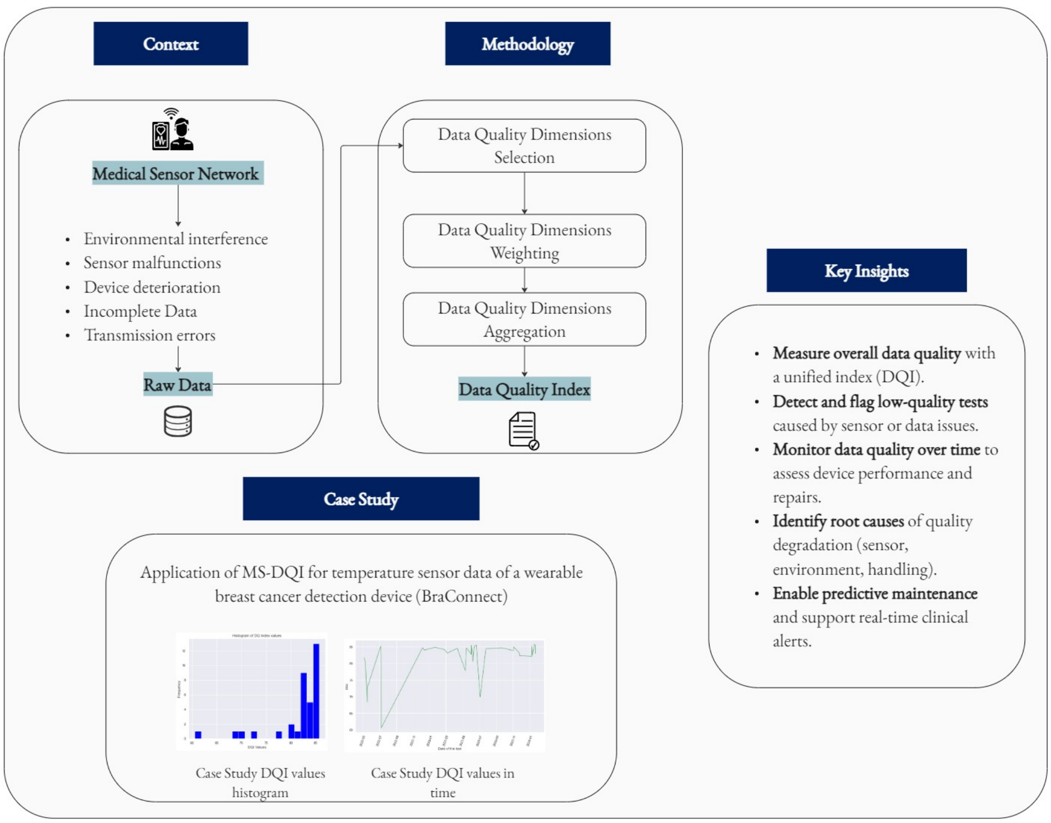
Graphical Abstract

MS-DQI: A Methodology for Data Quality Assessment in Medical Sensor Networks with a Case Study on a Temperature Sensor Network for Breast Cancer Detection

KETFI Raniya, AL MASRY Zeina, ZERHOUNI Noureddine, DEVALLAND Christine



Highlights

MS-DQI: A Methodology for Data Quality Assessment in Medical Sensor Networks with a Case Study on a Temperature Sensor Network for Breast Cancer Detection

KETFI Raniya, AL MASRY Zeina, ZERHOUNI Noureddine, DEVALLAND Christine

- Brief review of data quality assessment in medical data: We review existing methodologies in medical data quality assessment highlighting the gaps and the specific need for more generalizable assessment methodologies per data type

- Classification of common data quality issues in sensor networks

- Introduction of MS-DQI: We propose a novel methodology named MS-DQI for assessing data quality in medical sensor networks through predefined steps: Data Quality dimension selection, Data quality dimension weighting and dimension aggregation

- Scalable and Generalizable: The methodology can be adapted to various sensor types and medical applications, making it a flexible tool for enhancing data quality in health monitoring systems

- Case study to demonstrate the applicability of MS-DQI through a real-world study using temperature sensor data from wearable devices for breast cancer detection

MS-DQI: A Methodology for Data Quality Assessment in Medical Sensor Networks with a Case Study on a Temperature Sensor Network for Breast Cancer Detection

KETFI Raniya, AL MASRY Zeina, ZERHOUNI Noureddine

Université Marie et Louis Pasteur, SUPMICROTECH, CNRS, institut FEMTO-ST, F-25000 Besancon, France

DEVALLAND Christine

HOPITAL NORD FRANCHE-COMTE, Service d'anatomie et cytologie pathologiques, 100 Route de Moval, Trevenans, 90400, France

---

Abstract

The increasing use of medical sensors and wearable devices has created many opportunities for personalized health monitoring. However, ensuring the quality of data from sensor networks remains a challenge due to the variability in sensor performance and environmental factors. This paper introduces a comprehensive methodology for assessing data quality in medical sensor networks that comprises three major steps: first, defining a quantitative Data Quality Index (DQI) based on dimensions such as accuracy, completeness, availability, redundancy, precision, consistency and timeliness. Second, assigning weights to each dimension to reflect their relative importance. Last, aggregating these weighted dimensions into a single DQI. The proposed methodology is applied to a case study using temperature sensor data from a wearable breast cancer detection device.

The insights gained from this application include the evaluation of the dataset's data quality, the detection of defective tests with alarming data quality issues and analysis of the data quality evolution over time, highlighting the impact of device repair and changes. Additionally, the DQI helps identify the underlying causes of data quality issues, such as sensor malfunction, or environmental factors, and it can assist in predicting sensor or device failure. These insights highlight the utility of the proposed MS-DQI: Medical Sensors-Data Quality Index methodology in improving the reliability and performance of medical sensor network data.

Keywords:
Medical Sensor Network, Data Quality Assessment, Data Quality Index, Breast Cancer Detection, Temperature Sensors, Wearable Medical Devices, Sensor Data Reliability, Medical IoT

1. Introduction

In data-driven applications, a significant portion of the development process is dedicated to data preparation [1]. This preparation aims to increase the quality of the data; without this step, the insights derived even from advanced algorithms lack the necessary reliability. The International Organization for Standardization (ISO) defines data quality as "The degree to which a set of inherent characteristics of data fulfills requirements" [2]. This ISO definition emphasizes that data quality is determined by how well data meets certain defined conditions or objectives. In the medical field, data quality is crucial because accurate and reliable data directly impact patient care, clinical decisions and research outcomes. With the growing use of electronic health records (EHRs), digital medical imaging, genomics and data-driven healthcare applications, maintaining high data quality is crucial to ensure patient safety and improve clinical outcomes. On the other hand, medical wearable devices with sensors have emerged and revolutionized healthcare by bringing data-driven insights and personalized monitoring to patients' daily lives. Such wearable devices integrate advanced sensor technologies and data analytics to provide real-time health information and enable early detection of medical conditions [3]. The resulting data insights have the potential to transform healthcare delivery and improve patient outcomes; therefore, implementing data quality assurance measures is essential for validating and effectively using these devices.

One approach to managing data quality is to assess its level which involves analyzing various aspects of data, including its content, structure and context, to identify any issues or discrepancies that may affect its reliability or usability. Data quality assessment can be done by characterizing the data with various quality dimensions, which can be defined as specific attributes or characteristics used to measure the quality of data [4]. These dimensions collectively provide a comprehensive framework for evaluating various aspects of data accuracy, reliability, completeness and relevance. Commonly recognized data quality dimensions include Accuracy, Completeness, Consistency, Timeliness, Uniqueness and Accessibility. The definition of a dimension can vary depending on the context of use, for example, in a sensor network specific context, completeness can be defined in a spatial or temporal way. Spatial completeness refers to the extent to which the sensor network covers the entire area of interest, while temporal completeness refers to the extent to which the sensor network consistently records data over time without any gaps or interruptions. The work of [5] is one of the most cited works of data quality where authors see that data quality focuses on ensuring that data meets the needs and expectations of its users, is fit for its intended purpose and supports effective decision-making and business processes. They proposed an empirical framework to capture the aspects of data quality that are important to data consumers which resulted in four DQ (Data Quality) categories: 1-intrinsic DQ that consists of accuracy, objectivity, believability and reputation; 2-contextual DQ that consists of value-added, relevancy, timeliness, completeness and appropriate amount of data; 3-representational DQ consists of interpretability, ease of understanding, representational consistency and concise representation and 4- accessibility DQ consists of accessibility and access security. This work was the inspiration of many succeeding works that proposed their data quality frameworks based on the proposed dimensions. Other frameworks perceive data quality as a manifestation of different anomalies where they categorize data based on the occurrence of data quality problems [6].

The main contribution of this paper is proposing a methodology for data quality assessment by introducing a quantitative Data Quality Index (DQI) specifically designed for data collected from medical sensor networks addressing common challenges encountered in medical sensor data. The paper is structured as follows: Section 2 reviews the existing methodologies for assessing medical data quality and highlights their limitations, Section 3 details the possible data quality issues in medical sensor data. Section 4 presents the proposed methodology, detailing each step from the selection of data quality dimensions to the aggregation of the index. Finally, Section 5 applies the proposed methodology to a private dataset acquired from a wearable device equipped with thermal sensors for breast cancer detection, Section 6 and Section 7 conclude the paper by discussing key findings, limitations and perspectives for future work.

2. Medical Data Quality Assessment

Most data quality assessment work in the medical field focuses on EHRs (Electronic Health Records) which are digital versions of patients' paper medical records. The authors in [7] detailed the methods and dimensions used in EHR data quality assessment to enable their reuse for clinical research which could reduce the costs and inefficiencies associated with clinical research. In this work, the authors identified the most commonly used DQ dimensions as: completeness, correctness, concordance, plausibility and currency and seven general categories of data quality assessment methods: comparison with gold standards, data element agreement, data source agreement, distribution comparison, validity checks, log review and element presence. They also highlighted that there is a remarkable lack of uniformity or generalizability in the approaches employed to evaluate EHRs data quality. They continued their work in [8], proposing a guideline for EHR data quality assessment focusing on three data quality dimensions: completeness, correctness and currency of data. These dimensions are operationalized according to the three primary variables of EHR data: patients, variables and time. Each of the nine operationalized dimensions corresponds to a methodological recommendation for EHR data quality assessment. The framework was positively evaluated by experts in EHRs, although the authors noted that some of the experts expressed the need for more clarity on how to use the framework and to determine which operationalized dimensions and recommendations are relevant for a given context of use.

In the work of [9], a 6-step standardized process to assess data quality in a longitudinal data repositories in Transport Data Mart (TDM) is proposed to support comprehensive outcomes research efforts that aggregate patient data across the entire episode of care for a patient who is transported from one hospital to another. The approach included the following steps: 1- preliminary analysis: to check whether all appropriate records are included in the TDM, 2- documentation–longitudinal concordance: to measure longitudinal data agreement in the TDM, 3-breadth: to assess the presence of the following records for every patient hospital day for each encounter: vital signs, laboratory results, procedures and medications, 4- data element presence: by measuring the presence of a diagnosis, set of vital signs and lab panels for each hospital day, 5- density: to assess the number of records that had prior records available leading up to the incident hospitalization for the transferred patient and 6- prediction: considered as the ability to reuse EMR (Electronic Medical Record) data to predict something or find an association. The authors applied their framework on a TDM dataset from Cleveland Clinic Health System dataset for EMRs related to patient transfer requests and found the approach effective in establishing the meta-data for a longitudinal data repository that can be reused by other studies to answer different research questions.

The authors in [6] introduced a framework to categorize possible data quality problems in EHRs of emergency departments and described data quality assessment techniques for the identified problems. This work was inspired by [10]'s classification of data problems where data quality problems are split into missing data and non-missing data. Not missing data is further divided into wrong data and not incorrect but not directly usable data. They developed an R package "DAQAPO" that contains a set of generic functions to assess the quality of EHRs data, such as missing_values, incomplete_cases, inactive_periods, attribute_dependency and duration_outliers, these functions standardize and facilitate the process of identifying and quantifying potential data quality problems and the package is tested on a real-life dataset extracted from the EHR of the Emergency Department of a Belgian university hospital. While the developed data quality framework is a standardized data quality assessment framework for EHR data, it may limit the generalizability of the findings to other healthcare data or simulation studies.

In the work of [11], authors proposed a rule-based approach to data error identification in EHRs. The authors gathered over 6,000 data quality rules from different sources such as Observational Health Data Sciences and Informatics (OHDSI), the Healthcare Systems Research Network rules and the Sentinel data checking rules. These logical rules are categorized into twenty-two rule templates, providing a scalable framework for organizing and sharing data quality rules. The rule templates are further classified into five higher-level types: incompatibility, value out of range, temporal sequence error, incompleteness and duplication. This work illustrated that rule based data quality assessment can be applied in healthcare facilities to identify data errors. Nevertheless, assessing medical data quality based only on rules may not be sufficient to capture all types of data errors.

Besides work on EHR data, the work of [12] contributes to enhancing data quality in the field of Medical IoT (Internet of Things) by proposing a mechanism that automatically extracts high-quality data from heterogeneous IoT medical devices and transforms it into a common format for interoperability. The authors developed a data quality estimation index that is a mathematical aggregation of three features: availability, faulty data and Intraclass Correlation Coefficient (ICC) as shown in the equation (1) to estimate the overall quality for medical IoT device data in order to decide whether each device and as a result its derived data, are considered to be of good quality or not. To ensure reliable final results, the data must surpass the established threshold of 90%.

$$OverallQuality = ((Availability - FaultyData) * 0.7) + ICC * 3 \quad (1)$$

The proposed mechanism focuses on gathering data from diverse IoT medical devices, cleaning the data, assessing its quality levels and ensuring that only reliable, high-quality data is made interoperable for further analysis. The evaluation of the proposed mechanism is based on a specific scenario, which may not fully capture the complexities and variations that could arise when dealing with a wide range of IoT medical devices in real-world settings. Table 1 summarizes the reviewed studies according to their context, methodology, considered dimensions and their limitations.

Based on the literature in Medical Data Quality Assessment (MDQA), the effectiveness and applicability of existing methodologies are strongly influenced by the specific type of medical data being analyzed. Medical information systems manage a wide spectrum of data sources—including medical images, EHRs, genomic data and data from medical wearable devices—each of which presents distinct challenges for data quality assessment. Based on this analysis, several key insights and limitations can be drawn:

- **Data-type dependency:** MDQA research is highly dependent on the nature of the medical data under study. Different modalities present unique quality issues and evaluation requirements, for example, MDQA for wearable device data is influenced not only by device-specific technical characteristics but also by environmental and user-related factors such as patient mobility, signal noise, as well as the volume and velocity of continuous real-time data streams.

- **EHR dominance:** The majority of MDQA research focuses on EHR data, likely due to their prevalence, availability and central role in clinical decision-making. While this focus has driven methodological maturity in EHR assessment, it has also limited methodological development for other data types, particularly those emerging from modern IoT-based medical systems.

Research	Context	Methodology	Dimensions or DQ problems	Limitations
[9]	Electronic Health Records	6-step standardized process to assess data quality in a longitudinal registry	-Completeness -Correctness -Concordance -Plausibility -Currency	The need to establish data quality metrics for benchmarking acceptable levels of EMR data inclusiveness through testing and application
[8]	Electronic Health Records	A 3X3 matrix of data quality assessment recommendations based on data variables and data quality dimensions	-Completeness -Correctness -Currency	The need of the adoption of validated systematic methods for EHR data quality assessment
[6]	Electronic Health Records	A package of functions to detect data anomalies	-Missing data -Outliers -Inconsistent data -Imprecise data	The framework is not generalizable to other type of medical data
[12]	IoT medical device	Data quality estimation index	-Availability -Accuracy -Intraclass Correlation Coefficient	The framework does not take all data quality issues into consideration
[11]	Electronic Health Records	Rule based Data quality Assessment	-Incompatibility -Value out of range -Temporal sequence error -Incompleteness -Duplication	Difficult to maintain and poorly scalable when the number of rules is high, rules may not be exhaustive to capture all data errors

Table 1: Comparative summary of the reviewed Medical Data Quality Assessment (MDQA) methodologies and their limitations.

- **Limited generalizability and scalability:** Existing MDQA frameworks often rely on domain-specific assumptions, rigid rule sets, or fixed metric definitions, restricting their applicability across heterogeneous datasets and large-scale deployments. Such constraints face adaptation to new sensor modalities, evolving data formats and multi-center clinical environments.

- **Variable inclusion of medical expertise:** The involvement of healthcare professionals in the design of MDQA frameworks is inconsistent. Given that clinical context is critical for interpreting quality dimensions and setting acceptable thresholds; the absence of systematic medical input can lead to frameworks that are technically sound but clinically misaligned. This gap highlights the importance of interdisciplinary collaboration between data scientists, engineers and healthcare professionals.

- **Narrow focus in wearable sensor research:** MDQA studies addressing medical wearable device data have been disproportionately concentrated on specific signal types, notably electrocardiogram (ECG) and photoplethysmogram (PPG), as seen in [13], [14] and [15]. There is comparatively less work addressing other sensing modalities, such as thermal, despite their growing relevance in early disease detection and personalized monitoring.

- **Real-time applicability gap:** Many existing MDQA approaches are designed for retrospective analysis and lack provisions for incremental computation or streaming evaluation, which limits their applicability to real-time, high-frequency sensor data environments common in modern medical IoT applications.

- **Evaluation under-representation:** A significant portion of MDQA methodologies are validated on single datasets or within constrained clinical scenarios thereby limiting confidence in their robustness, reproducibility and transferability to diverse operational contexts.

As the above analysis indicates, current MDQA approaches address specific aspects individually but remain fragmented in scope and operational design. While some excel in well-structured domains such as EHRs, they often struggle to adapt to heterogeneous sensor data or to function efficiently in real-time, high-volume environments. The diversity of limitations outlined above underscores a persistent gap: the need for an assessment methodology that combines the strengths of existing approaches while overcoming their constraints in terms of scalability, adaptability and computational efficiency. This gap forms the basis for the methodological developments presented in the following sections. Building on the insights gained from the literature, it is essential to ensure that any proposed data quality assessment methodology addresses the full range of issues that can arise in medical sensor networks. Before presenting our approach, the next section outlines the most common data quality challenges encountered in such networks, providing the context for selecting relevant quality dimensions and guiding the design of the assessment framework.

3. Data Quality Issues in medical sensor networks

The availability of wearable and wireless body sensors and systems is rapidly increasing because of their potential to support real-time clinical and remote health monitoring as they enable the monitoring of health trends and the prediction/prevention of health deterioration [16]. Medical sensor data must be of high quality to ensure accurate diagnosis, treatment decisions and healthcare outcomes, yet in real-world applications, it suffers from different quality issues and errors that can occur at any part of the system. Connected and wearable medical devices are a part of IoT (Internet of Things) technologies, which describe the network of physical objects (things) that are embedded with sensors, software and other technologies to connect and exchange data with other devices and systems over the internet [17]. IoT applications or wearable devices are typically structured using a three-layer architecture: Perception Layer, Network Layer and Application Layer as shown in Figure 1. The perception layer concerns physical devices, sensors and other endpoints that collect data from the environment. The network layer is responsible for transmitting data collected by sensors to the next layer in the architecture; it includes various communication protocols, such as Wi-Fi, Bluetooth, Zigbee, cellular networks, etc. Finally, the application layer is where data is analyzed, processed and acted upon to derive insights, make decisions and trigger actions. Sensor errors generally originate in the first layer (Perception layer) where the degradation of sensors or the deterioration of the battery in these devices can contribute to data quality issues such as noisy measurements or unstable readings, missing measurements, sensor drifts, etc. Sensor errors can also be caused by environmental factors that may include temperature fluctuations, humidity, wind levels, electromagnetic interference and pressure changes [18].

Other transmission errors can occur in the network layer, for example, unstable wireless communication such as a Bluetooth card connection can result in both data loss and corruption. Sensor errors at the application layer may also result from processing-related issues, for example, errors can occur in pre-processing sensor measurements if data imputation is based on corrupted data, so the imputed data is also corrupted leading to incorrect interpretations or conclusions. In this layer, errors can also occur from integrating data from multiple sensors or sources, especially if the data formats, protocols, or semantics are inconsistent.

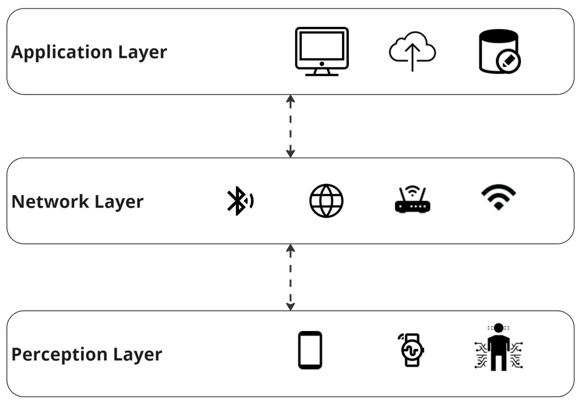


Figure 1: Three-layer architecture of IoT applications.

According to the International Standardization Organization (ISO), an error can be defined as “the result of a measurement minus the true value of the measured”. Our sensor quality issues classification was inspired by [10] where data quality problems were divided into missing and non-missing data. One of the most common data quality issues is **missing data**, which is also known as incomplete data and is addressed in the completeness data quality dimension [18]. In a sensor network context some measurements might be missing due to sensor malfunction at a certain timestamp or transmission errors, but we also can have the entire sensor attribute missing due to sensor defectuosity issues. Additionally, measurements may also be lost during the merging of data from different sources in the application layer.

Data that is non-missing is not always of a good quality, collected data may suffer from uncertainty issues due to environmental factors as mentioned previously. In a medical sensor context, sensors that are generally embedded in wearable devices (fitness trackers, wearable ECG monitors, or smart clothing that monitor vital signs, etc) can be affected by patient physical characteristics such as size in the case of smart clothing or by patient movements which can lead to inconsistencies in the sensors measurements. **Not missing data** issues can be **incorrect** data which refers to a measured information that does not reflect the true or correct values. We can characterize incorrect data into two main types: **inaccurate** data that violates logical rules or thresholds set by knowledge experts or data that is **inconsistent**, meaning data that is contradictory or conflicting with other data in terms of mutual dependency, it can be inconsistent temporally (within the same sensor over time) or spatially (with neighbor sensors). **Duplicate** or redundant data is also a common issue in sensor data, this redundancy can arise from various sources, such as overlapping measurements from multiple sensors capturing the same attribute or high-frequency sampling producing repetitive data points where changes are minimal. For medical devices, the problem of **outdated** sensor data introduces critical considerations for healthcare applications where the currency of information is very essential. Clinical trials are often extensive in duration which can lead to data becoming outdated due to the extended time required for validation, analysis and regulatory approvals. Additionally, the evolution of sensor types and their quality enhancements can make existing data less relevant or reflective of current capabilities, especially if new sensors are introduced during or after the trial period. Another common issue, specific to the medical context, is **imbalanced** data that presents a significant challenge when certain health conditions or outcomes are substantially less frequent than others, leading to skewed class distributions. However, imbalanced data is not typically considered a data quality problem in the traditional sense. Instead, it is a characteristic of the dataset that can pose challenges for certain types of data analysis and modeling tasks, particularly in machine learning. These common data quality issues in medical sensor data are summarized in Figure 2. The following section presents our methodology for medical sensor data quality assessment, which aims to address as many of these issues as possible.

4. Proposed Data Quality Index Methodology

A well-designed data quality assessment framework can help researchers and practitioners understand whether they have enough data and whether the data is of sufficient quality for their specific tasks [19]. Assessing data quality (DQ) with precise DQ characteristics (or dimensions) is a crucial step before defining a methodology to clean and improve the data quality. DQ

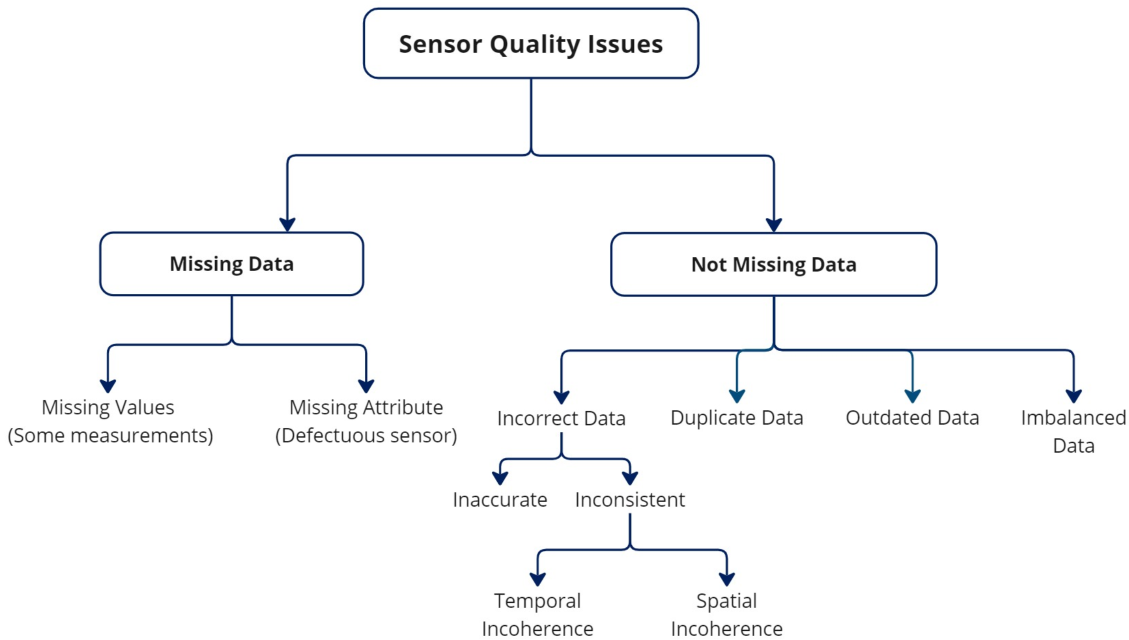


Figure 2: Classification of common data quality issues in medical sensor data.

dimensions can be defined as the specific attributes or characteristics used to assess and measure the quality of data. These dimensions collectively provide a comprehensive framework for evaluating various aspects of data accuracy, reliability, completeness and relevance.

Our methodology consists of developing a data quality index called MS-DQI to assess the quality of data within a medical sensor network. To this end, we started by selecting the appropriate dimensions that need to be used, as long as provided definitions align with our context and cover the sensor issues previously presented. A data quality index considers multiple dimensions of data quality, providing a comprehensive view of the strengths and weaknesses of the dataset. Most importantly, developing a standardized data quality index framework allows for consistent and uniform evaluations across different datasets. MS-DQI is designed as a comprehensive measure that aggregates several key dimensions of data quality, each weighted according to its relative importance within the context of the assessment and by medical and data experts. The resulting index is a score between 0 and 1, where 0 represents very poor quality and 1 indicates very high quality.

The proposed methodology consists of three global steps as presented in

Figure 3:

1. Select target DQ dimensions.
2. Assign weights to each selected DQ dimension.
3. Aggregate the weighted DQ dimensions into a single DQI score.

First of all, we note that the proposed DQI is conceived for data collected by a sensor network denoted by the ensemble S and each sensor s_i contributes to the entire sensor network such as:

$$S = \{s_1, s_2, ..., s_i\}, i \leqslant N \tag{2}$$

Here, N represents the total number of sensors in the network, and s_i represents the individual sensor within the network. We denote the N sensors reading over time for a test T over M time points by a matrix X of size N x M where x_{ij} denotes the measurement of sensor i at time j.

$$X_{NM} = \begin{pmatrix} x_{11} & .. & ... & x_{1n} \\ .. & .. & .. & .. \\ x_{m1} & .. & ... & x_{nm} \end{pmatrix} \tag{3}$$

4.1. DQ Dimensions Selection

The dimensions' selection phase in data quality assessment is very important for identifying key dimensions or aspects of data quality that are relevant to a specific context of use. While all data quality dimensions may initially appear useful, adapting them to the unique characteristics of a dataset or context is crucial. This step ensures that only the most relevant dimensions are selected, aligning with the goals and priorities of the specific use case. Involving stakeholders and experts in this step is essential, especially in medical data projects, as it provides a comprehensive understanding of objectives and requirements and ensures the selected dimensions reflect the specific needs of the application. Moreover, because data quality dimension definitions can be subjective, redundancies and overlaps may occur—for example, between timeliness and currency, or accuracy and validity. Addressing these overlaps requires synthesizing similar dimensions to develop a more streamlined and coherent DQI.

The literature on data quality often provides extensive lists of potential dimensions, which highlights the multifaceted nature of data quality. However, there's often less guidance on how to select and prioritize these dimensions effectively to develop a data quality assessment (DQA) framework designed to a specific context or use case. This might be justified by the fact that dimensions are highly context dependant and selecting dimensions for a DQA framework is a complex decision-making process that requires balancing multiple competing factors, such as relevance, measurability, feasibility and impact. In the proposed framework, we will start by listing potential data quality dimensions with their respective definitions, and then provide a medical IoT definition for each of the dimensions as well as a proposed metric that scores the dimension between 0 and 1. While developing the framework with a team of medical and clinical research experts, we observed that individuals without technical backgrounds often struggled to understand technical definitions. Therefore, we decided to include a third, simplified definition adapted to the specific context of medical sensor data. This additional definition is provided in the use case described in the following section.

First, we meticulously listed the most pertinent DQ dimensions. Among these, three dimensions—availability, timeliness and redundancy—are assessed globally at the test level X_{NM}. In contrast, dimensions such as completeness, precision, accuracy, uncertainty and consistency are evaluated at the individual sensor level s_i and then aggregated to represent the quality of the overall test X_{NM}. In most sensor network applications, not all sensors contribute equally to overall data quality or to the system's objectives. Therefore, we propose assigning weights w_i to each sensor s_i. This weighting approach reflects the relative importance of each sensor in capturing relevant information or detecting anomalies. For instance, sensors located in critical regions or those measuring parameters essential to system performance may be assigned higher weights. Moreover, sensor networks are dynamic systems subject to changes over time—such as sensor failures, replacements, or additions. Assigning adaptable weights allows the framework to accommodate these changes while maintaining accurate data quality assessments. The proposed dimensions are defined as follows:

1. **Availability:** This dimension related to IoT applications where the connected device can go down before its expected uptime. In the work of [12], the authors define this dimension as the ratio of the system UpTime to total time. Assessing the duration of uptime versus downtime provides insights into the reliability of the deployed sensors.

$$Avail_{X_{NM}} = \frac{Uptime}{OperatingTime} = \frac{UpTime}{M} \tag{4}$$

2. **Timeliness:** This dimension refers to the degree to which data values are of the right age in a specific context of use. It refers to the relevance or validity of data over time. Considering the sensor network being updated and improved over time, recent tests are considered to be better and more reliable. The freshness of a test is a way to quantify this dimension, the age of a test can also be used to measure this dimension.

$$Freshness_{X_{NM}} = 1 - \frac{Age}{MaximumAllowedAge} \tag{5}$$

3. **Redundancy**: This dimension refers to the amount of data items that have the same timestamp in IoT applications. Temporal redundancy is a fundamental and practical aspect within IoT and Wireless Sensor Networks (WSNs) due to the dense distribution of sensor nodes. This dense deployment results in a substantial volume of redundant data being generated [20].

$$Redun_{X_{MN}} = 1 - \frac{RepeatedRecords}{ExpectedRecords} = 1 - \frac{RepeatedRecords}{N} \tag{6}$$

4. **Completeness:** Completeness is one of the most commonly used dimensions and is defined as the degree to which subject data associated with an entity has values for all expected attributes and related entity instances [21]. In our case, it is calculated locally at the i^{th} sensor ($i \leqslant N$) level with the formula:

$$SensorCompl_i = \frac{Non - DefectiveMeasurements}{ExpectedMeasurements = M} \tag{7}$$

   where non-defective measurements can include non missing measurements, in-range measurements, non-null values and other rules, the global completeness of a test X_{NM} is calculated as follows:

$$Compl_{X_{NM}} = \frac{\sum_{i=1}^{N} w_i SensorCompl_i}{\sum_{i=1}^{N} w_i} \tag{8}$$

   w_i is the weight assigned to the i^{th} sensor ($i \leqslant N$).

5. **Precision:** According to [21] precision can be defined as the degree to which further measurements of the same phenomenon in a close time instant provide the same or similar results. We can calculate the

precision at the sensor i level with the ratio of the standard deviation (sigma) to the mean of the sensor measurements similar to a coefficient of variation, where a higher value indicates lower variability relative to the mean. The mean and standard deviation exclude null values (i.e., zeros or missing entries).

$$SensorPrec_i = 1 - \frac{\sigma}{\bar{x}_i} \tag{9}$$

$$\bar{x}_i = \frac{1}{m}\sum_{j=1}^{m} x_{ji} \tag{10}$$

$$Prec_{X_{NM}} = \frac{\sum_{i=1}^{N} w_i.SensorPrec_i}{\sum_{i=1}^{N} w_i} \tag{11}$$

w_i is the weight assigned to the i^{th} sensor ($i \leqslant N$).

6. **Accuracy:** According to [21], accuracy is the degree to which data has attributes that correctly represent the true value of the intended attribute of a concept or event in a specific context of use. It can be calculated as the ratio of the correct record to the collected record. We can define a correct record as one that meets a predefined set of rules or conditions depending on the context, example: not null records or respecting a predefined range of values or some other rules as deployed in rule-based DQA methodology by [11].

$$SensorAcc_i = \frac{CorrectRecords}{CollectedRecords} \tag{12}$$

$$Acc_{X_{NM}} = \frac{\sum_{i=1}^{N} w_i.SensorAcc_i}{\sum_{i=1}^{N} w_i} \tag{13}$$

w_i is the weight assigned to the i^{th} sensor ($i \leqslant N$).

7. **Uncertainty:** Uncertainty can be defined as the degree of doubt or variability that exists in data measurements. In our case, the variability for sensor measurements at instant t and $t+1$ should be minimized, so we quantify the uncertainty of sensor s_i with the following formula ($i \leqslant N$, $j \leqslant M$):

$$SensorUncert_i = \sqrt{\frac{\sum_{j=1}^{M-1}(x_{ij} - x_{ij+1})^2}{2.M.\bar{x}_i}} \tag{14}$$

$$Uncert_{X_{NM}} = 1 - \frac{\sum_{i=1}^{N} w_i.SensorUncert_i}{\sum_{i=1}^{N} w_i} \tag{15}$$

w_i is the weight assigned to the i^{th} sensor.
To maintain comparability across sensors and acquisition sessions, the sensor-level uncertainty values were normalized using min–max scaling. Each uncertainty value was divided by the maximum uncertainty observed across all sensors and all sessions in the dataset, ensuring that the resulting metric falls within the [0,1] interval. This normalization step makes the uncertainty measure dimensionless and prevents variations in absolute temperature levels or signal amplitude from influencing cross-sensor or cross-test comparisons.

8. **Spatial Consistency:** According to [21] consistency can be defined as the degree to which data has attributes that are free from contradiction, and are coherent with other data in a specific context of use. In a sensor network, sensors are considered to be consistent if neighbor sensors are as highly correlated as possible.

$$SensorConsis_i = \frac{\sum_{j=1}^{N} \lambda(s_i, s_k).PearsCoef(x_i, x_k)}{n} \tag{16}$$

Here $\lambda(s_i,s_j)$ denotes the Euclidean distanc between the sensor s_i and s_k $(i \leqslant N, k \leqslant N)$ and $PearsCoef(x_i, x_k)$ is the Pearson correlation coefficient of the variables x_i and x_k. If x_i and x_k are not normally distributed, alternative correlation metrics such as Spearman's or Kendall's Tau may be more appropriate [22].

$$Consis_{X_{NM}} = \frac{\sum_{i=1}^{N} w_i.SensorConsis_i}{\sum_{i=1}^{N} w_i} \tag{17}$$

This is not an exhaustive list of all DQ dimensions and metrics, researchers should explore and adapt appropriate metrics for each use case. As mentioned before these definitions should be adapted to a specific medical IoT context to to enhance clarity for target users. The proposed dimensions and their respective definitions are summarized in Table 2.

As mentioned earlier, the process of data quality dimensions selection is not widely discussed in the literature; however, we explored other fields that apply the concept of creating an index from multiple dimensions. We found a substantial amount of research in the development of Water Quality Index, Air Quality Index and Environmental impact assessment studies,

DQ Dimension	Standard definition	Medical IoT Definition
Availability	The ratio of the system uptime to total time	The time the device or the sensor spends working well compared to the total time it is expected to be operational.
Timeliness	The degree to which data has attributes that are of the right age in a specific context of use	Assesses the extent to which the device data attributes are appropriately up-to-date for a given context of use
Redundancy	The amount of data items that have the same timestamp.	Assesses how often the device records identical information at a specific point in time.
Completeness	The degree to which subject data associated with an entity has values for all expected attributes and related entity instances	Assesses how well it captures and transmits health data without missing or defective measurements
Precision	The degree to which further measurements of the same phenomenon in a close time instant provide the same or similar results.	Reflects the degree to which the device-recorded information may deviate from absolute certainty, acknowledging factors that introduce variability or imprecision.
Accuracy	The degree to which data has attributes that correctly represent the true value of the intended attribute of a concept or event in a specific context of use	It reflects how closely the recorded information aligns with the actual values
Uncertainty	The degree of doubt or variability that exists in data measurements	It assesses the variability between measurements taken by the sensors in time, variability between measurements between instant t and $t+1$ should be minimized.
Spatial consistency	The degree to which data has attributes that are free from contradiction and are coherent with other data in a specific context of use	The degree to which data from nearby sensors exhibits measurements that logically correspond with their physical proximity.

Table 2: Data quality dimensions context definitions.

which helped us gather some useful techniques for dimension selection. We provide a comprehensive overview of the different approaches available for DQ dimension selection by classifying them into three major categories: approaches based on expert knowledge, data, and data models. The techniques are explained below:

1. **Expert knowledge-based approaches:** Techniques in this category rely on the expertise and judgment of domain experts such as medical staff, clinical researchers or data experts to identify and prioritize data quality dimensions.
    - **Delphi method:** The Delphi method is a structured technique that involves iterative rounds of surveys and feedback from a panel of experts. Experts provide their opinions on relevant data quality dimensions, and the results are aggregated and fed back to the panel in subsequent rounds. This process continues until a consensus is reached on the most important dimensions. The authors in [23] used Delphi method to identify consensus on proposed data quality strategies from previous rounds that met or exceeded the acceptance threshold to construct subsequent round questions.
    - **Expert Panel judgment process:** This process brings together domain experts to collaboratively evaluate and select relevant dimensions [24]. Experts review relevant literature, discuss their insights and experiences and reach a consensus on the most relevant dimensions based on their collective judgment and expertise.
2. **Data-based approaches:** Techniques in this category leverage data-driven approaches to evaluate the relevance and significance of different dimensions based on the data being studied.
    - **Data availability:** This approach assesses the extent to which data is necessary to measure and monitor each dimension are accessible and obtainable.
    - **Correlation analysis:** Correlation analysis examines the relationships between different data quality dimensions to identify patterns, redundancies, or dependencies. Statistical techniques such as Pearson correlation coefficients or Spearman's rank correlation coefficients are commonly used to quantify the strength and direction of associations between dimensions.
    - **Data significance:** Data significance evaluates the impact of each dimension on the overall quality and usability of the data.

This involves assessing the relevance and importance of each dimension in achieving the objectives and requirements of the data users and stakeholders similar to conducting a sensitivity analysis.

3. **Model-based approaches:** Techniques in this category use predictive models or algorithms to prioritize dimensions based on their importance, such as:
   - **Feature importance ranking:** Machine learning models such as decision trees, random forests, or gradient boosting algorithms can be employed to rank dimensions based on their contribution to model performance. Feature importance scores or variable importance measures are computed to quantify the relative importance of each dimension in the predictive model.

Each technique offers unique advantages and considerations and the choice of methods may vary depending on the specific context, resources and expertise available.

4.2. DQ Dimensions weighting

Once relevant dimensions are identified, the weighting phase follows; where the goal is to assign relative importance to the selected dimensions based on their impact on data usability, decision-making and final objectives. Dimensions that are critical for strategic decision-making or application goals may be assigned higher weights, while dimensions with lower significance may receive lower weights. To ensure the index is representative and normalized across tests, the sum of weights v_k must be equal to 1 and the weights are in the range of [0,1] such as:

$$\sum_{k=1}^{l} v_k = 1$$

where v_k is the weight of the k_{th} dimension of a total of l dimensions. Here are some common techniques for weighting data quality dimensions;organized by their respective categories:

1. **Expert knowledge-based approaches:**
   - **Analytical Hierarchy Process (AHP):** A multi-criteria decision making approach that helps users prioritize dimensions by

structuring decision problems hierarchically and comparing pairwise preferences [25]. Experts evaluate the relative importance of criteria or factors by making pairwise comparisons between them. Each criterion is compared to every other criteria using a scale of relative importance, typically ranging from 1 to 9, where 1 refers to equal importance, and 9 refers to extreme importance. Reciprocal values are used for inverse comparisons (e.g., if dimension A is "twice as important as" dimension B, then dimension B is "half as important as" dimension A). In their study, [25] applied the Pairwise Comparison Matrix (PCM) technique derived from AHP for weight assignment. This method involves organizing comparison results into a matrix, where the diagonal elements are set to 1, indicating that a dimension has equal importance when compared with itself. Subsequently, the matrix is normalized to ensure that the sum of each column equals one. Finally, weights are computed as the mean of each row in the normalized matrix.

- **The Importance Scale Matrix (ISM)**: A decision-making technique used to prioritize criteria or parameters based on their relative importance. The comparisons are made without initially ranking the parameters. Once all pairwise comparisons are completed, the matrix of preferences obtained is analyzed using the 'eigenvector' method to assign weights to each parameter. The authors in [26] detailed this technique for parameter weight assignment for environmental impact evaluation. Although not widely used in data quality dimension weighting it can be generalized to any decision making problem.

2. **Data-based approaches:** Data-based approaches for data quality dimensions weighting support statistical techniques and empirical data to assign weights to different dimensions based on their importance and impact on data quality

   - **Entropy:** Entropy is a measure of uncertainty or disorder in a dataset. In the context of data quality dimensions weighting, entropy can be used to assess the variability or unpredictability of data quality across different dimensions. Dimensions with higher entropy values indicate greater uncertainty or variability and may be assigned lower weights, whereas dimensions with lower entropy values indicate more consistent or predictable data quality and may be assigned higher weights. Entropy-based methods help prioritize dimensions that contribute most to reducing uncertainty and improving data quality.

- **Machine learning:** A novel approach to assign weights to parameters is proposed by the work of [27] about creating a water quality index that takes several parameters in entry. The authors proposed to rank the parameters using a ML algorithms such as Random Forests to rank the parameters based on their contribution significance to the model. Once the parameters are ranked, their weights can be calculated using different mathematical and statistical attribute weighting methods such as the ones presented in Table 3 where l is the number of dimensions, k is the rank of the k^{th} dimension, v_k is the weight of k^{th} the dimension.

Attribute weighting methods	**Formula**
Rank Sum (RS) method	$v_k = \frac{l+1-k}{\sum_{d=1}^{l} d}$
Rank Reciprocal (RR) method	$v_k = \frac{\frac{1}{k}}{\sum_{d=1}^{l} \frac{1}{d}}$
Rank Order Centroid (ROC) method	$v_k = \frac{1}{l} \sum_{d=1}^{l} \frac{1}{d}$
Equal weights method	$v_k = \frac{1}{l}$

Table 3: Mathematical and statistical dimensions weighting methods.

- **Sensitivity analysis:** Sensitivity analysis assesses the sensitivity of data quality outcomes to changes in the weighting of different dimensions. By systematically varying the weights assigned to each dimension and observing the resulting changes in data quality metrics, sensitivity analysis helps identify which dimensions have the greatest influence on data quality and how changes in their weights affect overall data quality outcomes. Sensitivity analysis provides valuable insights into the robustness and stability of weight assignments.

4.3. DQ Dimensions Aggregation

In this phase, the goal is to obtain a quality index for each test X_{NM}; which is computed by aggregating the calculated dimensions, employing a weighted aggregation approach. A lot of weighted aggregation functions could be used in this phase, simple additive or multiplicative functions are the most commonly used in the literature. We list the most used ones in Table 4 where $DQI_{X_{NM}}$ is the DQI of the test X_{NM}, l is the number of the selected dimensions, v_k is the weight of k^{th} the dimension, D_k is the calculated metric for the k_{th} dimension such as in (4-17).

Aggregation Function	**Formula**
Additive function	$DQI_{X_{NM}} = \sum_{k=1}^{l} v_k D_k$
Multiplicative function	$DQI_{X_{NM}} = \prod_{k=1}^{l} v_k D_k$
Combined multiplicative and additive: This aggregation is used in developing a river water quality index by [28].	
Square root of the harmonic mean	$DQI_{X_{NM}} = \sqrt{\frac{l}{\sum_{k=1}^{l} D_k^2}}$
Minimum: Using the minimum function to calculate a quality index is a practical approach in situations where the overall quality is determined by the weakest link or poorest performance in one or more dimensions.	$DQI_{X_{NM}} = \min(D_1, D_2, \ldots, D_k)$
Weighted Quadratic Mean: The authors in [29] proposed calculating a water quality index with the quadratic mean, which could be adapted to our problem as follows:	$DQI_{X_{NM}} = \sqrt{\sum_{k=1}^{l} v_k D_k^2}$

Table 4: Data quality dimensions aggregation functions.

Using an aggregation function to calculate a DQI for each test provides a synthesized measure that reflects the collective impact of these quality aspects on the data's overall reliability and usefulness within the context of a test X_{NM}. The choice of aggregation function plays a critical role in how the Data Quality Index responds to datasets with different distributions or the presence of outliers. While the weighted additive function is commonly used in the literature because of its interpretability and alignment with expert-driven weighting, the MS-DQI framework is designed to accommodate alternative aggregation strategies. For example, multiplicative aggregation emphasizes the lowest-performing dimensions, thereby penalizing cases where a single weak dimension substantially undermines overall data quality. The minimum operator provides a non-compensatory approach, ensuring that poor performance in one dimension cannot be masked by higher scores in others. Conversely, quadratic or harmonic mean functions can reduce the influence of extreme values, offering more robustness to outliers. This modularity enables practitioners to select the aggregation function most appropriate for their dataset's distributional properties and the desired tolerance for compensability between dimensions.

The techniques used for each step of the proposed DQ Assessment are summarized in Figure 3. In this section, we introduced MS-DQI (Medical Sensors Data Quality Index), a comprehensive framework designed to assess data quality in medical sensor networks. The methodology is composed of three main steps:

1. Selecting appropriate data quality dimensions that are critical for the assessment of sensor data quality.
2. Assigning weights to these dimensions to reflect their relative importance, based on expert knowledge and data characteristics.
3. Aggregating the weighted dimensions into a single DQI score, providing a synthesized view of the dataset's quality.

Scalability was a key design consideration in the development of MS-DQI. The framework is built on a modular and hierarchical architecture, which allows for parallel and distributed computation across sensors, devices and time windows. Each data quality dimension is computed independently at the sensor level and can be aggregated locally or centrally, enabling the framework to scale seamlessly with the size of the network. Furthermore, the use of matrix operations and simple statistical measures makes MS-DQI computationally efficient and well-suited for real-time or high-frequency data environments, even on resource-constrained devices. This positions MS-DQI as a practical solution for large-scale deployments such as clinical monitoring systems, wearable networks, or wide sensor infrastructures.

In the following section, we apply the MS-DQI methodology to a dataset derived from the BraConnect (CBRA) clinical trial. This dataset contains temperature readings from wearable thermal sensors used for breast cancer detection. Through this case study, we will evaluate the quality of the collected data and demonstrate the practical utility of the MS-DQI in a real-world scenario.

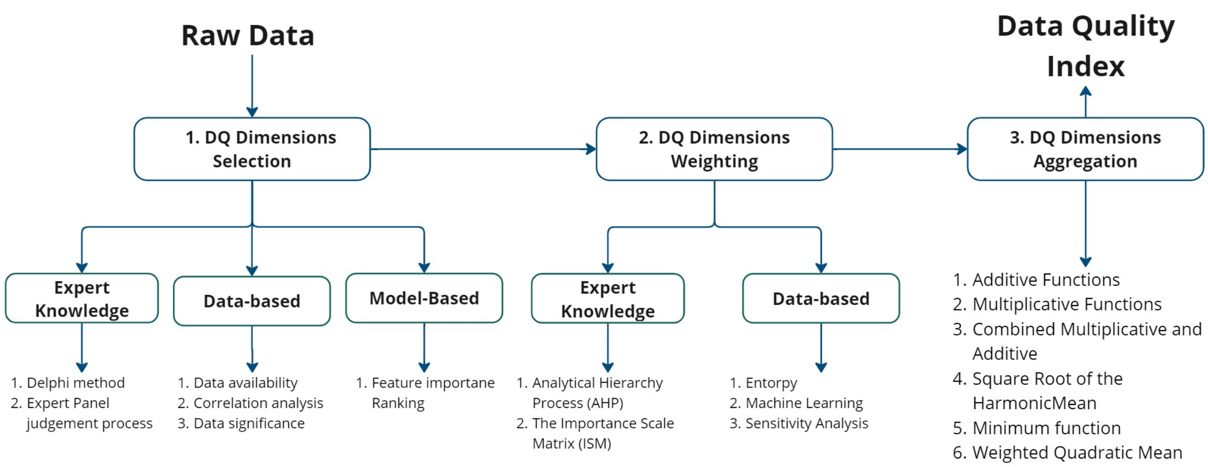


Figure 3: MS-DQI methodology workflow: dimension selection, weighting, and aggregation.

5. Case Study: Breast Cancer Dataset

The private dataset used in this study was acquired from a clinical trial registered with ClinicalTrials.gov (Identifier: NCT05294016). This trial tested a wearable device equipped with N thermal sensors for breast cancer detection. The dataset includes temperature readings from 35 patients and control subjects, all wearing a smart bra for z minutes, with measurements every 3 seconds from N sensors. Additional clinical data such as breast cancer presence, tumor size, mammography findings, and breast density were recorded.

Each test generates a CSV file with N features and a variable number of measurements (20*z measurements or M time points). The temperature data is stored as floating-point values with a precision of 10^{-3} degrees and ranges between -35°C and 55°C. The ratio of missing data varies across tests, as summarized in Table 5. Metadata relative to testing information such as test date, device label(multiple tests were conducted using different wearable devices), and ambient temperature were also available.

In the context of this case study, sensor weights were determined based on their anatomical placement on the wearable device. Given the higher diagnostic relevance of the breast surface in detecting thermal anomalies associated with underlying pathologies, sensors located directly on the breast were assigned greater weights. In contrast, sensors positioned in peripheral regions—such as under the armpits or between the breasts—were assigned relatively lower weights. This weighting strategy reflects the spatial hetero-

Data Element	Type	Range	Missing data
Timestamp	Integer	[0,20*z]	Variable with the test
$Sensor_i$(i $\leq N$)	Float (10^{-3}°C precision)	[-35°C,55°C]	Variable with the test

Table 5: BraConnect data elements characteristics.

geneity of clinical importance across sensor locations and ensures that data quality assessments are more sensitive to regions of higher diagnostic value.

5.1. Dimensions selection

The primary objective of calculating a DQI index for this dataset is to identify abnormal tests that did not meet expectations in terms of quality. The smart bra is equipped with a wireless sensor network that makes it prone to many DQ issues. Using the classification of quality issues in sensor data and insights given by the clinical research team responsible for conducting the tests we summarized the different issue causes observed in our experiment. These issues along with their causes can be covered with some of the proposed dimensions in our MS-DQI methodology. The summary of the encountered quality issues with their respective causes and dimensions is illustrated in Figure 4.

The highlighted dimensions in Figure 4 helped us map the observed data quality failure modes from the study (connection loss, defective sensors, noise, fit-related variation, device aging) to candidate dimensions. The dimensions that accounted for most of the identified data quality issues were: Availability, Completeness, Accuracy, Precision, Spatial Consistency, Redundancy and Timeliness. Then, we tested one of the proposed data-based approaches in the MS-DQI methodology to select relevant dimensions. Correlation analysis was run on all the calculated dimensions for all tests, results are presented in Figure 5. First we notice that Accuracy and completeness are highly correlated and that is due to the shared criteria definition (Null values in accuracy and missing data in completeness). Additionally, we notice that the redundancy row is null values because all tests redundancy value is 1 (no repeated time-stamped measurements). Finally, we can see that Precision and Uncertainty are highly correlated (0.94), indicating that a single dimension is sufficient to represent sensor imprecision. From this correlation analysis we selected the following discriminator dimensions set: Availability, Completeness, Precision, Spatial Consistency, Timeliness.

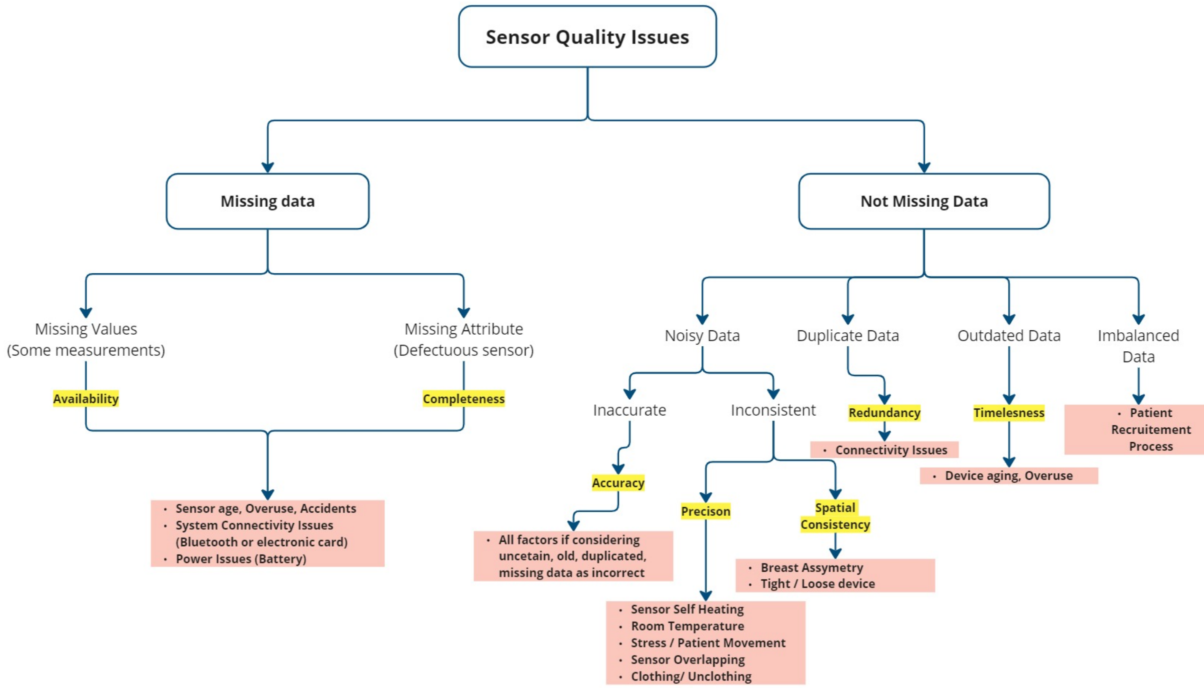


Figure 4: Mapping of observed sensor data quality issues to their causes and associated quality dimensions.

5.2. Dimensions weighting and aggregation

In this step, we employed the Analytic Hierarchy Process (AHP) as the expert-based weighting technique, in accordance with the proposed framework. To ensure consistency and clarity in the prioritization process, we first presented the formal definitions of each data quality dimension (Table 2) to all participating experts. Since some participants came from non-technical backgrounds (e.g., clinicians and members of the clinical research team), we supplemented these definitions with application-specific contextual descriptions specified to the connected bra scenario (Table 6). This approach aimed to facilitate a shared understanding of the dimensions across diverse expert profiles. The used scale is presented in [25] and in Table 7. The AHP-based survey was conducted with a multidisciplinary panel consisted of five members of the clinical research team, one medical expert, and two members of the data and engineering teams. The complete survey form is available at: Survey Form.

Once the survey was filled by the experts, the results are arranged in a

DQ Dimension	Standard definition	Medical IoT Definition	Connected Bra context
Availability	The ratio of the system uptime to total time	The time the device or the sensor spends functioning well compared to the total time it is expected to be operational.	Availability assesses how often the Bra measures data during the whole test time consistently. It is at 100% if the bra completed the expected time, lower if the test stopped before the expected time and 0 if the test failed to capture any data.
Timeliness	The degree to which data has attributes that are of the right age in a specific context of use	Assesses the extent to which the device data attributes are appropriately up-to-date for a given context of use	Assesses if the test measurements remain current and within an acceptable age reflecting the continuous enhancements to the device.
Redundancy	The amount of data items that have the same timestamp.	Assesses how often the device records identical information at a specific point in time.	The frequency of having identical thermal measurements with the same timestamp (same measurement instant)
Completeness	The degree to which subject data associated with an entity has values for all expected attributes and related entity instances	assesses how well it captures and transmits health data without missing or defective measurements	Assesses how well all thermal sensors capture data without missing any measurements (failure) during the test
Precision	The degree to which further measurements of the same phenomenon in a close time instant provide the same or similar results.	It reflects the degree to which the device-recorded information may deviate from absolute certainty, acknowledging factors that introduce variability or imprecision.	Assesses the fluctuations or imprecision caused by external interference or disturbances and noise in the temperature data during the test time
Accuracy	The degree to which data has attributes that correctly represent the true value of the intended attribute of a concept or event in a specific context of use	Reflects how closely the recorded information aligns with the actual values	Evaluates how closely the recorded temperature data represents the true temperature values of the breast during the test. Correct records are those that meet specific criteria. They are not missing, free from noise and sourced from non-defective sensors. Additionally, they adhere to the temperature range rules established by experts.
Spatial consistency	The degree to which data has attributes that are free from contradiction and are coherent with other data in a specific context of use	The degree to which data from nearby sensors exhibits measurements that logically correspond with their physical proximity.	Evaluates whether closely located sensors yield temperature measurements that align with their neighbor sensors during the test.

Table 6: Adapted definitions of data quality dimensions in the context of a connected bra sensor network.

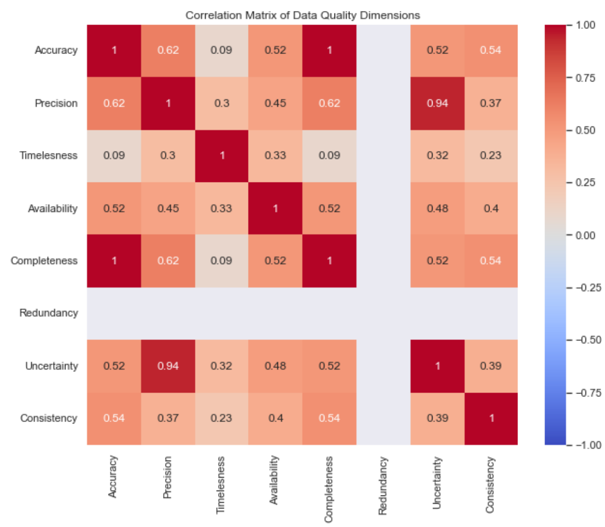


Figure 5: Correlation matrix of candidate data quality dimensions in the case study dataset.

matrix; where the diagonal is equal to 1, indicating that a dimension compared to itself is equally important. The values over the main diagonal are actually those given by the expert preferences, while the values below the main diagonal are reciprocals from those above it, i.e. given the l x l matrix A with elements a_{ko} where l is the number of dimensions, it satisfies :

$$a_{ko} \times a_{ok} = 1, \forall k \leq l, \forall o \leq l$$

An example from a verbal assessment matrix is shown in (18).

$$\mathbf{A} = \begin{array}{l} \\ \textbf{Precision} \\ \textbf{Timeliness} \\ \textbf{Availability} \\ \textbf{Completeness} \\ \textbf{Consistency} \end{array} \begin{array}{c} \begin{array}{ccccc} \textbf{Precision} & \textbf{Timeliness} & \textbf{Availability} & \textbf{Completeness} & \textbf{Consistency} \end{array} \\ \begin{pmatrix} 1 & 7 & 3 & 7 & 1 \\ 0.14 & 1 & 0.14 & 0.33 & 0.11 \\ 0.33 & 7 & 1 & 4 & 0.2 \\ 0.14 & 3 & 0.25 & 1 & 0.12 \\ 1 & 9 & 5 & 8 & 1 \end{pmatrix} \end{array} \tag{18}$$

The next step is to normalize the matrix so that the sum of every column is equal to one such as:

$$\bar{a_{ko}} = \frac{a_{ko}}{\sum_{k=1}^{l} a_{ko}}$$

Individual pairwise comparison matrices were aggregated using the geometric mean to obtain a single consensus matrix, following Saaty's recommendation for group decision-making [30]. The consistency of the aggregated

Intensity of importance	Definition	Explanation
1	Equal importance	Two activities contribute equally to the objective.
3	Moderate importance of one over another	Experience and judgment strongly favor one activity over another.
5	Essential or strong importance	Experience and judgment strongly favor one activity over another.
7	Very strong importance	An activity is strongly favored and its dominance demonstrated in practice.
9	Extreme importance	The evidence favoring one activity over another is of the highest possible order of affirmation.

Table 7: Fundamental scale for pairwise comparison [25].

expert judgments was verified using the AHP Consistency Ratio (CR). The resulting value was CR = 0.0803 that fall below the standard 0.10 threshold, indicating that the consolidated pairwise evaluations are coherent and statistically reliable. Finally, the weights are calculated as the mean of each row in the normalized matrix.

$$\bar{v_k} = \frac{\sum_{o=1}^{l} \bar{a_{ko}}}{l}$$

The resulted weights are illustrated in Table 8. In the dimensions aggregation step, we computed the DQI using a simple weighted additive function such as:

$$DQI_{X_{NM}} = \sum{}_{k=1}^{l} v_k D_k$$

$$DQI_{X_{NM}} = v_1.Prec_{X_{NM}} + v_2.Freshness_{X_{NM}} + v_3.Avail_{X_{NM}} + v_4.Compl_{X_{NM}} + v_5.Consis_{X_{NM}}$$

where X_{NM} is a test with N sensor at M time points (20*z minutes), l is the number of the selected dimensions, v_k is the weight of k^{th} the dimension, D_k is the calculated metric for the k_{th} dimension such as in (4-17).

	Precision	Timeliness	Availability	Completeness	Consistency
Weights	0.343546	0.033417	0.15425	0.057989	0.410797

Table 8: Final weights of selected data quality dimensions obtained via AHP analysis.

5.3. Results and discussion

An example of the results of the calculated DQI and dimensions from some tests are illustrated in Table 9.

Test	Precision	Timeliness	Availability	Completeness	Consistency	Mean	DQI
$Test_1$	88.40	76.11	97.67	97.63	65.63	85.09	80.600588
$Test_2$	97.45	76.39	97.50	98.95	68.82	87.82	85.079805
$Test_3$	98.35	76.93	98.00	99.60	68.76	88.33	85.497212
$Test_4$	99.25	77.89	1.83	99.36	65.82	68.83	69.782593
$Test_5$	98.14	79.33	97.33	97.33	66.62	87.75	84.391179

Table 9: Computed Data Quality Index (DQI) scores for individual tests in the case study dataset.

The histogram of the resulting DQI values through the different tests is illustrated in 6. The DQI ranged from a minimum of around 60.6% to a maximum of approximately 85.75%. Most of the values are clustered between 80% and 85%, indicating generally high data quality across the clinical tests. We also can notice that four tests have relatively a low DQI which is less than 75%. The lowest DQI is 60.6%, which is notably lower than the others, suggesting that this particular test has significantly lower data quality. While there is some variation in the data quality indices, most tests have values that are fairly close to each other, indicating some consistency in data quality. The tests that have lower DQI share a very low Availability score which means that the test has stopped a lot earlier than expected mostly due to connection errors. The general trend indicates that most clinical tests maintain a high level of data quality, with a few exceptions that could be investigated further to identify and address the root causes of poor data quality.

In order to study the variation of data quality over time, we plotted the DQI values according to the test dates, as shown in Figure 7. We observe that lower DQI values are primarily associated with tests conducted before July 2023. In contrast, DQI values tend to remain higher for tests performed after that date, which corresponds to the deployment of new prototype sets in September 2024.

This trend suggests a clear improvement in data quality following hardware and protocol updates. The higher and more stable DQI scores in recent

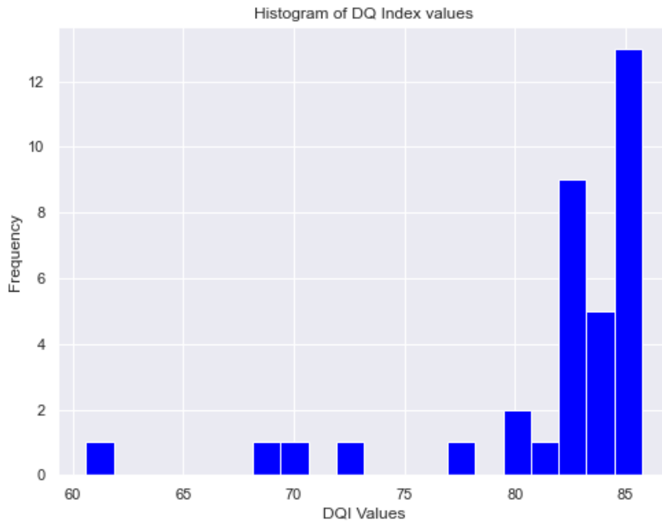


Figure 6: Distribution of Data Quality Index (DQI) scores across tests.

tests indicate that the newer prototypes, possibly due to better improved sensor calibration, connectivity and manufacturing processes, have positively impacted the reliability of the collected data.

Additionally, we observed a correlation between the patient BMI (Body Mass Index), overall DQI, and the consistency dimension in particular; as shown in Figures 8a and 8b. We can see that lower DQI and consistency values correspond to lower BMI values (less than 24 Kg/m^2); which can be explained by the use of a single bra size that works better with women with higher BMI which generally corresponds to a larger breast size. This indicates that device fit is a modifiable factor influencing data quality which supports the use of MS-DQI as a quality screening tool when deploying across varied body types, it can also help flag when fit adjustments or different device sizes are needed prior to clinical interpretation.

Our analysis shows that MS-DQI is capable of detecting both critical failures and more subtle degradations in sensor behavior. For example, Test 4 in Table 9 exhibits a low overall DQI despite relatively normal readings from most sensors, primarily due to low Availability from an early sensor shutdown. In other cases, we observed tests with high Completeness but low Precision and Spatial Consistency, indicating irregular signal noise and localized sensor drift. These findings demonstrate that the multi-dimensional formulation of MS-DQI allows it to reflect performance issues that may not be

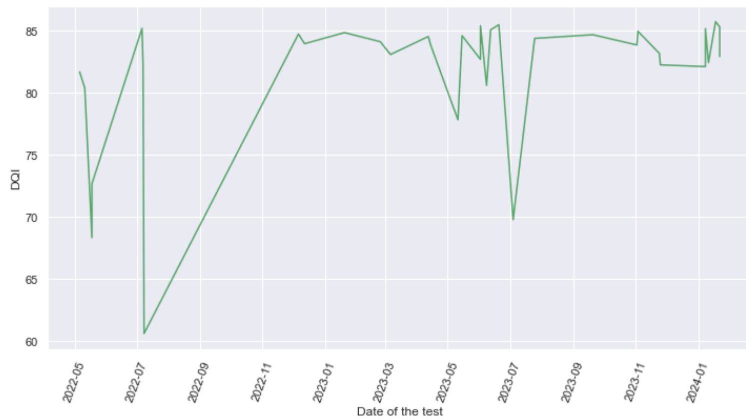
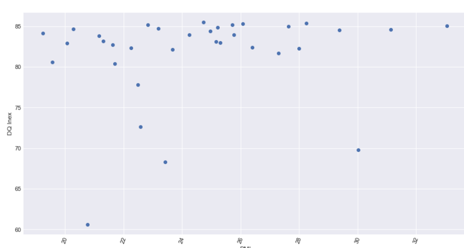
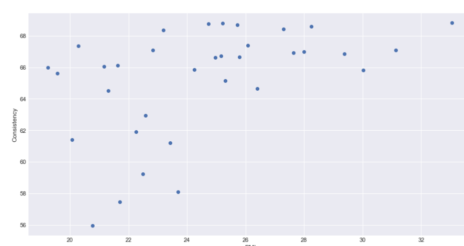


Figure 7: Evolution of Data Quality Index (DQI) scores over time.



(a) Impact of BMI on overall DQI



(b) Impact of BMI on consistency dimension

Figure 8: Body Mass Index Impact on Overall DQI and Consistency.

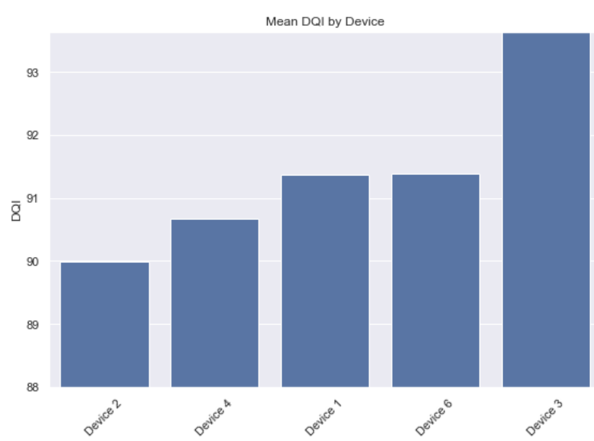


Figure 9: Mean Data Quality Index (DQI) scores across tested devices.

immediately visible in raw signal traces, thereby providing early indications of sensor malfunction.

The 35 tests were performed using different devices. The bar plot in Figure 9 illustrates the mean DQI recorded for each of the tested devices. We observe noticeable variation in DQI across devices: Device 3 achieved the highest mean DQI, indicating superior data quality, while Device 2 recorded the lowest. These discrepancies may stem from various factors, such as hardware limitations, sensor calibration, or sensor defects. This device-level analysis enables targeted improvements; devices with lower DQI values can be prioritized for maintenance, recalibration, or replacement. This prioritization may reduce unplanned downtime and improve overall system reliability. Moreover, the DQI serves as a quantifiable and standardized indicator for evaluating device performance over time. By establishing a baseline threshold for acceptable DQI levels, underperforming devices can be systematically flagged for further investigation or exclude them from critical applications such as clinical trials.

5.4. Sensitivity Analysis

The primary goal of data quality assessment is to ensure that insights derived from the data are both accurate and reliable. Since data quality encompasses multiple dimensions, understanding how each dimension impacts model performance is essential for validating their relevance. In this section, we investigate the influence of individual data quality dimensions on a baseline classification model designed to detect the presence of breast tumors using temperature data. This baseline model serves as a reference to evaluate how variations in data quality affect its predictive performance. By analyzing these effects, we aim to identify the most critical dimensions and establish quality thresholds beyond which model reliability may be compromised. To conduct this analysis, we followed the steps outlined below and illustrated in Figure 10:

1. **Modifying input data:** We began by modifying the dataset to simulate variations in specific data quality dimension metrics. First, we randomly selected m tests for modification to capture the inherent variability in data quality across different tests—variability that may stem from differences in devices, patient conditions, or environmental factors. For each selected test, we altered the data to independently vary each of the following quality dimensions:

- **Availability**: we deleted the last l rows with l being randomly generated for each test to simulate reduced availability.
- **Precision**: we added a Gaussian noise with a random noise standard deviation simulating measurement errors or inaccuracies commonly encountered in real-world sensor data.
- **Completeness**: we selected s sensors to set as defectuous (all measurements of the chosen sensors are set to null) with $s \leq 10$ per breast as the maximum allowed number of defectuous sensors is two per quadrant per beast in our current protocol (The breast is divided into 4 quadrant). Additionally, to simulate poor quality sensors that go off only at certain timestamps, randomly selected index pairs were set to null to simulate intermittent sensor dropout.

2. **Recalculation of Data Quality Dimensions:** Once we apply the modifications, we recalculate the data quality dimensions for each test and take the mean of all tests to represent the dimension level for the modified dataset.
3. **Model Training and Evaluation:** We train the same baseline model for all different dimension levels and evaluate the model's accuracy with a 5-fold cross-validation. We tested two models, the first one is a simple Extra Trees Classifier being trained on the mean temperatures over all test measurements and a second one is an LSTM network being trained with the whole test time series. The Extra Trees classifier was implemented using scikit-learn with 500 trees, the Gini impurity criterion, and a minimum of 2 samples per split and 1 per leaf. The LSTM baseline was implemented in TensorFlow. The model consisted of a single LSTM layer with 50 units, followed by a Dropout layer (0.2) and a sigmoid output neuron. It was trained with the Adam optimizer, binary cross-entropy loss, batch size 32, and 100 epochs. To ensure robust evaluation for both models, we applied 5-fold StratifiedKFold cross-validation, reporting the mean test performance across folds.

Availability Variation. We managed to have different datasets with different availability level ranging from 31% to 100% (the 100% availability is obtained with forecasting all measurements). Figure 11 illustrates the relationship between availability and model accuracy. We can notice that the availability level does not significantly impact the accuracy of the mean model

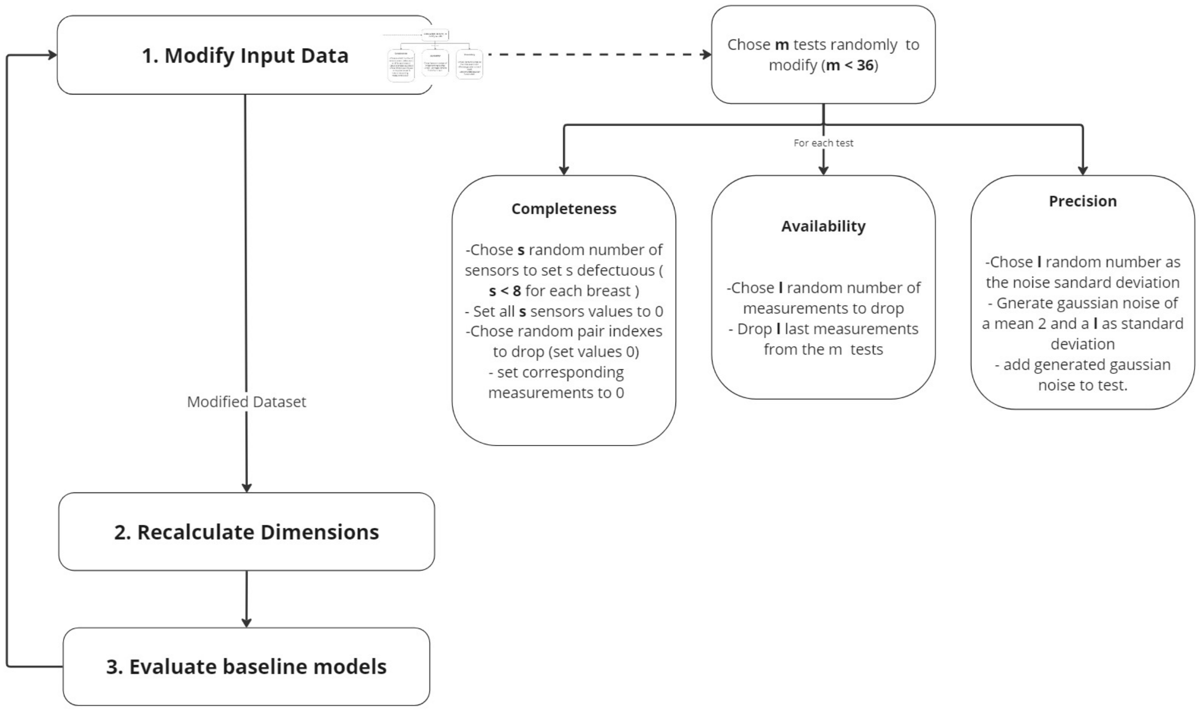


Figure 10: Sensitivity analysis workflow for varying data quality dimensions.

which suggests that the mean model's performance remains relatively stable across different availability levels. In contrast, the model trained on time series data exhibits a strong correlation with availability level, the accuracy of the time series model demonstrates shows a clear dependence on availability level. The highest model accuracy, reaching 90%, is attained only when the availability level is at its maximum (i.e., 100%), and the lowest accuracy (53,8 %) is obtained with the lowest availability level (i.e., 40,5%). Time series models are based on sequential data and rely on capturing temporal dependencies and patterns present in the data. When data availability is limited, especially for sequential observations over time, the model may struggle to learn and generalize from incomplete or random data points, which can lead to reduced model accuracy as important temporal relationships may be overlooked or misinterpreted.

Overall, the mean model implies that the mean may serve as a representative measure even when data availability is limited and the correlation between time series model accuracy and data availability underscores the importance of ensuring sufficient and consistent data availability for effective time series analysis and modeling. Strategies to address data availability issues may include data imputation techniques, forecasting methods, optimal test time and device design to overcome availability issues.

Completeness Variation. Figure 12 illustrates the accuracy effect on both mean and time series model. With completeness, we observe a positive correlation between completeness and model accuracy for both the time series model and the mean model; as expected where higher completeness levels tend to correspond to higher accuracy values, suggesting that more complete datasets result in improved model performance. The observed positive relationship between completeness and accuracy reaffirms the importance of data completeness in achieving reliable and accurate model predictions where higher data completeness ensures models have access to more information, allowing them to better capture underlying patterns and relationships in the data. Additionally, comparing the two models, we noted that the mean model generally achieves higher accuracy values compared to the time series model across various completeness levels. This difference in performance may suggest that aggregating temperature times series to their mean may encounter some over-fitting problems.

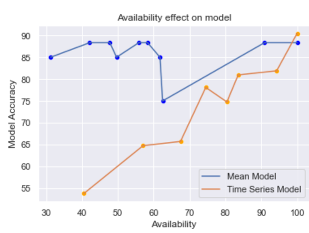


Figure 11: Effect of availability variation on detection model performance.

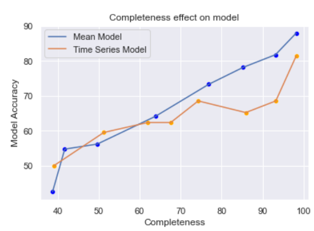


Figure 12: Effect of completeness variation on detection model performance.

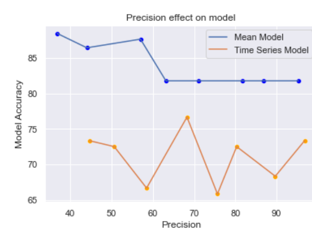


Figure 13: Effect of precision variation on detection model performance.

Precision Variation. Figure 13 illustrates the impact of precision variation on the detection model's accuracy. The datasets used exhibited precision levels ranging from 36.83% to 96.78%. We observe that the accuracy of the mean model decreases as precision increases. This result may suggest that the simplicity of the mean model allows it to perform better with noisier data, where added variability smooths out minor fluctuations that the model might otherwise miss. In contrast, the time series model displays fluctuating accuracy across all precision levels, without a clear trend. This suggests greater sensitivity to noise, likely due to its reliance on temporal patterns that can be distorted by variability in precision. Its performance remains unstable, reflecting the complexity of learning sequential dependencies under noisy conditions

MS-DQI validation. In order validate the proposed MS-DQI framework, we combined internal consistency checks, and sensitivity analysis. First, we examined whether MS-DQI scores aligned with known technical events documented during the clinical study, such as early test interruptions, sensor malfunction, and device-specific defects. Tests with low availability or defective sensors consistently produced low MS-DQI values, confirming that the index correctly captures real failure modes. Second, we assessed the robustness of MS-DQI by analyzing its relationship with detection models performance using two baseline classifiers (Extra Trees and LSTM). We observed that decreases in MS-DQI—particularly in the availability, precision, and completeness dimensions—were directly associated with a measurable drop in classification accuracy, indicating that MS-DQI reflects quality characteristics that meaningfully affect predictive performance. Finally, device-level analyses revealed that MS-DQI differentiates reliably between strong and weak prototypes, reinforcing its ability to generalize across heterogeneous sensor hardware. Together, these results confirm that MS-DQI is both robust and aligned with practical data-quality variations occurred in medical sensor networks.

6. Limitations and Perspectives

Recognizing the limitations of our study is essential for providing a transparent evaluation of its contributions and for guiding future research directions. First, the current implementation of MS-DQI has been applied to a homogeneous temperature sensor network, where all sensors were of the same type and embedded in a wearable bra prototype. In real-world deployments, heterogeneous sensor networks are increasingly common, integrating physiological signals such as ECG, PPG, EMG, or motion tracking. Extending MS-DQI to such contexts will require adapting or redefining certain quality dimensions and introducing modality-specific metrics, while preserving the framework's three-step structure: dimension selection, weighting and aggregation. Second, the list of data quality dimensions used in this study is not exhaustive and some dimension definitions rely on expert interpretation, which introduces subjectivity. Future work could involve the formalization of these definitions through clinical consensus and cross-domain validation, particularly as MS-DQI is applied to new modalities. Third, MS-DQI currently aggregates multi-dimensional quality information into a single scalar index using weighted additive functions. While this representation is useful for ranking and thresholding, it may obscure critical variations across dimensions. A promising direction would be to explore multi-dimensional representations of data quality, such as radar plots, quality vectors, or surface visualizations, to better support nuanced decision-making in clinical or operational contexts. The weighting strategy employed in the case study is based on the Analytic Hierarchy Process (AHP), favored for its transparency and expert interpretability. However, we acknowledge that other approaches—such as entropy-based weighting or machine learning-derived feature importance—may provide complementary insights. While a full comparative analysis falls outside the scope of this study, we propose to explore these alternative schemes in future work and assess their impact on quality ranking and sensitivity. This comparison is especially relevant for large-scale deployments, where expert involvement may be impractical. In this study, we adopted a simple additive formulation because of its interpretability and alignment with existing multidimensional quality indices. However, alternative aggregation functions—such as geometric or nonlinear weighted operators—may provide different sensitivity profiles or stronger penalization of low-quality dimensions. A systematic comparison of these aggregation schemes was not conducted here, but represents a promising perspective for future work, especially for applications requiring stricter control of extreme or uneven data-quality degradation.. Fourth, the real-time applicability and computational scalability of MS-DQI are key strengths that support its use in practical deployments. The framework relies on matrix-based operations for both metric computation and aggregation, which are computationally lightweight and well-suited for implementation on embedded or edge-computing platforms. This makes MS-DQI scalable to high-frequency, high-volume data environments. Furthermore, the modular architecture allows for incremental and distributed computation, enabling the framework to operate continuously in streaming contexts. These properties enable real-time feedback and alert systems, where data quality can be monitored dynamically, and automated notifications can be triggered when thresholds are breached (e.g., sudden drops in availability, increasing noise levels, or unacceptable latency). Integrating MS-DQI into such systems would enhance the reliability and responsiveness of clinical and remote monitoring applications, and this represents a promising direction for future development. Fifth, the methodology was evaluated through a single case study involving a connected bra with temperature sensors for breast cancer monitoring. To establish the generalizability of MS-DQI, it is necessary to test it across a broader range of medical sensor systems as well as non-medical IoT networks which would demonstrate the framework's adaptability and robustness across domains. Finally, the case study dataset used in this research is relatively small, which may limit the generalizability of the results. Scaling the methodology to larger datasets will not only support more robust validation but will also allow for sensitivity analyses on various design choices (e.g., weighting methods, aggregation strategies, dimension selection). Such analyses are essential to ensure the reproducibility and stability of MS-DQI in real-world deployments.

7. Conclusion

In this study, we introduced MS-DQI; a comprehensive and adaptable methodology for assessing data quality in medical sensor networks. By systematically integrating relevant data quality dimensions, assigning context-aware weights and aggregating these into a unified index, MS-DQI provides a standardized and interpretable metric for evaluating sensor data reliability. We demonstrated the practical utility of the methodology in identifying low-quality data, assessing device performance and understanding the impact of data quality on downstream detection models through a real-world case study involving temperature sensors in a wearable device for breast cancer detection.

The methodology not only provides a standardized and interpretable metric for data quality assessment but also allows for customization across different sensor types and medical applications. Overall, MS-DQI establishes a critical link between theoretical data quality frameworks and the practical requirements of medical IoT applications, contributing to the development of more reliable, explainable and clinically applicable IoT health monitoring systems. By addressing the challenges associated with sensor data variability and environmental factors, MS-DQI stands as a promising tool for ensuring the accuracy and effectiveness of medical sensors. Moreover, future work should focus on adapting the framework to heterogeneous sensors, validating it on larger and more diverse datasets, integrating real-time alert systems and evaluating alternative weighting and aggregation methods. These directions will further reinforce its robustness, generalizability, and clinical relevance. Ultimately, MS-DQI contributes to the development of more reliable, explainable, and trustworthy medical IoT solutions.

Declaration of Interest Statement

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

Funding Information

This research was conducted as part of the PhD thesis of KETFI Raniya. The authors gratefully acknowledge the financial support provided by Grand Besançon Métropole (GBM), whose contribution was essential for the successful completion of this work. This work has also been achieved in the frame of the EIPHI Graduate School (contract "ANR-17-EURE-0002").

Author contribution

The authors contributed to this work as follows: R.K. led the original draft preparation, both R.K. and Z.A. were in charge of reviewing and editing the manuscript. The methodology and conceptualization were developed by Z.A. and R.K. Data curation and analysis were done by R.K. Visualizations were collaboratively done by R.K. and Z.A., and validation was performed by Z.A. Finally, Supervision, project administration, and funding acquisition were managed by Z.A., N.Z., and C.D.

Data Availability Statement

Data used in this paper is confidential and was collected under the clinical trials registered with ClinicalTrials.gov (Identifier: NCT05294016).

References

[1] N. Paton, Automating data preparation: Can we? should we? must we?, in: 21st International Workshop on Design, Optimization, Languages and Analytical Processing of Big Data, 2019.

[2] ISO 9000:2015, 3.6.2, Quality management: The path to continuous improvement, 2015. URL: `https://www.iso.org/quality-management#:~:text=Quality`.

[3] M. A. Jabin, K. Ahmed, M. J. Rana, B. K. Paul, M. Islam, D. Vigneswaran, M. S. Uddin, Surface plasmon resonance based titanium coated biosensor for cancer cell detection, IEEE Photonics Journal 11 (2019) 1–10. doi:`10.1109/JPHOT.2019.2924825`.

[4] L. L. Pipino, Y. W. Lee, R. Y. Wang, Data quality assessment, Communications of the ACM 45 (2002) 211–218.

[5] R. Y. Wang, D. M. Strong, Beyond accuracy: What data quality means to data consumers, Journal of management information systems 12 (1996) 5–33.

[6] L. Vanbrabant, N. Martin, K. Ramaekers, K. Braekers, Quality of input data in emergency department simulations: Framework and assessment techniques, Simulation Modelling Practice and Theory 91 (2019) 83–101.

[7] N. G. Weiskopf, C. Weng, Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research, Journal of the American Medical Informatics Association 20 (2013) 144–151.

[8] N. G. Weiskopf, S. Bakken, G. Hripcsak, C. Weng, A data quality assessment guideline for electronic health record data reuse, Egems 5 (2017) 14.

[9] A. P. Reimer, A. Milinovich, E. A. Madigan, Data quality assessment framework to assess electronic medical record data for use in research, International journal of medical informatics 90 (2016) 40–47.

[10] W. Kim, B.-J. Choi, E.-K. Hong, S.-K. Kim, D. Lee, A taxonomy of dirty data, Data mining and knowledge discovery 7 (2003) 81–99.

[11] Z. Wang, S. Dagtas, J. Talburt, A. Baghal, M. Zozus, Rule-based data quality assessment and monitoring system in healthcare facilities, in: Improving Usability, Safety and Patient Outcomes with Health Information Technology, IOS Press, 2019, pp. 460–467.

[12] A. Mavrogiorgou, A. Kiourtis, K. Perakis, S. Pitsios, D. Kyriazis, Iot in healthcare: Achieving interoperability of high-quality data acquired by iot medical devices, Sensors 19 (2019) 1978.

[13] S. Vadrevu, M. S. Manikandan, Real-time ppg signal quality assessment system for improving battery life and false alarms, IEEE transactions on circuits and systems II: express briefs 66 (2019) 1910–1914.

[14] J. Witteveen, P. Pradhapan, V. Mihajlović, Comparison of a pragmatic and regression approach for wearable eeg signal quality assessment, IEEE journal of biomedical and health informatics 24 (2019) 735–746.

[15] J. Xie, L. Peng, L. Wei, Y. Gong, F. Zuo, J. Wang, C. Yin, Y. Li, A signal quality assessment–based ecg waveform delineation method used for wearable monitoring systems, Medical & Biological Engineering & Computing 59 (2021) 2073–2084.

[16] A. Kristoffersson, M. Linden, A systematic review on the use of wearable body sensors for health monitoring: A qualitative synthesis, Sensors 20 (2020) 1502.

[17] Oracle, 2024. URL: `https://www.oracle.com/uk/internet-of-things/what-is-iot/`.

[18] H. Y. Teh, A. W. Kempa-Liehr, K. I.-K. Wang, Sensor data quality: A systematic review, Journal of Big Data 7 (2020) 11.

[19] M. M. Rosli, E. Tempero, A. Luxton-Reilly, Evaluating the quality of datasets in software engineering, Advanced Science Letters 24 (2018) 7232–7239.

[20] W. M. Ismael, M. Gao, Z. Yemeni, Esrra-iot: Edge-based spatial redundancy reduction approach for internet of things, Internet of Things 14 (2021) 100388.

[21] ISO 25000, Iso/iec 25012: Data quality model, 2024. URL: `https://iso25000.com/index.php/en/iso-25000-standards/iso-2501`, accessed: 2024-09-17.

[22] M.-T. Puth, M. Neuhäuser, G. D. Ruxton, Effective use of spearman's and kendall's correlation coefficients for association between two measured traits, Animal Behaviour 102 (2015) 77–84.

[23] O. Pollard, A Delphi Study Analysis of Best Practices for Data Quality and Management in Healthcare Information Systems, Ph.D. thesis, Walden University, 2019.

[24] Project Management Institute, Expert judgment, 2024. URL: `https://www.projectmanagement.com/wikis/344587/expert-judgment`, accessed: 2024-09-17.

[25] J. H. Buelvas, D. Munera, N. Gaviria, Dq-man: A tool for multi-dimensional data quality analysis in iot-based air quality monitoring systems, Internet of Things 22 (2023) 100769.

[26] S. Goyal, V. Deshpande, Comparison of weight assignment procedures in evaluation of environmental impacts, Environmental Impact Assessment Review 21 (2001) 553–563.

[27] M. G. Uddin, S. Nash, A. Rahman, A. Olbert, Development of a water qualit y index model-a comparative analysis of various weighting methods, in: Mediterranean Geosciences Union Annual Meeting (MedGU-21). Istanbul, 2022, pp. 1–6.

[28] A. Abdul Hameed M Jawad, A. Haider S, M. Bahram K, Application of water quality index for assessment of dokan lake ecosystem, kurdistan region, iraq, Journal of Water Resource and protection 2010 (2010).

[29] M. G. Uddin, S. Nash, A. Rahman, A. I. Olbert, A sophisticated model for rating water quality, Science of the Total Environment 868 (2023) 161614.

[30] R. W. Saaty, The analytic hierarchy process—what it is and how it is used, Mathematical modelling 9 (1987) 161–176.