sources and achieve the physical or logical connection of data sets with different characteristics. Traditional data integration methods include schema matching approach, XML based data integration, and federal/middleware approach^[45]. However, due to the neglect of data semantic information, they struggle to handle complex semantic relationships between data.

4.4 Fusion

Compared to data from a single source, multi-source heterogeneous data focuses on data feature fusion, taking the similarities, differences, and complementarity of data characteristics as the entry point. From multiple

dimensions and stages, it can reflect the true state of an entity. Data fusion technology can integrate the inherent features of data from different sources, overcome structural differences between different data sources, and enable data from different sources to complement each other. This allows for more accurate extraction of data features.

Data integration focuses on the structural and semantic alignment of heterogeneous data sources, achieving a unified logical or physical connection across different datasets. Data fusion, in contrast, is built on the results of integration and emphasizes feature-level processing. It extracts and combines complementary features from different data sources, aiming to generate more representative and discriminative data representations for downstream tasks^[46].

To further strengthen the technical discussion on data fusion, more fine-grained distinctions among different fusion approaches are highlighted here. In particular, feature-based fusion has evolved significantly with the adoption of cross-modal transformer models. For example, visual-language pretraining frameworks can be adapted to industrial scenarios, where machine vision images are combined with sensor time-series signals to improve fault diagnosis and anomaly detection accuracy. These frameworks utilize attention mechanisms to align heterogeneous modalities and capture temporal-spatial correlations that traditional fusion methods cannot fully exploit.

Moreover, a systematic comparison of stage-based, feature-based, and semantic-based fusion methods is provided in Table 1, summarizing their technical characteristics, advantages, limitations, and representative industrial use cases.

4.5 Applications

Single-source data can only reflect the state of an object or system from a certain aspect, and has certain limitations. The widespread use of numerous Internet of Things (IoT) devices has increased the quantity and dimensions of multi-source heterogeneous data in intelligent manufacturing^[47]. At present, the application of multi-source heterogeneous data in intelligent manufacturing is mainly concentrated in several areas, including equipment maintenance, production guidance, and research and development design.

Monitoring the status of industrial equipment and predicting its next stage is a typical scenario for the application of multi-source heterogeneous data in intelligent manufacturing. It is widely used in equipment condition monitoring^[48,49], remaining useful life prediction^[50], and fault diagnosis^[51].

In the field of production guidance, the main focus is to mine the potential value of multi-source heterogeneous data in intelligent manufacturing, providing guidance for enterprise production and promoting production optimization. This mainly involves industrial quality

Table 1 Comparison of multi-source heterogeneous data fusion methods in intelligent manufacturing

Fusion Methods	Technical Features	Advantages	Limitations	Typical Industrial Use Cases
Stage-based fusion	Data are integrated at early, intermediate, or late stages using rule-based or statistical methods.	Simple and interpretable; effective for low-dimensional sensor data.	Low scalability; limited effectiveness for high-dimensional unstructured data.	Fundamental equipment monitoring and anomaly alerting (e.g., temperature or pressure sensors).
Feature-based fusion	Features are extracted from different modalities using deep neural networks (e.g., CNNs for images, RNNs for signals). Recently, cross-modal transformers such as vision-language pretraining frameworks have been applied.	Captures complementary information; supports real-time diagnosis; improved accuracy with advanced models.	Requires large labeled datasets; high computational cost.	Real-time fault diagnosis (e.g., combining vibration signals with thermal images) and predictive maintenance.
Semantic-based fusion	Ontologies, knowledge graphs, and semantic mapping are applied for cross-domain integration.	Provides high-level abstraction; enables knowledge transfer across systems and organizations.	Difficult to construct unified ontologies; low adaptability to rapidly changing data.	Cross-departmental data sharing, supply chain integration, and digital twin construction.

assessment^[52,53], production decision optimization^[54], workshop resource allocation^[55], and other aspects.

In the field of research and development design, multi-source heterogeneous data in intelligent manufacturing is mainly used as a system driving force, with further development and utilization based on this data. It is primarily applied in the development of digital twin models^[56,57], equipment condition monitoring^[58,59], and other areas.

Furthermore, cross-domain integration and emerging trends have been emphasized. Multi-source heterogeneous data has been increasingly integrated with supply chain management, energy utilization, and workforce coordination to achieve system-wide optimization. Applications in predictive maintenance, green manufacturing, and carbon emission management have been recognized as critical directions. With the advancement of large-scale models and federated learning, multi-source heterogeneous data is expected to provide stronger support for full life-cycle management and sustainable value creation. In this way, the evolution of intelligent manufacturing is anticipated to be further promoted.

5 Case Study

Using China's furniture industry as an example, issues such as data collection and analysis difficulties, data chain connectivity challenges, and low design efficiency have been identified as major obstacles to the transformation and upgrading of intelligent manufacturing. The full utilization of multi-source heterogeneous data generated during production and manufacturing processes is regarded as essential. The enabling of data value is considered crucial for the digital transformation of the furniture industry.

This case study was derived from a joint project between the research group and a furniture manufacturing

enterprise. The research object was the furniture assembly board production line. This factory mainly focuses on the processing and manufacturing of assembly boards, covering processes such as raw material sorting, robotic stacking, finger-jointing, sanding, and curing.

The production process is as follows: The raw wood materials are sorted by color, texture, and other methods to select those with similar colors and textures. Different types of wood are then stacked using robotic arms. A finger-jointing machine is used to join the raw wood materials, and a finger-jointer combines wood blocks of similar sizes into long strips. These long strips then go through sanding and defect repair. Finally, long strips of the same size are assembled in a certain proportion to form wooden boards, which are then transferred to a conditioning room for curing, as shown in Fig.7.

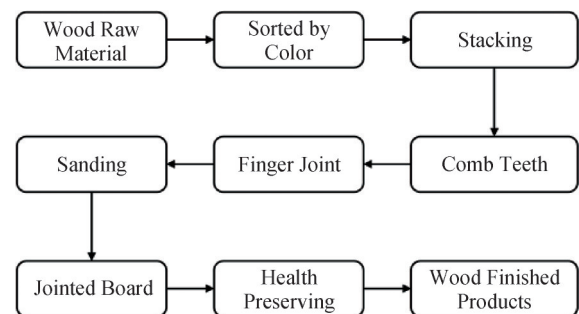


Fig.7 The procedure of wood processing

During these processes, large volumes of multi-source heterogeneous data are generated. Equipment-side data are collected through PLC systems and industrial sensors, including status, current, voltage, power, and fault information. Process-side data includes material batches, alarms, dispatching, and reporting streams. Environmental and energy data record temperature, humidity, and energy consumption per output. Visual data

are captured by station cameras, providing images for surface quality and defect inspection. All data are aggregated via local networks and 5G gateways and uploaded in real time to the enterprise cloud, ensuring temporal consistency and traceability.

A modular acquisition scheme is adopted. Each data source is connected through standardized interfaces. Timestamps are used for cross-source alignment. For storage, a layered design is employed. Online data are directed to time-series and cache layers to support real-time monitoring and early warning. Historical data are stored in distributed and relational databases, such as HBase and SQL Server, to enable retrospective analysis and reporting. Integration is achieved through codebooks and master data dictionaries, ensuring semantic alignment across systems for attributes such as energy, power, working hours, and material codes. Feature-level fusion is performed. Missing values are imputed, anomalies are suppressed, and redundant features are reduced through variance and correlation constraints. Standardization and alignment are carried out before concatenation, resulting in task-oriented joint feature representations. Applications focus on process quality assessment, energy anomaly detection, and resource optimization, with results fed

back through visual dashboards. The system design aligns with the enterprise's existing *5G + edge-cloud collaboration + distributed* storage architecture.

On the analytical side, Hadoop and distributed storage platforms are employed to organize and standardize the heterogeneous data. HBase is used for unstructured and semi-structured data, while SQL Server supports structured data, achieving cloud-based management. Task-driven feature fusion is adopted. For quality-related tasks, interpretable models such as gradient boosting trees are applied for process and energy evaluation and threshold-based warnings. Lightweight convolutional networks are used for defect image scoring, and probabilities are combined with time-series features. For energy-related tasks, short-term forecasting is performed through autoregressive models or lightweight RNNs, targeting energy consumption per unit output and key operating indicators. Feature engineering follows unified alignment, normalization, and robust scaling. Redundant features are removed by correlation and variance thresholds. Class imbalance is mitigated through undersampling and cost-sensitive training. Rolling windows are used for training, validation, and evaluation to ensure reproducibility.

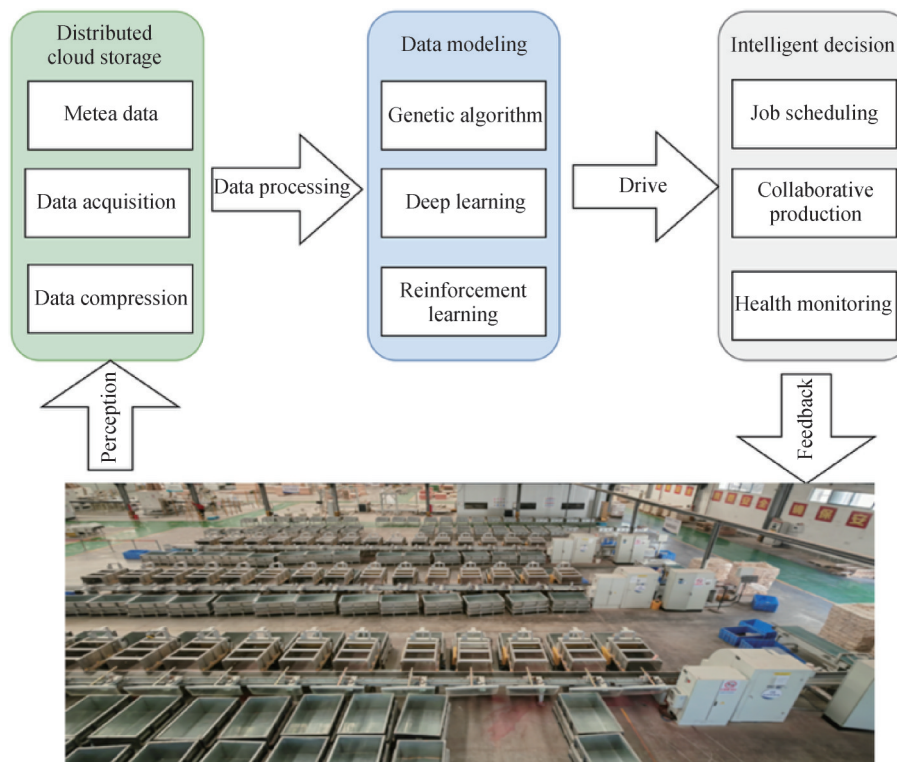


Fig.8 The application of multi-source heterogeneous data in an intelligent factory

In this case, the proposed framework is mapped onto a real production line scenario to demonstrate its potential applicability. Multi-source features derived from the production process were conceptually aligned with tasks such as quality assessment, anomaly detection, and energy management. For instance, modular acquisition and layered storage ensured that equipment-side, process-

side, environmental, and visual data could be captured and organized in a timely and consistent manner. Integration and feature-level fusion were then applied to generate task-oriented feature representations, which could support applications including process quality monitoring, energy anomaly detection, and resource optimization.

This case study demonstrates the feasibility and practical relevance of the proposed framework. Its key characteristics lie in modular acquisition, layered storage, semantic integration, feature-level fusion, and application-driven closed-loop feedback. Together, these features provide a conceptual pathway for intelligent manufacturing enterprises to achieve quality improvement, energy optimization, and predictive maintenance, and highlight the potential of the framework for broader application in discrete manufacturing scenarios.

6 Future Directions

6.1 Storage of Unstructured Data

Multi-source heterogeneous data in intelligent manufacturing encompasses structured, semi-structured, and unstructured data (e. g., images, videos, product drawings, BOM diagrams). Storage management must support unified storage and efficient retrieval across all these types, including specialized engineering data such as 3D model files. However, existing research predominantly addresses general unstructured data (images, videos) and overlooks domain-specific formats. Therefore, future work should develop storage frameworks that accommodate the professional characteristics and complex scenarios of manufacturing data.

6.2 Integration and Fusion of Multiple Data Types

The integration of multi-source heterogeneous data in intelligent manufacturing primarily addresses structural and semantic inconsistencies across datasets. Current studies mostly focus on structured data, but traditional schema-matching and middleware methods remain insufficient to handle the complexity of semantic conflicts in real industrial environments. Future research should therefore emphasize semantic-enhanced integration approaches, such as ontology construction and knowledge graph alignment, to achieve more reliable cross-source consistency and provide a solid foundation for subsequent processing.

Data fusion, in contrast, is conducted after integration and focuses on feature-level representation. By combining complementary information across modalities (e. g., vibration signals, thermal images, and textual logs), fusion can generate more discriminative features that better reflect the true state of industrial entities. However, most existing work is still limited to sensor-level fusion. The potential of multimodal deep fusion, particularly cross-modal transformers and generative models, remains underexplored. Future directions should include exploring fine-grained feature alignment, adaptive weighting strategies, and robustness enhancement under missing-modality scenarios, so as to

fully exploit the value of multi-source heterogeneous data.

In summary, integration resolves *how data can be connected and made consistent*, while fusion resolves *how features can be jointly represented and utilized*. A clear distinction between these two stages ensures that heterogeneous data can support more accurate fault diagnosis, predictive maintenance, and digital twin applications in intelligent manufacturing.

6.3 Lack of High-quality Multi-source Heterogeneous Datasets

Currently, publicly available datasets in the manufacturing industry typically only provide one-dimensional time-series data collected from multiple sensors, lacking truly diverse heterogeneous data such as images, videos, and text. Therefore, future research efforts should focus on constructing high-quality multi-source heterogeneous datasets from real industrial scenarios to drive further development of technologies related to multi-source heterogeneous data in intelligent manufacturing. This limitation also underscores the necessity of coordinated efforts and innovative technical solutions, which are further discussed in Section 6.4.

6.4 Emerging Solutions and Pathways

To address the aforementioned challenges, several promising research and application directions can be further pursued.

A combined architecture can be adopted, where edge nodes preprocess unstructured files for real-time access and rapid response, while cloud systems focus on long-term archiving and integrity verification. In particular, blockchain-based mechanisms can be introduced to ensure the immutability and traceability of archived manufacturing data.

Industry-academia collaboration should be encouraged to build standardized multi-source heterogeneous data benchmark datasets. These datasets may include time-series sensor signals, CAD files, and production logs from sectors such as automotive and electronics. To guarantee compliance with privacy and data security requirements, federated learning and privacy-preserving data sharing frameworks can be employed.

Generative artificial intelligence, especially diffusion models, can be leveraged to enrich scarce non-structured data and enhance the robustness of fusion processes. By synthesizing missing or under-represented modalities, generative models can significantly improve the resilience of data-driven intelligent manufacturing systems under real-world incomplete data scenarios.

Overall, addressing these future challenges through advances in storage, integration, dataset construction, and emerging solutions will significantly enhance the utilization of multi-source heterogeneous data, enabling more reliable fault diagnosis, predictive maintenance, and digital twin applications in intelligent manufacturing.

7 Conclusion

This paper first elucidates the connotation, characteristics, and differences and similarities between multi-source heterogeneous data in intelligent manufacturing and industrial big data. It then analyzes and discusses the key technologies for processing multi-source heterogeneous data in intelligent manufacturing and proposes a main technical system framework. Finally, it explores the future development directions of multi-source heterogeneous data in intelligent manufacturing. As the volume of multi-source heterogeneous data in intelligent manufacturing gradually increases, the manufacturing industry urgently needs new technologies to realize the management of massive multi-source heterogeneous data and eliminate the barriers between different types of data, thereby promoting the transformation of traditional manufacturing and facilitating the development of intelligent manufacturing.

Author Contribution:

Minghao Zhu: Supervision, Project administration, Writing-review & editing. Pengfei Yang: Analysis and interpretation of data, Writing-original draft, Software. Bo Gao: Conceptualization, Writing-review & editing, Supervision. Xuehan Li: Conceptualization, Writing-review & editing, Supervision. Letian Wang: Analysis and interpretation of data, Writing-original draft, Software. All authors have read and agreed to the published version of the manuscript.

Foundation Information:

This research was funded by the National Natural Science Foundation of China, grant number 62172033.

Conflicts of Interest:

The authors declare no competing interests.

Data Availability:

The authors declare that the main data supporting the findings of this study are available within the paper and its Supplementary Information files.

Dates:

Received 4 July 2025; Accepted 15 December 2025; Published online 31 March 2026

References

- [1] Liu Y, Sun J, Jiang Y, et al (2020). 4C Model: Value Discovery in Big Data [J]. *Journal of Management World* , 36 (02): 129-138+223. <https://doi.org/10.19744/j.cnki.11-1235/f.2020.0026>
- [2] Chen S, Mao X, Wang C (2022). Missing Data Completion in the Big Data Era: Challenges and Solutions [J]. *Journal of Management World* , 38(01): 196-207. <https://doi.org/10.19744/j.cnki.11-1235/f.2022.0015>
- [3] Nagorny K, Lima-Monteiro P, Barata J, et al (2017). Big Data Analysis in Smart Manufacturing: A Review [J]. *Network and System Sciences*, 10(3): 28. <https://doi.org/10.4236/ijcns.2017.103003>
- [4] Lin Y, Chen R, Jin T (2020). Research on multi-source heterogeneous data fusion technology for complex information system [J]. *China Measurement & Test* , 46(07):1-7+23.
- [5] Cai B, Huang L, Xie M (2017). Bayesian networks in fault diagnosis[J]. *IEEE Transactions on Industrial Informatics* , 13 (5): 2227-2240. <https://doi.org/10.1109/TII.2017.2695583>
- [6] Zhang Q (2018). Adaptive Kalman filter for actuator fault diagnosis[J]. *Automatica* , 93: 333-342. <https://doi.org/10.1016/j.automatica.2018.03.075>
- [7] Sun J, Wang P, Luo Y, et al (2019). Surface defects detection based on adaptive multiscale image collection and convolutional neural networks[J]. *IEEE Transactions on Instrumentation and Measurement* , 68(12): 4787-4797. <https://doi.org/10.1109/TIM.2019.2899478>
- [8] Huang J, Jiang Z, Liu J, et al (2024). Demystifying and extracting fault-indicating information from logs for failure diagnosis[C]. 2024 IEEE 35th International Symposium on Software Reliability Engineering (ISSRE). IEEE, 2024: 511-522. <https://doi.org/10.1109/ISSRE62328.2024.00055>
- [9] Toorajipour R, Sohrabpour V, Nazarpour A, et al (2021). Artificial intelligence in supply chain management: A systematic literature review[J]. *Journal of Business Research* , 122: 502-517. <https://doi.org/10.1016/j.jbusres.2020.09.009>
- [10] Carvalho D, Pereira E, Cardoso J (2019). Machine learning interpretability: A survey on methods and metrics[J]. *Electronics* , 8(8): 832. <https://doi.org/10.3390/electronics8080832>
- [11] Chen C, Song M (2019). Visualizing a field of research: A methodology of systematic scientometric reviews [J]. *PLoS one*, 14(10): e0223994. <https://doi.org/10.1371/journal.pone.0223994>
- [12] Kong L, Peng X, Chen Y, et al (2020). Multi-sensor measurement and data fusion technology for manufacturing process monitoring: a literature review [J]. *International journal of extreme manufacturing* , 2(2):022001 (27pp). <https://doi.org/10.1088/2631-7990/ab7ae6>
- [13] Gawde S, Patil S, Kumar S, et al (2023). A scoping review on multi-fault diagnosis of industrial rotating machines using multi-sensor data fusion [J]. *Artificial Intelligence Review* , 56 (5). <https://doi.org/10.1007/s10462-022-10243-z>
- [14] Lee J, Bagheri B, Kao H (2015). A cyber-physical systems architecture for industry 4.0-based manufacturing systems[J]. *Manufacturing letters* , 3: 18-23. <https://doi.org/10.1016/j.mfglet.2014.12.001>
- [15] Zhang L, Luo Y, Tao F, et al (2014). Cloud manufacturing: a new manufacturing paradigm[J]. *Enterprise Information Systems*, 8(2): 167-187. <https://doi.org/10.1080/17517575.2012.683812>
- [16] Li B, Jiang G, Li N (2024). Research on large-scale structured and unstructured data processing based on large language

- model[C]. Proceedings of the International Conference on Machine Learning, Pattern Recognition and Automation Engineering. **2024**: 111-116. <https://doi.org/10.1145/3696687.3696707>
- [17] Chkirkbene Z, Hamila R, Gouissem A (**2024**). Large language models (llm) in industry: A survey of applications, challenges, and trends[C]. 2024 IEEE 21st International Conference on Smart Communities: Improving Quality of Life using AI, Robotics and IoT (HONET). IEEE, **2024**: 229-234. <https://doi.org/10.48550/arXiv.2412.16089>
- [18] Liu Q, Qu Q, Zhao D, et al (**2021**). Dynamic quantitative assessment method of chemical safety risk based on multi-source heterogeneous data fusion [J]. *CIESC Journal* , 72(03): 1769-1777.
- [19] Sun R, Ren Y (**2024**). A multi-source heterogeneous data fusion method for intelligent systems in the Internet of Things [J]. *Intelligent Systems with Applications* , 23: 200424. <https://doi.org/10.1016/j.iswa.2024.200424>
- [20] Hu Y, Jia Q, Yao Y, et al (**2024**). Industrial internet of things intelligence empowering smart manufacturing: A literature review[J]. *IEEE Internet of Things Journal* , 11(11): 19143-19167. <https://doi.org/10.1109/JIOT.2024.3367692>
- [21] McKinney M, Garland A, Cillessen D, et al (**2025**). Unsupervised multimodal fusion of in-process sensor data for advanced manufacturing process monitoring[J]. *Journal of Manufacturing Systems* , 78: 271-282. <https://doi.org/10.1016/j.jmsy.2024.12.003>
- [22] Khan M, Wu X, Xu X, et al (**2017**). Big data challenges and opportunities in the hype of industry 4.0 [C]. 2017 IEEE International Conference on Communications (ICC), pp. 1-6. <https://doi.org/10.1109/ICC.2017.7996801>
- [23] Zhang L, Xie Y, Luan X, et al (**2018**). Multi-source heterogeneous data fusion [C]. 2018 International conference on artificial intelligence and big data (ICAIBD), IEEE. pp. 47-51. <https://doi.org/10.1109/ICAIBD.2018.8396165>
- [24] Wang G, Liu J, Li G, et al (**2021**). System protection of a pipe corridor power cabin based on multi-source heterogeneous data fusion [J]. *Power System Protection and Control* , 49(07): 103-109. <https://doi.org/10.19783/j.cnki.pspc.200904>
- [25] Hu Y, Piao X, Sun Y, et al (**2013**). Multi-source heterogeneous data fusion method and its application in object positioning and tracking [J]. *Scientia Sinica (Informationis)* , 43(10): 1288-1306. <https://doi.org/10.1360/112013-120>
- [26] Yan J, Meng Y, Lu L, et al (**2017**). Industrial big data in an industry 4.0 environment: Challenges, schemes, and applications for predictive maintenance [J]. *IEEE Access* , 5, 23484-23491. <https://doi.org/10.1109/ACCESS.2017.2765544>
- [27] Wang J, Zhang W, Shi Y, et al (**2018**). Industrial Big Data Analytics: Challenges, *Methodologies, and Applications* [J]. <https://doi.org/10.48550/arXiv.1807.01016>
- [28] Li X, Jing T, Li R, et al (**2022**). BDRA: Blockchain and decentralized identifiers assisted secure registration and authentication for VANETs [J]. *IEEE Internet of Things Journal* , 10(14): 12140-12155. <https://doi.org/10.1109/JIOT.2022.3164147>
- [29] Li X, Jing T, Wang X, et al (**2024**). Two-Stage Offloading for an Enhancing Distributed Vehicular Edge Computing and Networks: Model and Algorithm [J]. *IEEE Transactions on Intelligent Transportation Systems*, **2024**. <https://doi.org/10.1109/TITS.2024.3424852>
- [30] Wen C, Lv F (**2024**). Review on Deep Learning Based Fault Diagnosis [J]. *Journal of Electronics & Information Technology* , 42(01): 234-248.
- [31] Jiang G, Yang Y, Wang Z (**2024**). Fault Diagnosis of Gearbox of CNC Machine Tool Based on Multimodal Ensemble Convolutional Neural Network [J]. *Machine Tool & Hydraulics* , 52(08): 202-207.
- [32] Gao L, Lei W, Cao Y, et al (**2024**). Research on Deep Multimodal Transfer Learning in Bearing Fault Diagnosis [J]. *Machinery Design & Manufacture* , (04): 145-148+153. <https://doi.org/10.19356/j.cnki.1001-3997.20231018.006>
- [33] Kong Y, Li C, Chu D (**2022**). Cross modal storage and retrieval system for multi - source heterogeneous data [J]. *Journal of Nanjing University (Natural Science)* , 58(03): 377-385. <https://doi.org/10.13232/j.cnki.jnju.2022.03.002>
- [34] Zhang X, Xue D, Zheng J, et al (**2021**). The Transmission and Parse Technology of Multi-source Heterogeneous Data Based on OPC UA in Intelligent Manufacturing [J]. *Mechanical & Electrical Engineering Technology* , 50(01):1-7.
- [35] Leng J, Jiang P (**2019**). Dynamic scheduling in RFID-driven discrete manufacturing system by using multi-layer network metrics as heuristic information[J]. *Journal of Intelligent Manufacturing* , 30, 979-994.
- [36] Costa F, Genovesi S, Borgese M, et al (**2021**). A review of RFID sensors, the new frontier of internet of things[J]. *Sensors* , 21,3138. <https://doi.org/10.3390/s21093138>
- [37] Lessmeier C, Kimotho J K, Zimmer D, et al (2016). Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification [C]. PHM Society European Conference. <https://doi.org/10.36001/phme.2016.v3i1.1577>
- [38] Lei Y, Li N, Guo L, et al (**2018**). Machinery health prognostics: A systematic review from data acquisition to RUL prediction [J]. *Mechanical systems and signal processing*, 104, 799-834. <https://doi.org/10.1016/j.ymssp.2017.11.016>
- [39] Park J, Chi S (**2016**). An implementation of a high throughput data ingestion system for machine logs in manufacturing industry [C]. 2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN), IEEE. pp. 117-120. <https://doi.org/10.1109/ICUFN.2016.7536997>
- [40] Zheng S, Bo C, Du F (**2023**). Spatiotemporal video data acquisition and application for intelligent manufacturing [J]. *Journal of Electronics and Information Science* , 8, 8-12. <https://doi.org/10.23977/jeis.2023.080202>
- [41] Sun Y, Jiang H, Ding G (**2022**). Data modeling, integration and storage technology for production process management

- and control [J]. *China Mechanical Engineering* , 33, 356. <https://doi.org/10.3969/j.issn.1004-132X.2022.03.012>
- [42] Kong Y, Li D, Li C, et al (2021). A multi-source heterogeneous data storage and retrieval system for intelligent manufacturing [C]. 2021 IEEE International Conference on e-Business Engineering (ICEBE), IEEE. pp. 82-87. <https://doi.org/10.1109/ICEBE52470.2021.00032>
- [43] Wang Y, Li K, Gan S, et al (2019). Missing data imputation with OLS-based autoencoder for intelligent manufacturing [J]. *IEEE Transactions on Industry Applications* , 55, 7219-7229. <https://doi.org/10.1109/TIA.2019.2940585>
- [44] Joswiak M, Peng Y, Castillo I, et al (2019). Dimensionality reduction for visualizing industrial chemical process data [J]. *Control Engineering Practice* , 93, 104189. <https://doi.org/10.1016/j.conengprac.2019.104189>
- [45] Almutairi M M, Yamin M, Halikias G (2021). An analysis of data integration challenges from heterogeneous databases [C]. 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom), pp. 352-356.
- [46] Zheng Y (2015). Methodologies for cross-domain data fusion: An overview [J]. *IEEE transactions on big data* , 1, 16-34. <https://doi.org/10.1109/TBDDATA.2015.2465959>
- [47] Jia F, Xiong G, Zhu F, et al (2019). Research and implementation of industrial Internet of things communication system based on MQTT [J]. *Chinese Journal of Intelligent Science and Technology* , 1(03):249-259.
- [48] Liu W, Tian Z, Jiang X, et al (2022). A milling cutter state recognition method based on multi-source heterogeneous data fusion [J]. *The International Journal of Advanced Manufacturing Technology* , 122, 3365-3378. <https://doi.org/10.1007/s00170-022-10017-5>
- [49] Wang X, Yan J (2024). Deep learning based multi-source heterogeneous information fusion framework for online monitoring of surface quality in milling process [J]. *Engineering Applications of Artificial Intelligence* , 133, 108043. <https://doi.org/10.1016/j.engappai.2024.108043>
- [50] Bai R, Noman K, Yang Y, et al (2024). Towards trustworthy remaining useful life prediction through multi-source information fusion and a novel LSTM-DAU model [J]. *Reliability Engineering & System Safety* , 110047. <https://doi.org/10.1016/j.res.2024.110047>
- [51] Yan J, Hu Y, Guo C (2019). Rotor unbalance fault diagnosis using dbn based on multi-source heterogeneous information fusion [J]. *Procedia Manufacturing* , 35, 1184-1189. <https://doi.org/10.1016/j.promfg.2019.06.075>
- [52] Xiao Q, Shan M, Xiao X, et al (2020). Evaluation model of industrial operation quality under multi-source heterogeneous data information [J]. *International Journal of Fuzzy Systems* , 22, 522-547. <https://doi.org/10.1007/s40815-019-00776-x>
- [53] Liu L, Wan X, Li J, et al (2022). An improved entropy-weighted topsis method for decision-level fusion evaluation system of multi-source data [J]. *Sensors* , 22, 6391. <https://doi.org/10.3390/s22176391>
- [54] Yu J, Gu S, Wang J, et al (2020). The intelligent decision-making based on multisource heterogeneous data fusion in manufacturing [C]. 2020 12th International Conference on Advanced Computational Intelligence (ICACI), IEEE. pp. 155-159. <https://doi.org/10.1109/ICACI49185.2020.9177661>
- [55] Zhou B, Bao J, Li J, et al (2021). A novel knowledge graph-based optimization approach for resource allocation in discrete manufacturing workshops [J]. *Robotics and Computer-Integrated Manufacturing* , 71, 102160. <https://doi.org/10.1016/j.rcim.2021.102160>
- [56] Miao Q, Zou W, Liu L, et al (2020). Intelligent workshop digital twin virtual reality fusion and application [C]. *Advanced Manufacturing and Automation IX 9th*, Springer. pp. 585-592. https://doi.org/10.1007/978-981-15-2341-0_73
- [57] Ren Z, Wan K, Zhang R, et al (2023). Digital twin portrait: a fusion and application method of multi-source twin data for flexible manufacturing line [J]. *IEEE Journal of Emerging and Selected Topics in Industrial Electronics* . <https://doi.org/10.1109/JESTIE.2023.3317792>
- [58] Huang Y, Tao J, Sun G, et al (2023). A novel digital twin approach based on deep multimodal information fusion for aero-engine fault diagnosis [J]. *Energy* , 270, 126894. <https://doi.org/10.1016/j.energy.2023.126894>
- [59] Xia J, Huang R, Chen Z, et al (2023). A novel digital twin-driven approach based on physical-virtual data fusion for gearbox fault diagnosis [J]. *Reliability Engineering & System Safety*, 240, 109542. <https://doi.org/10.1016/j.res.2023.109542>