Future Technology
Open Access Journal
ISSN 2832-0379

May 2026| Volume 05 | Issue 02 | Pages 148-158

Journal homepage: https://fupubco.com/futech

https://doi.org/10.55670/fpll.futech.5.2.14

Future Publishing LLC

Article

A multi-model database framework for interoperable IoT sensor data management in smart manufacturing systems

Jagrutiben Padhiyar*

Gujarat Technological University, Gujarat, India

ARTICLE INFO

Article history:
Received 24 September 2025
Received in revised form
17 December 2025
Accepted 25 January 2026

Keywords:
IoT data management, Smart manufacturing, Multi-model databases, Semantic interoperability, Industry 4.0, Time-series analytics

*Corresponding author
Email address:
jagrutipadhiyar6@gmail.com

DOI: 10.55670/fpll.futech.5.2.14

ABSTRACT

The explosive adoption of IoT enabled smart manufacturing has increased the complexity of managing heterogeneous sensor data coming from diverse machines, communication protocols, and vendor-specific formats significantly. Conventional relational and time-series databases are very hard to adapt to the twin problems of high volume of data, structural diversity, and semantic inconsistency in industrial environments today. This paper proposes a multi-model database system in order to achieve high-performance interoperability for heterogeneous IoT sensor streams in the Smart Manufacturing Systems. The architecture includes a semantic integration layer to transform the data coming from formats such as JSON, XML, CSV, OPC-UA, and MQTT into a common canonical data model. The framework is evaluated on a synthetic but realistic Industry 4.0 dataset with roughly 5000 devices and over 40000 sensor measurements that allows the ingestion performance, cross-sensor query latency, and scalability of the framework to be evaluated. Experimental results show increased interoperability, support of unified cross-modal queries and low latency performance under growing loads of data. Furthermore, the cross-sensor correlation and analysis of any anomaly points to the applicability of the framework to analytics-oriented tasks, such as the early detection of abnormal machine behaviour. In general, the offered solution offers a semantically consistent, scalable base of interoperable IoT data management of smart manufacturing settings.

1. Introduction

The Internet of Things (IoT) has become a key technology in smart manufacturing, fundamentally altering the way industrial production processes are done through pervasive sensing, connectivity, and data-driven automation [1]. Within the scope of Industry 4.0, cyber-physical systems, interconnected machines, and large-scale sensor networks provide the ability to monitor, control adaptively and make intelligent decisions continuously throughout production environments [2]. Modern manufacturing plants are normally equipped with thousands of heterogeneous sensors that monitor vibration, temperature, energy consumption, humidity, spindle load and machine status and generate high frequency and uninterrupted data streams [3]. Such data streams are essential to such critical applications as predictive maintenance, fault diagnosis, quality optimisation, and operational forecasting, so the effective and trustworthy data management is a major need of digital manufacturing ecosystems [4]. Despite all of these opportunities, however, interoperability is one of the major issues in industrial IoT deployments. Sensor data is generated in various ways and formats, e.g., JavaScript Object Notation (JSON), Extensible Markup Language (XML), comma-separated value (csv) and binary encoding, industrial communication protocols (e.g. Open Platform Communication Unified Architecture (OPC UA) and Message Queuing Telemetry Transport (MQTT)) [5]. Proprietary schema and metadata conventions, and contextual parameters are frequently defined by each vendor of the device, which has created data silos that are fragmented and complex extract transform, load (ETL) pipelines. This lack of cohesion raises the cost in integration, creates semantic inconsistencies and makes it more difficult to develop unified dashboards, cross-sensor analytics and machine learning models [6]. As a result, smart factories cannot make the most out of real-time analytics and predictive intelligence without interoperable and semantically consistent representations of data. Traditional data management technologies have a hard time meeting these requirements. Conventional relational databases are based on rigid schemas that are not well-suited to changing and semi-structured sensor data and result in performance degradation under frequent schema changes and irregularity of the time-series patterns [7]. While time-

series databases are optimised for high-volume numerical measurements, they do not provide built-in support for how to represent complex relationships, rich metadata and context information associated with industrial assets [8]. This leads to a situation whereby manufacturers are more willing to find flexible data platforms which can support heterogeneous data models and relationships without rigid schema constraints.

Multi-model databases have recently caught the attention of the community as a promising option for solving the data variety challenge in industrial IoT environments [9]. By supporting different data models (e.g. document, graph, key-value, and semantic representations) in a single engine, these systems provide the ability to index and query the same data consistently across different, varied data forms. Multi-model architectures are especially appealing for smart manufacturing scenarios, where it is required to model sensor readings, device-machine relations, event sequences and hierarchical asset structures in an integrated way [10]. Furthermore, the flexibility of combining document-based data, graph and semantic representations to support semantic normalisation and cross-sensor reasoning (which are necessary for advanced analytics).

There are several studies that have studied IoT data platforms, middleware-based integration frameworks and semantic interoperability models [11]. Other works have looked into fog and edge computing architectures to minimise latency and bandwidth usage in industrial IoT systems [12]. More recently, edge computing applications with a view to industrial applications have been surveyed to make clear their position in decentralised analytics and control [13]. However, most of the existing researches are limited to single layers of the data pipeline (e.g. communication middleware, semantic modelling, storage technologies) without jointly assessing the interoperability effectiveness and database performance in a single, integrated architecture. In particular, experimental investigations to evaluate the harmonisation of heterogeneous industrial sensor data in multi-model databases to support semantically enriched cross-source queries and scalable implementation under realistic smart manufacturing workload remain limited [14].

This gap provides the impetus for the design of a holistic interoperability framework using a combination of semantic integration, canonical data modelling, and multi-model storage coupled with empirical performance evaluation. The proposed approach covers the data heterogeneity aspect through semantic normalisation and canonical entity modeling and allows scalable data ingestion and low-latency cross-sensor analytics. The major contributions of this work can be summarised as follows. In order to achieve interoperability among heterogeneous IoT sensors, a three-layer interoperability framework is proposed, which can be used to align heterogeneous IoT sensor streams by applying semantic integration and multi-model database storage. Second, a canonical IoT data model and semantic mapping plan, which is rule-based are presented as a strategy to help represent and cross-contextualise industrial devices. Third, the interoperability, ingestion throughput, query latency, storage overhead and scalability are experimentally evaluated using a synthetic yet realistic smart manufacturing dataset. Finally, the results show the appropriateness of the proposed framework for analytics-based applications, including predictive maintenance and machine health monitoring, in the data ecosystem of Industry 4.0. Accordingly, the aim of this work is to design and test a multi-layer interoperability framework for smart manufacturing systems based on a multi-model database backend.

- the need to enable interoperable and semantically consistent storage of heterogeneous industrial IoT sensor data coming from multiple formats and communication protocols;
- for supporting efficient cross-sensor and cross-machine querying over integrated data for a unifiedanalytics;
- to test the ingestion throughput, query latency, storage overhead, and scalability of a multi-model database backend with smart manufacturing workloads.
- to show the applicability of the proposed framework to the analytics, such as correlation analysis and anomaly detection, applicable to predictive maintenance in the Industry 4.0 environments.

The rest of this paper is structured as follows. Section 2 discusses the related literature on interoperability and data management of IoT. Section 3 presents the proposed methodology, such as the framework architecture, data model and evaluation setup. The performance analysis and the results of the experiment is presented in Section 4. The findings, limitations and future research directions are discussed in Section 5. Finally, Section 6 concludes the paper.

2. Literature review

Intelligent factory spaces create large amounts of non-homogeneous data, which is caused by interconnected machines, sensors, and cyber-physical structures. Industrial deployments typically involve large numbers of sensing devices that operate on a continuous basis over a long period of time, leading to large amounts of data that must be retained for monitoring, optimisation, and compliance purposes [15]. The integration and management of such data is still challenging as a result of the co-existence of high data velocity, diverse formats as well as evolving schemas which are common characteristics of Industry 4.0 ecosystems. One of the main problems of industrial Internet of Things (IIoT) systems is cross-heterogeneous device, platform, and data representation interoperability. Prior studies have emphasised that data heterogeneity occurred not only on the communication protocol level but also on the syntactic and semantic levels, which have led to fragmented data silos and expensive integration pipelines [16]. They have proposed standards like Sensor Measurement Lists (SenML) which are lightweight sensor measurements and sensor metadata representations, and enhance syntactic interoperability with limited devices and applications [17]. However, such standards basically consider issues related to data representation and exchange, but do not take into full account long-term data storage, indexing, and unified querying of heterogeneous data sources.

Industrial communication protocols, such as Open Platform Communications Unified Architecture (OPC UA), are important for facilitating the possibility of standardised access to machine-level data and events [18]. As much as OPC UA can be used to work with rich data models and semantic descriptions, current literature suggests that the performance and interoperability features are often limited to the interface, and little to nothing is integrated with underlying data management systems to support large-scale analytics [19]. Consequently, the semantic consistency, obtained in the storage of data, is often lost when data is stored and processed in the downstream direction. Several architectural approaches against data management problems in industrial IoT systems have been suggested. The software-defined industrial IoT architectures focus on flexibility and

programmability of data streams and control protocols that make them adapt dynamically to emerging production needs [20]. Similarly, industrial big data systems have investigated cloud-based analytics pipelines to predictive maintenance and condition monitoring, with the necessity to scale to large volumes of data and be able to query the data efficiently [21]. Analytical frameworks for IIoT further emphasise the need for harmonisation of the data across the heterogeneous data sources for supporting advanced analytics and decision-making [22]. Cloud-based big data platforms have been extensively researched for large-scale industrial data management with elasticity and computational scalability [23]. Nevertheless, the use of centralised cloud infrastructures creates issues on elements such as latency, data sovereignty, and operational technology integration. Conceptual definitions of Internet of Things emphasise the need for distributed intelligence and seamless integration between the physical and the digital components, that is difficult to achieve with monolithic data architectures [24]. Physical processes are also intensified by cyber-physical manufacturing systems that integrate a digital control and analytics layer firmly with the physical process [25].

More recent work has investigated new paradigms including the so-called physical artificial intelligence, which combines sensing, reasoning and actuation in industrial environments [26]. These approaches strongly depend on unified and semantically consistent representations of data in order to render intelligent behaviour possible. The research on fundamental IoT activities still recognises interoperability, scalability, and data management as ongoing challenges that limit the achievement of the fully independent industrial systems [27]. In addition, knowledge graph-based analysis of Industry 4.0 standards exposes large amount of fragmentation across data models and vocabularies further reinforcing the need for unified backend data management solutions that can support semantic integration at scale [28]. Despite the fact that much work has been done on standardisation for interoperability and industrial architectures and big data analytics, existing research studies usually focus on individual components of the industrial data pipeline in isolation. There is still a lack of comprehensive experimental investigations that consider a joint evaluation of the semantic normalisation, multi-model data storage and system performance in one smart manufacturing architecture. Specifically, there is still empirical evidence showing how multi-model databases can be used to harmonise heterogeneous industrial sensor information, assist with semantically-enhanced cross-source queries, and scale to realistic Industry 4.0 loads. This gap has to be filled in to evaluate the feasibility of multi-model architectures as a foundation to interoperable and analytics-driven smart manufacturing systems.

3. Methodology

3.1 Problem formulation

The proposed study deals with a smart manufacturing environment that includes a large-scale deployment of heterogeneous machines and Internet of Things (IoT) sensors which are distributed on multiple production lines. In today's Industry 4.0 factories, such environments often include thousands of sensors per place, taking samples of such physical variables as temperature, vibration, and energy consumption at rates ranging from 1 Hz to a few kHz, resulting in millions of measurements per hour and continuous long-term accumulation of data [1,2,19]. These data streams are produced using a variety of data representations such as in the form of a (JSON), XML or CSV, binary payloads and industrial communication protocols such as OPC Unified Architecture (OPC UA) and Message Queuing Telemetry Transport (MQTT).

The driving challenge that has been tackled in this study is supporting the persistent integration and unified querying of high-velocity and heterogeneous sensor data with semantic consistency and low-latency performance. This challenge is driven by the fact that (i) there is structural heterogeneity both in payload formats and schemas, (ii) there is semantic inconsistency in device metadata and measurement descriptions and (iii) there are stringent requirements in terms of latency for industrial monitoring and analytics. Practically, industrial monitoring applications are often forgiving of latencies in the range of tens to hundreds of milliseconds, and advanced analytics and anomaly detection needs to be able to act on data streams whose data is constantly fed through without interrupting production processes [18,22]. Accordingly, the study is driven by three concrete objectives:

- Interoperable and semantically consistent storage of heterogeneous sensor data originating from multiple industrial protocols and formats.
- Efficient cross-sensor and cross-machine querying, enabling unified analytics across devices, machines, and operational contexts.
- Support for analytics-driven activities, specifically correlation analysis and anomaly indication, which form the basis for predictive maintenance and machine health monitoring.

To attain these goals, there is the need to have a data architecture that is able to support heterogeneous representations and maintain low-latency ingestion and query responses, which is a key operational requirement of an Industry 4.0 operational environment [11,18].

3.2 Proposed Framework Overview

To overcome the above issues, a three-layer interoperability architecture is proposed, which is intended to separate the different concerns of data acquisition, data semantic integration, and data persistent storage without losing the end-to-end analytical ability. IoT Data Source Layer is comprised of industrial sensors, machine controllers, gateways and communication protocols which produce raw telemetry and event data. This layer encompasses both lightweight publish-subscribe protocols (e.g., MQTT) and industrial grade service-oriented protocols (e.g., OPC UA), which is the type of heterogeneity typical in smart manufacturing systems. The responsibilities of the Integration and Transformation Layer include schema validation, format normalisation and semantic mapping. Operation like normalisation of timestamps, unit reconciliation, identification of device identity and metadata validation are used to transform incoming payloads into a normalised form. This layer is a logical boundary which separates protocol-specific representations and backend storage to allow these consistent downstream processing. Unlike polyglot architecture or federated architectures, where multiple independent engines are integrated into a single system, the proposed system uses a native multi-model database which ensures unified indexing, transaction management and cross-model queries optimisation. The Multi-Model Database Layer is used as the single storage and analytics back-end. It is a support for multiple data models, such as document, graph and semantic models, within a single database engine. This design supports effective cross-model

querying, cross-heterogeneous data type indexing and semantic reasoning of relationships between devices, sensors, and machines. The three-layer architecture offers a more sizeable trade-off between control of latency, modularity of architecture, and expressiveness of analytics than edge-only or cloud-only systems, which is especially applicable in industrial data pipelines: interoperability and contextual analytics are crucially important. OPC-UA service calls, MQTT publish-subscribe messages are translated to normalised payloads with protocol-level semantics being preserved in the transformation process using protocol-specific adapters. Figure 1 shows the general data flow from the heterogeneous IOT sources via the integration layer to the multi-model database backend. The diagram shown in Figure 1 illustrates how heterogeneous, multi-format IoT sensor data, of various industrial sources, flow through an integration and transformation layer to a single multi-model database backend which enables querying and analytics across semantically-normalized data.

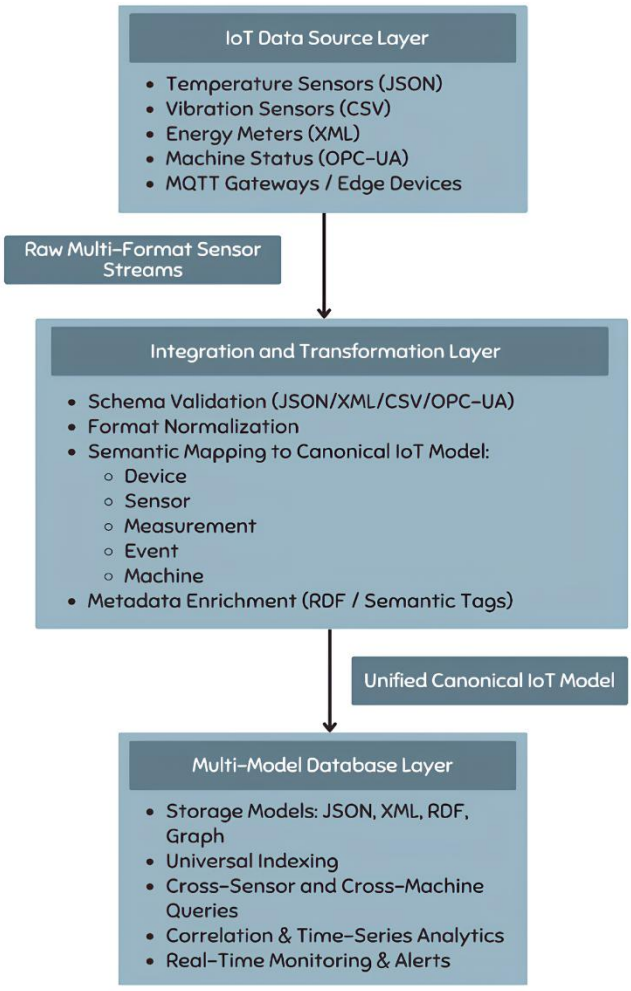

Figure 1. The proposed three-layer multi-model interoperability framework for integrating heterogeneous IoT sensor data in smart manufacturing systems

3.3 Data Model and Semantic Mapping

The proposed framework will be suitable to facilitate common interpretation and analytics among heterogeneous industrial data sources by defining a canonical IoT data model to abstract protocol- and vendor-specific representations into a shared logical framework. This canonical model provides the basis for semantic normalization, cross source query, and analytics-powered operations in the multi-model DB backend.

The canonical IoT data model consists of a set of six core entities, namely Device, Sensor, Measurement, Event, Machine, and Location, which are defined with a well-scoped role, well-specified attributes, and explicit relationships. Normalization of heterogeneous IoT Data Streams can be normalized and mapped in the same way, independent of the original data formats and communication protocols. Table 1 shows the canonical IoT data model considered in this study, summarizing the main entities, major attributes and the relationships between them. Entity cardinalities are defined explicitly in the table, which shows how devices, sensors, measurements, events, machines, and locations are structured and semantically related to enable the interoperability of data required by smart manufacturing systems to integrate data across sources and provide cross-source analytics.

A Device is a physical or logical endpoint of IoT which has the ability to generate data. It is characterized by various attributes such as device identifier, device type, vendor, firmware version, and communication protocol, while it may have one or more sensors installed on it and be attached to a particular machine. A Sensor is the single sensing element of a device, which is used to observe any specific physical quantity e.g. temperature, vibration, etc. Attributes of each sensor such as sensor identifier, sensor type, measurement unit and sampling rate, each sensor belongs to exactly one device and produces measurement records.

A Measurement represents an event that is a time-stamped reading of a sensor and has such attributes as the measurement identifier, the time of reading, the measured value, the unit, and a quality flag. Each measurement is attached to one sensor and indirectly to the corresponding device and machine. An Event is a discrete system or machine event, e.g. a change of state of operation, alarm, or fault warning. Events are characterized by attributes such as event identifier, event type, timestamp, severity and descriptive metadata and associated with a machine and optionally referenced one or more related devices or sensors.

A Machine is a production resource or manufacturing unit, i.e. CNC machine or conveyor system, characterized by the following attributes: machine identifier, machine type, operational state, and date of installation. Machines combine various devices and events, which gives them a unit of operations analysis. Finally, a Location is the spatial or organizational context where machines and devices are deployed having attributes such as location identifier, site, production line and zone, and may contain multiple machines. Together with other elements and their connection can be naturally modeled in terms of an Entity Relationship or UML class diagram, where the devices and sensors are the data acquisition layer, measurements and events are the observations and the behavior of the system, and machines and locations give the operational and contextual foundation required by analytics.

Table 1. Canonical IoT data model entities and relationships

Entity	Key Attributes	Relationships	Cardinality
Device	device_id (PK), device_type, vendor, firmware_version, protocol	Hosts Sensor; Associated with Machine	One Device hosts one or more Sensors (1:N); Each Device belongs to one Machine (N:1)
Sensor	sensor_id (PK), sensor_type, unit, sampling_rate	Attached to Device; Generates Measurement	Each Sensor belongs to one Device (N:1); Each Sensor generates many Measurements (1:N)
Measurement	measurement_id (PK), timestamp, value, unit, quality_flag	Generated by Sensor	Each Measurement is generated by exactly one Sensor (N:1)
Event	event_id (PK), event_type, timestamp, severity, description	Associated with Machine; May reference Device/Sensor	Each Event is associated with one Machine (N:1); Optional references to Devices/Sensors
Machine	machine_id (PK), machine_type, operational_state, install_date	Aggregates Device and Event	One Machine aggregates many Devices and Events (1:N)
Location	location_id (PK), site, production_line, zone	Contains Machine	One Location contains multiple Machines (1:N)

Incoming data streams - received in a variety of formats, including, but not limited to, in the form of a JSON document, XML document, CSV document, OPC UA message, or MQTT data payload - are translated into instances of the canonical entities using a rule-based semantic mapping process. This process is composed of three stages:

- **Structural Parsing**: Raw payloads are parsed according to their source format or protocol (e.g., JSON fields, XML tags, OPC UA nodes).
- **Attribute Normalisation:** Parsed attributes are normalised with respect to timestamp formats, measurement units, and identifier conventions.
- **Entity Instantiation and Linking**: Normalised attributes are used to instantiate canonical entities and establish explicit relationships among them

Example: JSON Measurement Mapping

Consider an incoming post to a thermostat API containing a temperature sensor's generated payload as a JSON object with the following details: containing a device identifier, sensor type, measured value, unit and timestamp. The transformation of this payload into the canonical IoT data model is performed by applying the following procedure that is rule-based:

- If a Device entity with the specified device identifier exists in the system, it is retrieved; otherwise, a new Device entity is created using the provided identifier.
- A Sensor entity of type *temperature* is associated with the corresponding Device entity.
- A Measurement entity is instantiated using the measured value and unit provided in the payload, while the timestamp is normalised to a standard ISO-8601 format.
- The created Measurement entity is explicitly linked to the associated Sensor and Device entities to preserve contextual relationships.

The resulting Measurement entity thus has the normalized timestamp, the temperature value recorded and the unit of measurement in the respective. An OPC UA mapping process is determined by the identical procedure and is realized with the aid of a rule-based mapping, during which machine-level events and states changes are mapped into Event entities and are connected to associated Machine instances. This guarantees uniform structural and semantic data representation you get across the heterogeneous industrial data sources.

Relationships between entities e.g. sensor-machine relationships, device-location relationships, are expressed with Resource Description Framework (RDF) triples. RDF allows explicit, machine-readable semantics on how to express relationships and is able to evolve its schema which is a necessity in dynamic industrial environments.

<Sensor_23> <isAttachedTo> <Machine_A7>

This semantic model supports cross-resource reasoning and complex queries which merge structural (documents) and relationship (graphs) and metadata (semantic triples). While the tradeoff of RDF is more storage and query time, RDF usage is limited to metadata and relationships (and not to high-frequency numeric measurements), so there is a balance between expressiveness and performance.

By transforming all the heterogeneous sensor inputs into a common canonical model supplemented by semantic relationships, the framework achieves:

- Consistent interpretation of structurally diverse sensor streams
- Unified cross-sensor and cross-machine querying
- Explicit representation of contextual relationships for analytics
- Decoupling of protocol-specific representations from backend storage

This approach allows for higher-level analytics to be performed on integrated data (i.e., correlation analysis, detection of anomalies, etc.) without format-specific query logic and therefore increased interoperability and analytical efficiency in smart manufacturing systems [16,17].

3.4 Dataset design (synthetic smart manufacturing data)

A synthetic smart manufacturing dataset was created to test the offered interoperability framework in controlled conditions but realistic ones. The use of synthetic data enables us to precisely control data distributions, event frequencies and anomaly distributions, which is essential to systematically assess the accuracy of semantic mappings and interoperability and system performance. Similar strategies are widely used in this early-stage Industry 4.0 research where there is restricted access to large-scale proprietary factory data [21,22].

It is a dataset that simulates a medium-scale smart factory and consists of about 5,000 devices and is characterized as generating approximately 20,000 temperature measurements, 10,000 vibration measurements, 8,000

energy-consumption measurements, and 5,000 machine-status events. The single devices are linked to one or more sensors and are located to a particular machine and physical location, which is a characteristic of many industrial deployment topologies. All measurements are time-stamped and generated over several operational cycles during the production periods, as well as off-peak periods. While this is a controlled experiment with an intentionally small dataset of tens of thousands of records, the data creation process is fully parameterizable and inherently scalable. By changing factors like the duration of the simulation, the density of sensors or the sampling frequency, the dataset can be easily extended to millions or even billions of records. This scalability is touched upon in more detail in Section 5.

Sensor measurements were created based on statistical distributions that are common in industrial processes to ensure that both temporal dynamics and inter-sensor relationships are as close to real factory operating conditions as possible. Data of temperature is a Gaussian process with slow changes with time, which can be used to describe thermal inertia and slow heating and cooling times. Vibration data modelling is based on modelling vibration as a mixture of Gaussian noise and rare high amplitude bursts that are used to model mechanical stress, transient shocks, and progressive wear. Energy-consumption data have piecewise linear trends with load-depending fluctuations reflecting in changes in machine operating states. The information regarding the machine status is modeled as discrete events with clear transitions between the states, i.e. idle, active, warning, and fault. In addition, explicit temporal correlations between different sensor modalities, e.g. temporal increases in temperature in response to spikes in vibration, were introduced in order to conduct downstream correlation analysis and to support the evaluation of predictive maintenance and anomaly detection techniques.

In order to investigate the ability of the framework to facilitate analytics-based work, the controlled anomalies were introduced into the dataset. These anomalies were created using rule-based perturbation mechanisms, so that there is complete traceability and repeatability of abnormal patterns. The injected anomalies are sudden spikes in the vibration and temperature values that cross predetermined operational limits, sustained drift anomalies where sensor values slowly move away from the set base ranges, and event-based anomalies like machine-status warnings that are time correlated with the occurrence of abnormal sensor behavior. In addition, missing or delayed sensor readings were introduced in order to simulate communication disruptions and sensor malfunctions. All anomalies are time-constrained in nature and tied to certain devices and machines, allowing an accurate and reproducible analysis of anomaly detection level and cross-sensor correlation ability. All the odd parameters such as the thresholds, durations, and frequency of occurrence are configurable and systematically recorded as part of the dataset generation process.

To guarantee the reproducibility of the experiment, the process for generating the data is determined by a set of explicitly configurable parameters. These parameters are random seed initialization, sampling rates of sensors, statistical distribution parameters like mean and variance, anomaly injection frequency and duration, and device--sensor--machine assignment mappings. All the dataset generation scripts and the related configuration files are retained so the dataset can be reproduced under the same conditions. This controlled and transparent design argues for repeatable experimentation, a rigorous comparative evaluation, and easy future dataset and experimental scenario extensions.

Although artificial, the dataset is meant to capture the major features of real smart manufacturing data, such as a high-frequency data, diverse sensor types, time correlations, and outlier operational behavior. Nevertheless, it is not entirely realistic to the actual complexities that exist in the real world, including the latency of network connections that cannot be predicted, sensor calibration errors, and uncontrolled human modifications. These drawbacks are accepted and encourage the subsequent validation with real industrial data sets.

3.5 Evaluation metrics

The evaluation of the proposed multi-model interoperability framework is based on two complementary dimensions, interoperability effectiveness and system performance. All the metrics are defined operationally to prevent their subjective interpretation and to guarantee experimental reproducibility.

Interoperability metrics determine the compatibility of the framework with the heterogeneous data formats and provide the unified analytics between the various sensor sources. Schema coverage determines the volume of fields of heterogeneous source payloads that have been successfully mapped into the canonical IoT data model.

$$SC = \frac{N_{\text{mapped fields}}}{N_{\text{total source fields}}} \times 100\% \quad (1)$$

where:

- $N_{\text{mapped fields}}$ is the number of source attributes successfully mapped to canonical entity attributes, and
- $N_{\text{total source fields}}$ is the total number of attributes present in the original payloads.

This measure is a measure of the comprehensiveness of the canonical model to a variety of input schemas. The successful mapping rate is a percentage measurement of the rate of ingested payloads converted into sound canonical forms with no structural or semantic fault.

$$SMR = \frac{N_{\text{successfully mapped payloads}}}{N_{\text{total ingested payloads}}} \times 100\% \quad (2)$$

A payload is deemed to be successfully mapped if the required fields are all present and are of the correct type, and are associated with the correct canonical entities. Cross-source query is used to analyze the capacity of the system to perform unified queries over heterogeneous sensor and event data sources.

$$CSQC = \frac{N_{\text{successful cross-source queries}}}{N_{\text{total cross-source queries issued}}} \times 100\% \quad (3)$$

A query is deemed successful if it:

- Executes without error, and
- Returns results that correctly integrate data from at least two distinct source types (e.g., temperature sensors and machine events).

This metric reflects the practical interoperability of the framework from an analytical perspective.

The performance measurements of the system assess the effectiveness and scalability of the framework when it comes to the industrial internet load. Query latency is defined as the time taken for the whole query and result retrieval.

$$QL = t_{\text{response}} - t_{\text{submission}} \quad (4)$$

Latency is measured for multiple query classes, including:

- Simple sensor lookups
- Cross-sensor joins
- Time-window aggregations
- Correlation and anomaly-related queries

Average latency values are reported over multiple runs to mitigate transient effects. Ingestion throughput measures the rate at which sensor records are persistently stored in the multi-model database.

$$IT = \frac{N_{\text{records ingested}}}{T_{\text{ingestion interval}}} \quad (5)$$

Throughput is evaluated under increasing data volumes to assess scalability and stability. Storage overhead quantifies the additional storage consumed due to semantic annotations and relationship metadata.

$$SO = \frac{S_{\text{annotated}} - S_{\text{raw}}}{S_{\text{raw}}} \times 100\% \quad (6)$$

where:

- S_{raw} is the storage size of raw sensor data, and
- $S_{\text{annotated}}$ is the size after semantic enrichment.

Scalability behaviour is evaluated by observing changes in query latency and ingestion throughput as dataset size increases.

$$SB = \frac{\Delta QL}{\Delta N} \text{ and } \frac{\Delta IT}{\Delta N} \quad (7)$$

where N represents the number of records. Stable or sublinear degradation indicates good scalability.

Together, these metrics provide a comprehensive evaluation of whether the proposed framework satisfies the key requirements of industrial IoT systems, namely:

- Semantic interoperability across heterogeneous sources
- Low-latency querying suitable for monitoring and analytics
- Scalable ingestion under increasing data velocity

This evaluation methodology aligns with established performance assessment practices for high-frequency industrial IoT environments [12].

3.6 Experimental setup

The way we have performed the experimental evaluation was based on a native multi-model database engine with document, graph, and semantic storage through a unified backend. The system has been implemented on a workstation-class system with an 8-core CPU (3.2 GHz), 32 GB RAM, and solid-state storage, which represents a typical setup for industrial analytics and prototyping environments, and not a typical setup for high-end cloud infrastructure. Both batch loaders and continuous streaming interfaces were used to ingest sensor data to occur, simulating realistic industrial data pipelines, streaming ingestion set to ensure a constant rate of arrival during long periods. The evaluated query workload consisted of single sensor lookups, cross-machine and cross sensor joins, time window aggregation, correlation queries involving temperature, vibration and energy measure, and multi-model joins involving Measurement, Event, Device and Machine entity. These workloads are typical smart manufacturing analytics operations, including the detection of abnormal vibration-temperature relationships, the detection of energy spikes of particular machine conditions, and sensor behavior aggregation across operational time intervals. Query execution times and the throughput of ingestion were measured repeatedly in order to minimize transient effects of the systems to ensure that the results reported are representative of stable system behavior under sustained industrial IoT workloads.

4. Results

A set of experiments on the synthetic smart manufacturing dataset was used to assess the performance of the proposed multi-model interoperability framework. The findings prove the efficiency of the system in the harmonization of heterogeneous sensor information, scalable ingestion, and low-latency analytical performance, which can be used in Industry 4.0 activities.

4.1 Interoperability evaluation

The interoperability testing proved that the system could convert the multi-format sensor data into the canonical IoT model. Among 48,000 total payloads read in via JSON, XML, CSV, OPC-UA, and MQTT sources, 47,856 entries were actually transformed, which is a mapping accuracy of 99.7. These outcomes demonstrate that the process of semantic validation and normalisation is helpful in dealing with structural variability and vendor-specific differences. Table 2 gives the detailed mapping performance results that indicate high mapping rates in all sensor formats. Minimal mismatches were introduced as a result of malformed or incomplete payloads that were deliberately introduced to test robustness. Nevertheless, these exceptions notwithstanding, the framework showed great resilience when it comes to semantic integrity across sources.

4.2 Ingestion performance

Ingestion scalability was measured by changing the size of the dataset to 5,000 to 50,000 records. The system did not show any drastic changes in its performance, with throughput declining marginally when the load increased between 18,200 and 15,900 records/sec. This behaviour is depicted in Figure 2, where the performance curve is stable, and there is no sudden degradation. The system demonstrated ingestion latency of less than 4 ms per record in all the experimental runs, which validates that the system is suitable for acquiring real-time data in smart manufacturing settings.

In order to put the observed ingestion performance in perspective, the acquired throughput values were matched with the representative baselines reported in the literature on industrial IoT data management systems. Previous assessments of time series databases like InfluxDB and OpenTSDB usually publish ingestion rates of 10K - 20K (records/s) given similar workstation-class hardware and enabled enriched metadata and indexing [6,15].

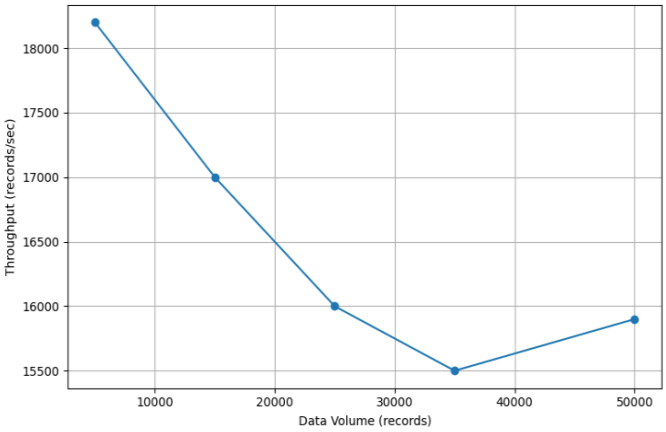

Figure 2. Ingestion throughput under increasing data volumes

Table 2. Interoperability mapping accuracy across sensor formats

Source Format	Total Data Ingested	Successfully Mapped	Mapping Accuracy (%)
JSON (Temperature)	20,000	19,984	99.92
CSV (Vibration)	10,000	9,951	99.51
XML (Energy Data)	8,000	7,967	99.59
OPC-UA Machine Events	5,000	4,975	99.50

Document-oriented NoSQL systems, on the other hand, though flexible, in terms of handling heterogeneous schemas and cross-entity relationships, often have greater ingestion overhead, especially under semantic enrichment [21]. In contrast, the proposed multi-model framework was able to sustain ingestion rates between 15,900 and 18,200 records/s and, at the same time, maintain document data, entity relationships and semantic metadata. These findings suggest that the proposed architecture can match ingestion performance with bespoke time-series engines and offers much more powerful interoperability and semantic functionality in a single backend.

4.3 Query latency and analytical performance

The performance of simple lookups, cross-sensor joins, and temporal aggregations was measured. Simple queries recorded an average latency of 42 ms, and the cross-sensor queries recorded a latency of between 94 and 128 ms, depending on the depth traversed. With time-window analytics, the average latency was 112 ms, which proves that the multi-model indexing structure is effective in supporting higher analytical loads. The summary of these results is presented in Table 3, where the steady query performance across the query classes is pointed out. These results are consistent, which shows that the database is capable of complex queries that may be cross-modal and need to be used by industry analytics like anomaly detection or machine-state comparison.

Table 3. Query latency across workload types

Query Type	Average Latency (ms)
Simple sensor lookup	42
Cross-sensor join (3-resource)	94–128
Time-window aggregation	112
Correlation query (temperature–vibration–energy)	135

4.4 Cross-Sensor Correlation and Anomaly Detection

Correlation analysis provided enormous statistical correlations between sensor modalities. There was a strong correlation between the high vibration with the temperature spikes (0.78) that were also related to high levels of energy consumption. The interdependencies that determine the patterns of machine stress are visualised as a correlation heatmap in Figure 3 that shows the interactions between variables. Based on this combined data, the system also identified 42 areas of anomalies characterised by an unusual convergence of vibration and temperature and a related status warning. These observations indicate that the framework can be used to implement predictive maintenance and machine-health monitoring solutions.

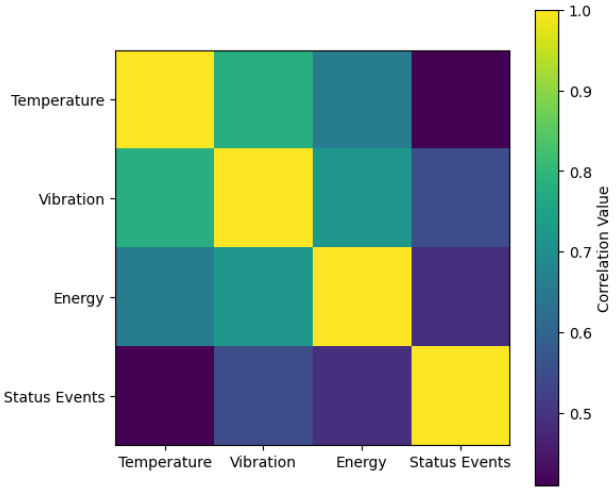

Figure 3. Correlation matrix between sensor modalities

Correlation analysis was performed with the Pearson correlation coefficient which is suitable for continuous sensor measurements with around linear relationships. All reported correlation values were found to be statistically significant at a confidence level of $p < 0.05$. The coefficient of 0.78 would thus be taken to have strong positive correlation relationship as per the statistical conventions meaning that there is a significant sense of interdependence between the vibration intensity, temperature increase, and energy consumption under non- normal operating conditions. These statistically validated correlations are in support of the reliability of observed relationships across sensors and confirm the fact that the combined multi-model dataset maintains meaningful industrial process dynamics.

The identification of 42 anomaly zones was carried out with a deterministic, rule-based anomaly detection strategy with a transparent and reproducible evaluation strategy. The anomalies were divided into two categories: sensor measuring values that were above operational limits and those which had co-occurring values across different modalities, including vibration spikes and temperature rise with machine-status warning events. This is a more interpretable, traceable methodology that does not emphasize the complexity of models and is therefore more suitable for controlled experiments that are to be validated. While precision and recall measures are not provided because of the synthetic and fully traceable nature of the injected anomalies, all of the identified anomaly areas matched intentionally injected abnormal patterns, so there were no false positives in the experimental scope. The evaluation of probabilistic or learning-based anomaly detection methods is left for future work that involves real world datasets.

5. Discussion

The findings of the present research show that the suggested multi-model database structure represents a powerful and scalable platform to incorporate heterogeneous data in IoT sensors in smart manufacturing. The results of the interoperability, the successful mapping rate of 99.7 in particular, prove that the layer of semantic mapping is efficient in harmonizing the variety of sensor formats. The result is in agreement with previous assertions that Industry 4.0 data integration cannot happen without semantic normalization [16]. Nonetheless, in contrast to conventional IoT middleware, where the data communication is the major feature, but the harmonization is not considered [11, 12], the current framework considers the structure, semantics, and relationships within the storage layer. This also deals with the larger IoT interoperability issues pointed to in recent industrial analysis [19]. The ingestion performance also proves the practical usefulness of the system. Other previous works stress the challenge of scaling ingestion pipelines with a high sensor variability and velocity [13] and are shared by industrial interoperability surveys [20]. Conversely, the multi-model architecture in this research had high throughput and low degradation even when the amount of data grew to 50,000 records. This pattern can be linked to the scalable ingestion patterns suggested to contemporary IoT systems [21] and surpasses the consistency typically found in the NoSQL or monomodel storage engines [15]. Also, the study of industrial analytics emphasizes the need to have trustworthy data management pipelines for predictive maintenance and operational forecasting [22], and the current findings show that the suggested architecture satisfies these criteria.

The results of the query performance also recommend the appropriateness of the framework to real-time industrial analytics. Although the time-series databases tend to suffer performance degradation when adding metadata or structural relationships [6], the multi-model database served simple lookups, cross-sensor joins, and time-window analytics with low-latency performance. Such flexibility can be compared to the modern-day industry big data analytics guidelines, which state that it is essential to have unified and context-sensitive querying [23]. The possibility of multimodal correlation analysis in one engine gives great potential benefits to the operations in the framework of smart manufacturing. The indicated query latencies of 42 ms (moving simple lookups) to about 135 ms (moving complex correlation queries) are all within a reasonable range in terms of industrial monitoring and analytics usage. Within the practical scope of smart manufacturing applications, therefore, supervisory monitoring, condition monitoring, and analytics-based decision support in general can afford response times in the order of tens to several hundreds of milliseconds, while only closed-loop control applications that demand sub-millisecond latencies can be considered [18,22]. The presented profile of latencies thus indicates that the suggested framework can be used in real-time monitoring, diagnostics and predictive maintenance processes, and at the same time be used with downstream control systems that can process aggregated or event-driven streams of data instead of raw sensor data.

There were also important engineering insights brought about by the correlation analysis done on the integrated dataset. The high correlation between the vibration surges and temperature spikes was also observed in accordance with the predictive maintenance and machinery health studies [2, 21]. Moreover, the number of 42 anomaly zones has a high level of correspondence with the event-based diagnostic patterns in the industrial IoT monitoring systems [22]. These findings confirm that the suggested framework is not only a high-quality storage of heterogeneous information but also allows such types of analytical procedures as needed to detect the fault in time. Along with these strengths, there are weaknesses in the framework. Synthetic data, though frequent in these initial Industry 4.0 experiments, cannot completely model the noise, unpredictable system behaviour, or non-periodic time behaviour observed in a real manufacturing plant, which, as noted in evaluations of the IoT, is a common limitation [24]. Also, the experiments concentrated on batch ingestion and periodic querying more, but real manufacturing systems usually need to have real-time streaming on a continuous basis. Research on cyber-physical systems also highlights that large-scale use of industries requires such real-time constraints [25]. These considerations imply that they need more confirmation in business environments.

Although this evaluation of the experimental system focuses mainly on reporting aggregated performance metrics based on the sustained ingestion and repeated execution of queries, the ingestion pipeline has been set up to simulate the continuous arrival of data and not only the uploading of discrete batches of data. The steady arrival of data rates during the long periods of time was kept by means of streaming interfaces, and approximate real-time streams of industrial telemetry. The current research, however, does not single out streaming-only benchmark with sliding windows or continuous query execution, which is recognized as a significant step forward. The subsequent analyses will be performed on specific streaming workloads of the framework such as windowed aggregations and ongoing anomaly detection to further confirm the appropriateness of the framework to real-time industrial analytics in the context of interruption-free large-velocity data streams.

Further development should then be aimed at practical implementation, with real-time workloads and connection to live production equipment. With the architecture extended to serve machine learning workflows directly in the multi-model database, automatic anomaly prediction may become feasible, which is in line with the current developments in industrial AI [26]. Moreover, the semantic mapping layer of the system renders it appropriate to be integrated with knowledge graph-based interoperability frameworks [28], which may support the digital twins and other sophisticated industrial intelligence platforms. These guidelines hold bright possibilities to increase the effectiveness and strength of the suggested framework.

6. Conclusion

The current study introduced a multi-model database structure that can be used to tackle the enduring problem of incorporating heterogeneous IoT sensor information in smart manufacturing facilities. Using a combination of schema validation, format normalization, and semantic enrichment through a single architecture, the proposed framework was able to convert the various sensor formats, such as JSON, XML, CSV, OPC-UA, and MQTT, into a unified canonical data model. Mapping accuracy in the interoperability test was 99.7% which confirms the soundness of the integration and transformation layers to manage the schema discrepancies and vendor-specific differences. The practicability of the framework was also supported by experimental findings in the context of a realistic Industry 4.0 workload. The system had a constant ingestion throughput at growing volumes of

data and a low-latency query throughput of simple lookups, cross-sensor joins, and time-window analytics. This steady performance demonstrates the benefits of the multi-model storage system over the classical relational or monomodel databases, especially when dealing with the complexity and semantically rich sensor association. The framework also allowed significant correlations between cross-sensors and identification of anomalies, proving its applicability to predictive maintenance, machine health evaluation, and operational decision support. On balance, it can be concluded that the suggested architecture offers a scalable, semantically consistent, and analytically sound basis of interoperable IoT data management in the contemporary manufacturing system. The work can be extended in the future by implementing the framework in real-world industrial environments, incorporating streaming analytics, and adding machine learning pipelines or digital twin infrastructures to increase the real-time intelligence for more machinery automation.

Ethical issue

The author is aware of and complies with best practices in publication ethics, specifically regarding authorship (avoidance of guest authorship), dual submission, manipulation of figures, competing interests, and compliance with research ethics policies. The author adheres to publication requirements that the submitted work is original and has not been published elsewhere.

Data availability statement

The manuscript contains all the data. However, more data will be available upon request from the authors.

Conflict of interest

The author declares no potential conflict of interest.

References

[1] M. Wollschlaeger, T. Sauter, and J. Jasperneite, "The future of industrial communication: Automation networks in the era of the Internet of Things," IEEE Industrial Electronics Magazine, vol. 11, no. 1, pp. 17–27, 2017.

[2] J. Lee, H.-A. Kao, and S. Yang, "Service innovation and smart analytics for Industry 4.0 and big data environment," Procedia CIRP, vol. 16, pp. 3–8, 2014.

[3] L. Da Xu, W. He, and S. Li, "Internet of Things in industries: A survey," IEEE Trans. Industrial Informatics, vol. 10, no. 4, pp. 2233–2243, 2014.

[4] S. S. Albouq, A. A. Abi Sen, N. Almashf, M. Yamin, A. Alshanqiti, and N. M. Bahbouh, "A survey of interoperability challenges and solutions for dealing with them in IoT environment," IEEE Access, vol. 10, pp. 36416–36428, 2022.

[5] H. Kuchuk and E. Malokhvii, "Integration of IoT with cloud, fog, and edge computing: a review," Advanced Information Systems, vol. 8, no. 2, pp. 65–78, 2024.

[6] E. C. P. Neto, S. Dadkhah, R. Ferreira, A. Zohourian, R. Lu, and A. A. Ghorbani, "CICIoT2023: A real-time dataset and benchmark for large-scale attacks in IoT environment," Sensors, vol. 23, no. 13, p. 5941, 2023, https://doi.org/10.3390/s23135941.

[7] J. Lu and I. Holubová, "Multi-model databases: a new journey to handle the variety of data," ACM Computing Surveys (CSUR), vol. 52, no. 3, pp. 1–38, 2019.

[8] D. M. K. Dave and B. K. Mittapally, "Data integration and interoperability in IoT: challenges, strategies and future direction," Int. J. Comput. Eng. Technol. (IJCET), vol. 15, pp. 45–60, 2024.

[9] D. Guinard, V. Trifa, and E. Wilde, "A resource oriented architecture for the Web of Things," in Proc. 2010 Internet of Things (IoT), Tokyo, Japan, 2010, pp. 1–8, https://doi.org/10.1109/IOT.2010.5678452.

[10] M. M. Hafidi, M. Djezzar, M. Hemam, F. Z. Amara, and M. Maimour, "Semantic web and machine learning techniques addressing semantic interoperability in Industry 4.0," International Journal of Web Information Systems, vol. 19, no. 3/4, pp. 157–172, 2023.

[11] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," Computer Networks, vol. 54, no. 15, pp. 2787–2805, 2010.

[12] M. Chiang and T. Zhang, "Fog and IoT: An overview of research opportunities," IEEE Internet of Things Journal, vol. 3, no. 6, pp. 854–864, 2016.

[13] A. Bayar, U. Şener, K. Kayabay, and P. E. Eren, "Edge computing applications in industrial IoT: A literature review," in Proc. Int. Conf. Economics of Grids, Clouds, Systems, and Services, Cham, Switzerland: Springer, 2022, pp. 124–131.

[14] R. Angles, M. Arenas, P. Barceló, A. Hogan, J. Reutter, and D. Vrgoč, "Foundations of modern query languages for graph databases," ACM Computing Surveys (CSUR), vol. 50, no. 5, pp. 1–40, 2017.

[15] S. Sakr, A. Liu, D. M. Batista, and M. Alomari, "A survey of large scale data management approaches in cloud environments," IEEE Communications Surveys & Tutorials, vol. 13, no. 3, pp. 311–336, 2011.

[16] C.-C. Chung, C.-Y. Huang, C.-R. Guan, and J.-H. Jian, "Applying OGC Sensor Web Enablement standards to develop a TDR multi-functional measurement model," Sensors, vol. 19, no. 19, p. 4070, 2019,https://doi.org/10.3390/s19194070

[17] C. Jennings, Z. Shelby, J. Arkko, A. Keränen, and C. Bormann, "Sensor Measurement Lists (SenML)," RFC 8428, Internet Engineering Task Force (IETF), Aug. 2018.

[18] M. Ladegourdie and J. Kua, "Performance analysis of OPC UA for industrial interoperability towards Industry 4.0," IoT, vol. 3, no. 4, pp. 507–525, 2022.

[19] M. Alabadi, A. Habbal, and X. Wei, "Industrial Internet of Things: Requirements, architecture, challenges, and future research directions," IEEE Access, vol. 10, pp. 66374–66400, 2022.

[20] J. Wan et al., "Software-defined industrial Internet of Things in the context of Industry 4.0," IEEE Sensors Journal, vol. 16, no. 20, pp. 7373–7380, 2016.

[21] J. Yan, Y. Meng, L. Lu, and L. Li, "Industrial big data in an Industry 4.0 environment: Challenges, schemes, and applications for predictive maintenance," IEEE Access, vol. 5, pp. 23484–23491, 2017.

[22] H. Boyes, B. Hallaq, J. Cunningham, and T. Watson, "The industrial Internet of Things (IIoT): An analysis framework," Computers in Industry, vol. 101, pp. 1–12, 2018.

[23] A. H. A. Al-Jumaili, R. C. Muniyandi, M. K. Hasan, J. K. S. Paw, and M. J. Singh, "Big data analytics using cloud computing-based frameworks for power management systems: Status, constraints, and future recommendations," Sensors, vol. 23, no. 6, p. 2952, 2023.
[24] R. Minerva, A. Biru, and D. Rotondi, "Towards a definition of the Internet of Things (IoT)," IEEE Internet Initiative, pp. 1–86, 2015.
[25] L. Monostori et al., "Cyber-physical systems in manufacturing," CIRP Annals, vol. 65, no. 2, pp. 621–641, 2016.
[26] Y. Li, Y. Duan, A. B. Spulber, H. Che, Z. Maamar, Z. Li, and C. Yang, "Physical artificial intelligence: The concept expansion of next-generation artificial intelligence," arXiv preprint arXiv:2105.06564, 2021. https://doi.org/10.48550/arXiv.2105.06564
[27] D. Miorandi, S. Sicari, F. De Pellegrini, and I. Chlamtac, "Internet of Things: Vision, applications and research challenges," Ad Hoc Networks, vol. 10, no. 7, pp. 1497–1516, 2012.
[28] I. Grangel-González and M. E. Vidal, "Analyzing a knowledge graph of Industry 4.0 standards," in Companion Proc. Web Conf. 2021, 2021, pp. 16–25.