IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 18 July 2025, accepted 11 August 2025, date of publication 20 August 2025, date of current version 15 September 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3601031

RESEARCH ARTICLE

A Robust Prioritization Framework of Data Quality Dimensions to Improve ML-Driven Healthcare Systems Using AHP and Sensitivity Analysis

AREEN METIB AL-HGAISH[1,2], (Member, IEEE), RODZIAH BINTI ATAN[1], RAZALI YAAKOB[3], AND MOHD HAFEEZ OSMAN[1], (Member, IEEE)

[1]Department of Software Engineering and Information Systems, Universiti Putra Malaysia, Serdang 43400, Malaysia
[2]Electrical Engineering Department, Al-Zaytoonah University of Jordan, Amman 11733, Jordan
[3]Department of Computer Science, Universiti Putra Malaysia, Serdang 43400, Malaysia

Corresponding author: Rodziah Binti Atan (rodziah@upm.edu.my)

This research was funded by Universiti Putra Malaysia, under the Geran Putra -Inisiatif Putra Siswazah (GP-IPS) (GP-IPS/2023/9748200).

ABSTRACT The success of Machine Learning (ML) models in healthcare relies heavily on the quality of data used. High-quality data are crucial for improving the predictive capabilities and the overall performance of ML systems. Despite this, research on data quality in healthcare and ML remains limited, with varying definitions of issues and dimensions across contexts. This study introduces a structured, expert-driven framework for prioritizing data quality dimensions critical to ML performance in healthcare. In contrast to performance evaluation studies involving machine learning algorithms or classifiers, this research does not encompass the training or comparison of predictive models. It addresses a key gap in ML data control by integrating ISO/IEC 25012 with Priestley's classification and using the Analytic Hierarchy Process (AHP) to evaluate 15 dimensions based on expert judgment. The findings identify Completeness (21.25%), Accuracy (15.53%), Consistency (14.31%), Currentness (14.82%), and Precision (13.82%) as the most influential dimensions for ML healthcare outcomes. A One-at-a-Time (OAT) sensitivity analysis with ±17.6% perturbation confirms the robustness of prioritization despite expert input variability. Key contributions include: 1) a tailored framework for ML healthcare data; 2) AHP-based dimension prioritization; 3) validation through sensitivity testing; 4) insights into data quality's impact on ML fairness and transparency; and 5) practical guidance for data governance and resource allocation. Future work will apply this framework to clinical datasets to validate its effectiveness in enhancing ML model performance and generalizability.

INDEX TERMS Data quality, machine learning, healthcare, data dimensions, predictive models, analytic hierarchy process, sensitivity analysis.

I. INTRODUCTION

Technological advancements and the growing availability of extensive datasets have propelled significant progress in Machine Learning (ML). This rapid evolution has notably impacted healthcare, offering unprecedented opportunities for innovation in clinical decision-making, patient management, and disease prevention. Despite these advancements, the success of ML models largely depends on the quality of data employed for training and evaluation. High-quality data are essential for developing robust models, while poor data can significantly impair model performance. This reliance on data quality is succinctly captured by the saying, "garbage in, garbage out." [1].

The associate editor coordinating the review of this manuscript and approving it for publication was Jad Nasreddine.

In the healthcare domain, the use of ML systems is rapidly increasing, with Electronic Health Record (EHR) databases providing numerous opportunities for secondary application of clinical data [2], [3], [4]. Such secondary applications span educational initiatives, clinical research, quality assurance, public health surveillance, resource allocation, and various commercial activities [5], [6]. Recent research has focused on aggregating EHR data and applying Artificial Intelligence (AI) techniques to create models that support decision-making, foster medical innovations, and achieve diverse objectives [7], [8].

ML models are now common in biomedical research and clinical practice. They support risk modeling, screening, diagnosis, therapy-response prediction, prognosis, and intensive care units (ICUs) mortality prediction [9], [10], [11]. However, the quality of training data affects the reliability and performance of the models. Data deficiencies, such as inaccuracies, incompleteness, and biases, can result in flawed ML predictions, directly leading to harmful clinical outcomes. Therefore, identifying and prioritizing the most impactful data quality dimensions is essential to enhancing ML reliability and protecting patient safety [12]. The success of ML systems in healthcare relies heavily on the quality of EHRs, which are crucial for enhancing clinical decisions and improving patient care.

Despite extensive research on healthcare data quality, a comprehensive understanding of the associated issues remains elusive [13]. Although significant progress has been made in applying ML to healthcare systems, a notable gap persists in studies addressing diverse dimensions of data quality [14]. This gap makes it difficult to identify and resolve data-quality challenges, highlighting the need for a structured approach to prioritize the most important dimensions. Furthermore, inconsistencies in terminology across studies exacerbate these challenges, complicating comparisons and discussions of data quality dimensions and impeding the resolution of specific issues [15], [16], [17], [18]. Resolving these inconsistencies is essential for establishing a shared vocabulary and developing effective strategies to improve data quality in ML-driven healthcare.

Another challenge is the absence of a standard framework to rank data-quality dimensions, which hampers healthcare organizations' ability to enhance ML outcomes. Previous research often addresses single dimensions or provides broad advice without identifying which dimensions are most important for healthcare ML. This fragmented and incomplete understanding highlights the urgent need for developing a well-structured framework. We therefore introduce an expert-driven, validated prioritization framework to guide resources toward the most impactful areas. Such a framework would help identify which data quality issues require immediate attention and resources, supporting better decision-making and more effective management of data quality challenges. This research is guided by the following primary research question:

- **Which data quality dimensions should be prioritized to enhance the performance and reliability of ML-driven healthcare systems?**

To support this inquiry, the study explores the following sub-questions:

- "How can the ISO/IEC 25012 data quality model be effectively adapted and restructured to address the specific needs of ML applications in healthcare?"
- "How can the AHP be used to prioritize these data quality dimensions based on expert input systematical?"
- "How reliable is the prioritization when evaluated using a context-driven sensitivity analysis method based on expert judgment variability?"

These questions collectively delineate the scope of this research and substantiate the chosen methodological approach. They facilitate the development of a rigorous prioritization framework specialized for ML-driven healthcare, harmonizing technical performance requirements with systematic, expert-driven assessment. Eventually, this approach will help focus on each data quality dimension directly and organize its related issues. This ensures a more coherent, targeted, and scientifically grounded initiative to enhance data quality, thereby improving ML performance in vital fields like healthcare systems.

Prior AHP/MCDM studies in healthcare usually focus on choosing technologies or features, not data-quality dimensions [70], [73], [78]. Likewise, research using AHP and fuzzy AHP for data quality evaluation has mainly been carried out in non-healthcare contexts, such as metadata assessment in open data portals [75] or public sector data analysis [77].

This study adopts the ISO/IEC 25012 standard [19], part of the system and software quality requirements and evaluation (SQuaRE) series of international standards, to standardize terminology and systematically address associated issues. This standard defines 15 data quality dimensions and is organized into three categories. Building on insights from Priestley's research on the data quality requirements for ML [20]. We restructured these 15 dimensions into four categories (Intrinsic, Contextual, Representational, and Accessibility), more aligned with ML applications, enabling greater clarity and practical relevance. This reclassification improves alignment with ML-driven healthcare systems. Using the Analytic Hierarchy Process (AHP), we structured these dimensions and categories hierarchically: the top level prioritizes the dimensions, the second level contains the categories, and the third level lists the dimensions, facilitating systematic evaluation and ranking of data quality challenges in ML.

The AHP-based approach highlights and prioritizes the data quality dimensions for ML-driven healthcare systems. This process establishes new benchmarks by minimizing bias and laying the foundation for reliable ML applications. To validate the stability of this prioritization, we conducted a one-at-a-time (OAT) sensitivity analysis, a method frequently utilized to evaluate stability and reliability in Multi-criteria Decision-making (MCDM) frameworks [21]. This study applies a ±17.6% perturbation to dimension weights, derived from expert-derived pairwise comparison matrices based on the standard deviation of aggregated group judgments at the category level.

While this research is limited to theoretical validation using expert input, our future studies will empirically apply and deeply evaluate these prioritized dimensions on a real healthcare dataset. Such implementation is expected to reveal the direct influence of improved data quality on the predictive model's performance of ML-driven healthcare systems. Incorporating systematic evaluations into regulatory approval processes could accelerate the approval of ML technologies and build trust in healthcare innovations.

This paper does not propose new ML algorithms nor evaluate model performance metrics. Instead, it concentrates on prioritizing data quality dimensions through decision-analytic techniques and experts' inputs. Empirical validation is beyond the scope of this study and is designated for future research.

The remainder of this paper is organized as follows. Section II highlights the novelty and key contributions of this study. Section III provides preliminary knowledge and defines the core data quality concepts. Section IV reviews the related literature and background, establishes the research context, and identifies gaps in existing studies. Section V explains the methodology, including the adoption of ISO/IEC 25012 standards, re-categorization of data quality dimensions, and application of the AHP. Section VI presents the discussion and results, focusing on the prioritization of dimensions and their implications for ML performance. Section VII addresses contributions, limitations, and directions for future research. Finally, Section VIII concludes the paper.

II. STUDY NOVELTY AND KEY CONTRIBUTIONS

This study establishes a systematic framework for prioritizing data quality dimensions in ML-driven healthcare systems. It addresses a critical gap caused by lacking a structured, ML-specific model for evaluating data quality. By integrating the ISO/IEC 25012 standard, a data quality model, with Priestley's categorization of ML-relevant data quality dimensions, this study provides an adaptive model to address the challenges and demands inherent in ML-driven healthcare environments. The ISO/IEC 25012 standard, initially developed for traditional software systems, was strategically restructured into four ML-relevant categories (Intrinsic, Contextual, Representational, and Accessibility) to enhance clarity and alignment with healthcare ML requirements.

Expert evaluations conducted through detailed pairwise comparisons by 14 experienced domain specialists in ML, data quality, and healthcare systems substantially enhanced the framework's comprehensiveness and practical relevance. The AHP provided methodological rigor and precision in ranking and prioritizing the 15 identified dimensions. Additionally, the robustness of the AHP-based prioritization was rigorously validated through a sensitivity analysis employing the OAT perturbation technique, involving ±17.6% adjustments derived from expert judgments.

The outcomes revealed a distinct prioritization boundary, marked by a significant 9.7% gap between the fifth- and sixth-ranked dimensions, Precision and Efficiency. This gap demonstrates strong expert consensus around the top five dimensions: Completeness, Accuracy, Consistency, Currentness, and Precision. These five dimensions consistently maintained their positions among the top ranks across all 120 sensitivity analysis scenarios, reflecting high stability and resilience to variations in expert input.

Meanwhile, the middle-level dimensions exhibited moderate ranking changes but did not exceed the fifth rank, while the lowest-level dimensions remained entirely unchanged across all scenarios. This differentiated behavior reinforces the reliability of the proposed prioritization and supports a tiered approach to improving data quality.

These findings provide empirical justification for focused investment in the top-priority dimensions and highlight the framework's robustness as a generalizable and decision-supportive tool. Furthermore, this study contributes a validated reference point for future ML-driven healthcare systems seeking reliable, interpretable, and performance-focused data quality interventions by isolating sensitivity-tested, high-impact dimensions.

Key contributions include:

- Establishing the first structured and expert-validated prioritization framework for ML-driven healthcare data quality needs.
- Developing an adaptive framework that explicitly aligns data quality dimensions with the requirements of ML-driven healthcare systems.
- Clarifying and standardizing terminology facilitates consistent and meaningful comparisons across studies, enabling more effective data quality evaluations.
- Providing actionable insights for healthcare organizations to enhance data quality and directly improve ML model accuracy, reliability, and interpretability.
- Establishing a methodology for addressing multifaceted data quality issues, bridging theoretical gaps, validating the stability of prioritized dimensions, and offering practical guidelines for future implementation.
- Introducing a sensitivity-confirmed prioritization threshold that distinguishes the top five critical data quality dimensions, ensuring strategic focus for resource allocation in healthcare ML applications.
- Demonstrating the robustness of AHP-based prioritization under expert uncertainty using a validated OAT perturbation method, supporting reproducibility and reliability in real-world settings.

This research lays the foundation for future studies, setting a benchmark for data quality evaluation in ML applications and paving the way for more hardy and efficient healthcare solutions.

III. PRELIMINARY KNOWLEDGE

This section clarifies the essential data quality concepts vital for understanding their impact on the performance of ML models in sensitive areas such as healthcare. It lays the conceptual groundwork for the development and prioritization framework presented in this study and supports the reader in contextualizing the methodological decisions.

- **Data Quality:** It refers to the data state defined by various attributes that assess its suitability for operational use, decision-making, and strategic planning. According to Wang and Strong [17], data quality is characterized by "fitness for use," highlighting the importance of evaluating data against its intended purpose to ensure that it meets specific application needs. High data quality is essential in healthcare settings, where incorrect or incomplete data can lead to life-critical decision errors.
- **Data Quality Dimensions:** These specific and measurable characteristics define the quality of data, such as accuracy, consistency, and other relevant dimensions. These dimensions break down data quality into actionable components, enabling the precise identification of areas that require improvement to enhance data performance [22]. These serve as evaluation criteria to determine if the data is sufficient for use.
- **Data Quality Categories:** Data quality can be categorized to illustrate how data dimensions are managed, handled, and utilized. These categories encompass various aspects of data quality assessment and improvement, reflecting the different ways in which data can be structured and maintained [23]. Effective data management involves assessing data content's relevance to user needs and expectations.
- **Data Quality Issues:** Issues regarding data arise when they fail to meet standards set by relevant dimensions or categories. Common problems include inaccuracies, missing values, inconsistent formatting, and outdated information [24]. Such issues can introduce noise, bias, or inefficiency into ML models, severely limiting their predictive performance and reliability. Addressing these issues is crucial for improving the data quality, thereby enhancing the accuracy and effectiveness of models and analyses based on data [25]. Addressing data quality challenges is vital for maintaining model reliability and ensuring accurate analytical results. Recent reviews have emphasized the significant impact of data issues, such as missing data and imputation strategies, highlighting the broader implications of data completeness and quality on predictive model performance [26].

To establish a solid foundation for improving data quality in ML-driven healthcare systems, this study operationalized these concepts through a structured literature review and applied the AHP to evaluate the relative importance of each dimension. This study enables future targeted improvements in data management strategies by proposing a framework grounded in established standards and validated through expert judgment. Addressing these fundamental principles is a prerequisite to enhancing the performance, interpretability, and safety of ML models deployed in real-world healthcare environments.

IV. LITERATURE REVIEW AND BACKGROUND

This section covers the core concepts, recent developments, and key methodological trends at the intersection of ML and healthcare. It situates the current research within the broader academic discourse by analyzing a diverse body of literature that explores data quality dimensions, ML model reliability, and decision-making frameworks. While prior studies have offered valuable insights, many have tackled data quality in a fragmented or narrowly focused manner, often addressing isolated issues or specific ML applications without providing a unified prioritization strategy.

This study employs a structured, holistic approach to address fragmentation by integrating the AHP. This provides a way to rank data quality dimensions, highlighting critical attributes to improve ML performance in healthcare. Unlike previous methods lacking quantifiable prioritization, it offers a transparent, expert-driven evaluation.

This review synthesizes prior work's key conceptual challenges and themes, merging theory with practice. It summarizes established findings and offers a cohesive framework for enhancing ML outcomes through targeted data quality improvements. Combining rigor with relevance, it builds on existing literature while addressing critical gaps.

A. MACHINE LEARNING IN HEALTHCARE

The adoption of ML in healthcare has emerged as a groundbreaking development that harnesses advanced computational capabilities and extensive datasets to transform medical practices. Early predictive models, such as the Framingham risk score introduced in 1967, set the stage for sophisticated algorithms used today [27]. Furthermore, AI plays a pivotal role in medical imaging, assisting in diagnostics and emphasizing its essential contribution to contemporary healthcare.

ML applications also extend to EHRs, where they produce valuable insights that improve patient risk-scoring systems, forecast disease onset, and optimize hospital operations [28]. As these systems process increasing volumes of data, their algorithms evolve, resulting in continuous improvement in healthcare applications. Building on this progress, recent studies have introduced methods to maintain and enhance the quality of medical drug data by leveraging clustering techniques within a data lake environment. This approach optimizes data utility for medical decision-making by offering alternative drug recommendations for patient needs [29], [30]. Alzyadat et al. [31] highlight clustering techniques to uncover hidden relationships in structured datasets, facilitating more efficient decision-making processes. Integrating k-means clustering with data pre-processing methods improved the interpretability and stability of drug-related datasets and advanced healthcare analytics.

ML addresses toxicity-related challenges in drug delivery systems by proposing a GAN-augmented CNN model capable of generating synthetic microscopic images to train toxicity classification systems. This innovation was designed to identify toxic nanocarriers during controlled and targeted drug delivery, effectively addressing the limitations posed by scarce experimental datasets [32].

Recent advances in healthcare decision-making have utilized sophisticated decision-making frameworks. For instance, Akhtar et al. developed a novel IoT-based methodology using fractional fuzzy Hamacher aggregation operators, enhancing the precision and reliability of healthcare selection processes through aggregation techniques [33]. This trend highlights the increasing complexity and precision required in healthcare-related decision support systems.

AI models achieved over 91% accuracy in diagnosing lung diseases from chest X-rays using preprocessing, segmentation (LinkNet), and CNN classifiers like DenseNet201, demonstrating ML's potential for quick, dependable respiratory screening [34]. Similarly, Al-Yousef et al. enhanced early breast cancer detection with a fusion of MFFNN, LDA, SVM, and KNN via majority voting, improving diagnostic accuracy in BI-RADS mammography systems [35]. AI and ML have improved predictions for health emergencies, disease conditions, and immune responses. Despite some skepticism regarding their practical applications and interpretations, the use of ML in healthcare is rapidly expanding [36]. The advent of platforms such as PyTorch, DeepLearning4J, TensorFlow, and Keras has facilitated the development and application of these algorithms, making them increasingly accessible for various clinical purposes [37]. ML's capabilities extend beyond diagnostic support and enhance decision-making processes. AI applications have improved case triage, image scanning and segmentation, and disease risk prediction [38], [39], [40]. In addition, AI techniques uncover complex relationships within medical data that are challenging to express using traditional methods [41].

Hybrid algorithms have demonstrated remarkable efficacy in improving disease prediction. A multilayered hybrid algorithm (MLHA) combining supervised and unsupervised learning techniques, such as SVM, random forest, and k-means clustering, integrated with XGBoost in a two-layer architecture, achieved superior accuracy in Type II diabetes classification. This demonstrates the potential of hybrid ML systems to address complex medical-prediction challenges [42].

ML contributes to clinical decision making while reinforcing the principles of evidence-based medicine. By identifying hidden risk factors and gaps in healthcare, ML improves the precision of risk scores and assists the healthcare sector in managing risks more effectively [43], [44]. This technology enables the integration and analysis of larger datasets and supports decision-making processes with minimal human intervention. However, the application of ML in healthcare presents several challenges. A major hurdle is the availability of high-quality training and testing data sets. Accurate and reproducible ML predictions depend on large and reliable datasets. Healthcare data often suffer from issues such as incompleteness, heterogeneity, and an imbalance between data richness and sample sizes, which complicate the development and interpretation of ML models. These challenges must be addressed carefully because of the sensitive nature of healthcare data [28]. As ML continues to advance in the healthcare sector, it promises more rapid diagnosis and improved patient care. However, the field must overcome the complexities related to data quality and algorithmic reliability to fully harness ML's potential for ML in healthcare, which is partly addressed in this research.

B. DATA QUALITY IN ML-DRIVEN HEALTHCARE SYSTEMS

The quality of training data is vital for the efficiency, accuracy, and complexity of ML-driven healthcare tasks. Poor data quality can cause faulty conclusions and reduce AI effectiveness, which relies on large datasets for model development. ML has the potential to greatly improve patient care by enhancing diagnostics, optimizing treatments, and streamlining administrative tasks [45]. However, despite improving ML models, a significant gap exists in enhancing data quality. A thorough understanding of the dataset is essential for its effective utilization; insufficient comprehension can lead to inaccurate analyses and unreliable decision-making. Recent research highlights the negative impact of biased data on research outcomes. Biases in EHR data can distort ML and AI models, resulting in skewed results that adversely affect patient care [46], [47]. Such biases can lead to incorrect risk assessments or treatment recommendations [48], [49], perpetuating systemic disparities in healthcare at individual and population levels. Addressing these biases requires a structured framework that focuses on data quality, which is critical for ML-driven healthcare systems [50], [51]. Bias in EHR data highlights broader data quality issues that can arise from errors in data entry, documentation inaccuracies, problems with EHR software, or barriers to accessing care. EHR data often encompasses a wide range of elements, including laboratory results, vital signs, demographics, medications, and medical history [52], [53]. Healthcare data often exhibit a range of challenges, including heterogeneity, temporal variations, spatial variations, sparsity, incompleteness, noise, irregularities, and inaccuracies [54].

One prevalent issue in clinical research is missing data, which occurs when variable values are not recorded for every subject. Common strategies for handling missing data include complete-case analysis, in which missing data records are excluded from the analysis, and mean imputation, in which the average of the available data substitutes missing values. However, these techniques can introduce biases [55]. Missing data can be categorized into three types: "missing completely at random," where the probability of missing data is unrelated to any measured or unmeasured variables; "missing at random," where the likelihood of missing data is related to observed variables but not to the missing values themselves; and ''not missing at random,'' where the probability of missing data is related to the unobserved value itself [56].

The growing complexity of healthcare datasets, fueled by advancements in memory and computing power, presents additional challenges. Datasets often become imbalanced because of the prevalence of rare events among numerous cases in the majority class, compounded by the multiclass nature of clinical problems and various diagnosis codes. Inter-disease heterogeneity further complicates dataset analysis. This bias towards the majority class can adversely impact ML training models, potentially leading to suboptimal care for patients in minority classes [57].

Moreover, outliers in the healthcare data pose significant challenges. ML algorithms designed to detect unusual physiological readings can facilitate rapid emergency intervention and provide new insights into health conditions. For instance, Edin [58] assessed five ML algorithms, including two unsupervised and three supervised techniques, for their ability to identify anomalies in heart-rate data.

The diversity of healthcare data stems from the complex and varied nature of the information generated by medical services and research activities. The continuous evolution of medical terminology, the extensive volume of data produced through automated processes, and the need for thorough data analysis to support decision-making emphasize the critical role of effective health system computerization and knowledge management [59].

Data quality encompasses multiple concepts, including accuracy, validity, reliability, completeness, readability, timeliness, accessibility, and confidentiality [15], [60]. It can be compromised at various stages, such as data collection, coding, and standardization, and is affected by a range of technical, organizational, behavioral, and environmental factors [61]. Problems such as manual processes, data diversity, duplication, and data collection and processing errors can render health data inadequate for researchers, administrators, and healthcare professionals [62]. Maintaining high data quality is vital for supporting informed decision-making, enhancing service delivery, and producing reliable health status evidence, improving patient care [63]. Inadequate data quality can adversely affect continuity of care [64], patient safety [65], and research outcomes [66].

C. DIMENSIONS AND CATEGORIES OF DATA QUALITY IN ML-DRIVEN HEALTHCARE

This study utilized the ISO/IEC 25012 standard from the SQuaRE series to address data quality challenges in ML applications within healthcare. This internationally recognized standard offers a comprehensive framework and detailed quality dimensions, including Accuracy (ACCY), Credibility(CRD), Traceability (TRC), Completeness (CMP), Currentness (CUR), Efficiency (EFF), Understandability (UND), Consistency (CON), Precision (PRC), Recoverability (RCV), Availability (AVL), Portability (PORT), Accessibility (ACCS), Confidentiality (CONF), and Compliance (COMP) [19]. This standard provides a model for assessing the data quality in various technological contexts.

The ISO/IEC 25012 model was chosen because of its broad acceptance and proven effectiveness in standardizing data quality dimensions, which are crucial for the success of ML models in various healthcare applications. Poor data quality can severely impact ML system performance, leading to inaccurate predictions and suboptimal clinical decisions [67], [68].

Despite advancements in ML technologies, there is a notable gap in comprehensive studies that address data quality dimensions. The variability in terminology across existing studies complicates the comparisons and understanding of these dimensions, leading to confusion and inconsistencies in their applications [15], [16], [17], [18] Employing ISO/IEC 25012, this study aims to clarify these terminologies and provide a standardized framework for effectively prioritizing critical data quality dimensions.

To refine our approach further, we integrated Priestley's classification [20], which categorizes data into four distinct categories: Intrinsic (INT), Contextual (CTX), Representational (REP), and Accessibility (ACC). This classification is relevant to the complexities of ML applications in healthcare in terms of handling data quality issues.

- Intrinsic: This category focuses on data accuracy, origin, and cleanliness, emphasizing the true value, traceability, and error-free state of data. This is essential for ML applications where data lineage and auditability are crucial for interpreting and verifying ML outcomes.
- Contextual: Pertains to the relevance and appropriateness of data for its intended use in specific healthcare ML tasks. It examines whether the data is suitable, complete, timely, or relevant.
- Representational: This area focuses on how data are presented, ensuring clarity, interpretability, and standardization across different systems.
- Accessibility: This encompasses the ease of data access and use underpinned by security and legal frameworks to ensure data protection and privacy compliance.

The incorporation of the ISO/IEC 25012 standard, along with the adapted classification from Priestley, enables a thorough evaluation of the data quality dimensions that are crucial for the effective operation of ML systems in healthcare. This approach, combined with AHP, allows for methodological assessment and prioritization of the importance of these dimensions. Table 1 presents the categories of data quality dimensions distributed across the 15 ISO/IEC 25012 dimensions.

D. THE ANALYTIC HIERARCHY PROCESS (AHP)

Developed by Thomas L. Saaty in the 1970s, AHP is a systematic mathematical framework for solving complex decision-making problems [69] It functions by organizing

TABLE 1. Categorization and description of data quality dimensions adapted for ML-driven healthcare systems.

Category	Dimension	Description
INT	ACCY	This refers to how well healthcare data reflects true patient information, vital signs, and other critical clinical metrics without errors within ML systems.
	CRD	The level of trustworthiness and believability of healthcare data sources, which ML systems utilize for making clinical predictions.
	TRC	The ability to track and audit the origin, transformation, and usage of healthcare data within ML systems for accountability.
CTX	CMP	This denotes the degree to which data is fully represented, ensuring that all necessary data points are included and that the dataset is comprehensive enough to support the specific tasks of ML models.
	CUR	The timeliness of healthcare data reflects its relevance and recency for ML processing and decision-making.
	EFF	The degree to which data is processed efficiently in terms of resource usage.
REP	UND	The clarity with which healthcare data is presented makes it interpretable by algorithms and human users.
	CON	The degree of uniformity and coherence of healthcare data ensures that ML systems receive consistent inputs.
	PRC	The granularity of data is exactly or described without rounding errors in healthcare data, enabling ML systems to enhance performance.
	RCV	The capacity to restore healthcare data in its entirety and accuracy for ML use after any disruptive events.
ACC	AVL	The readiness and reliability of healthcare data for use by ML systems at any required time, ensuring continuous operational capability.
	PORT	The flexibility of healthcare data to be transferred and used across different ML platforms and systems without quality loss.
	ACCS	The ease with which healthcare data can be accessed by authorized personnel and ML systems, considering various user roles and requirements.
	CONF	The extent to which patient data is kept secure from unauthorized access within ML environments, upholding privacy and ethical standards.
	COMP	The degree to which healthcare data adheres to legal standards, ethical guidelines, and data protection regulations relevant to ML applications.

decisions into a hierarchical structure of goals, criteria, sub-criteria, and alternatives. AHP's power lies in its pairwise comparison method, which converts qualitative judgments into quantitative values, thereby streamlining the decision-making process.

Recently, the AHP has been increasingly recognized for its significance in healthcare, where complex decision-making processes are multifaceted. This assists decision-makers in prioritizing health policies and allocating resources to manage healthcare services effectively [70], [71], [72]. Furthermore, the selection of ML algorithms, a critical aspect of advancing healthcare technologies, has also benefited from the structured decision-making provided by AHP [73].

The relevance of AHP extends to the realm of ML in healthcare, particularly in optimizing algorithms and model selection. By applying AHP, researchers can prioritize various features and parameters in ML models, ensuring that the most significant predictors of health outcomes are considered [74]. This is particularly useful in scenarios where the dimensions of data quality must be assessed and ranked to improve the performance of the predictive models.

The applicability of AHP extends beyond healthcare, impacting various sectors such as the semiconductor industry and public sector data intelligence. Research has utilized AHP, including its variants such as fuzzy AHP, to evaluate metadata quality in open data portals, assess data quality dimensions in the semiconductor sector, and identify success factors for implementing data intelligence in the public sector. These diverse applications emphasize AHP's adaptability in handling uncertain and imprecise data and its ability to enhance decision-making by prioritizing crucial elements such as project management, information systems, and data quality [75], [76], [77].

Moreover, AHP has proven valuable in strategic prioritization across AI-related fields. For example, a recent study used AHP to identify and rank critical factors in AI-driven drug discovery, demonstrating its effectiveness in handling complex multi-criteria decisions involving algorithm performance, interpretability, and data quality [78]. These studies reflect the increasing adoption and methodological rigor of structured MCDM techniques like AHP in decision-intensive areas such as healthcare and ML.

E. SENSITIVITY ANALYSIS IN PRIORITIZATION FRAMEWORKS

Sensitivity analysis validates decision-making frameworks like AHP by assessing how small input changes affect criteria rankings, such as expert comparisons. This procedure is integral in high-risk applications like ML-driven healthcare, where decision accuracy directly impacts clinical outcomes.

In high-risk environments like healthcare, even minor weight changes assigned by experts can alter the prioritized criteria, which may affect the accuracy of predictive models. Research has shown that minor adjustments in the weights of top-level categories can influence the ultimate criteria ranking [79], [80], [81]. These findings confirm that sensitivity analysis is essential for ensuring the reliability of prioritization outcomes, particularly when frameworks are built upon expert input, which may naturally vary due to subjective perspectives or domain-specific knowledge.

Recent studies emphasize the necessity of sensitivity testing in fuzzy and MCDM models. For instance, Alballa et al. [82] incorporated sensitivity analysis into an enhanced CODAS method to assess the model's stability under various weight scenarios, demonstrating its effectiveness in location decision-making. Similarly, Rahim et al. [83] utilized a quasi-rung orthopair fuzzy COPRAS model and conducted weight variation analysis to validate the outcomes of supplier selection. These contributions highlight sensitivity analysis as a crucial validation step in prioritization frameworks, highlighting its significance in confirming the reliability and consistency of results across different domains.

In line with these established practices, recent advances in methodological research [84], [85], [86] have recommended the use of the OAT perturbation method as a preferred and effective sensitivity analysis technique within the context of AHP. This method involves an organized process where one weight input is individually modified at a time, while simultaneously adjusting the other weights proportionally to ensure that the sum of weights remains unchanged. This process allows researchers to assess how isolated changes in a single input affect rankings. This structured perturbation enhances transparency by revealing the decision process under different uncertainties, aiding in model validation and testing.

These studies demonstrate that integrating sensitivity analysis is the best practice and a crucial step in modern prioritization methods. It enhances severity through systematic assessment of result stability and reliability. Moreover, it improves transparency in expert-driven decisions, assisting stakeholders in understanding the impact of weights and assumptions. Eventually, this practice promotes more trustworthy and practical solutions, especially in critical fields like healthcare, logistics, and intelligent systems, where precision and transparency are crucial.

V. METHODOLOGY

This study employed the AHP methodology to rank and prioritize data quality dimensions critical to ML-driven healthcare systems. The process began with identifying and structuring relevant data quality dimensions based on the ISO/IEC 25012 standard, ensuring a comprehensive and internationally recognized foundation. Subsequently, the selected dimensions were reorganized into application-specific categories tailored to meet the needs of ML applications in healthcare. Finally, AHP was used to gather expert input, assess consistency, compute local and global weights, and produce a final prioritized ranking of dimensions based on their evaluated significance.

To confirm the robustness of these rankings and mitigate potential bias from expert subjectivity, a sensitivity analysis was conducted using the OAT perturbation method. This approach evaluated the stability of dimension rankings under $\pm 17.6\%$ variations in category-level weights, validating the reliability of the prioritization framework under realistic uncertainty.

Figure 1 presents an overview of the proposed methodology flowchart that encapsulates the key stages and processes involved in achieving the research objectives. The stages are organized as follows:

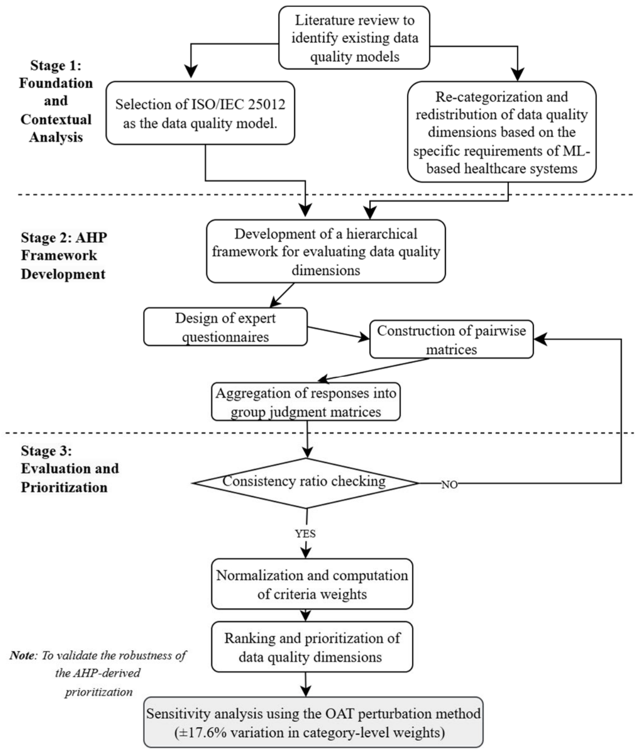

FIGURE 1. Flowchart of the AHP-driven data quality dimension prioritization methodology and sensitivity validation.

A. FOUNDATION AND CONTEXTUAL ANALYSIS

1) LITERATURE REVIEW

An in-depth examination of existing studies, data quality dimensions, and models was conducted to identify suitable data quality dimensions for ML-driven healthcare systems. This process involved analyzing recent research efforts to understand how AHP and sensitivity analysis have been utilized in prioritization tasks across various fields, including healthcare, AI systems, supplier selection scenarios, etc. The review aimed to identify methodological gaps and limitations in current practices. Furthermore, it emphasized the significance and relevance of the approach taken in this study by illustrating how similar techniques have been applied and the challenges encountered in various contexts. This thorough examination provided a strong foundation for the study's methodology and highlighted the current research on data quality prioritization in complex systems.

2) SELECTION OF ISO/IEC 25012

This study selected data quality dimensions based on the ISO/IEC 25012 model, which provides a comprehensive framework for evaluating data quality. Since a specific model for data quality in ML applications has not yet been developed, ISO/IEC 25012 is a foundational reference for identifying and prioritizing the essential dimensions affecting ML-driven models in healthcare settings. All 15 dimensions of ISO/IEC 25012 were retained to ensure comprehensive coverage of the domain. These dimensions formed the basis for a structured expert evaluation and subsequent prioritization using AHP.

3) RE-CATEGORIZATION OF DATA QUALITY DIMENSIONS

This study focuses on the structured organization of data quality dimensions into pertinent categories. The foundational work by Priestley et al. [20] We identified four main categories of data quality requirements suitable for ML applications and customized these categories to address the demands of ML-driven healthcare systems: intrinsic, contextual, representational, and accessibility.

This re-categorization was instrumental in establishing the hierarchical structure needed by the AHP technique and enabled subsequent testing through category-based perturbations. This adaptation is important because it enhances effective data management and utilization, particularly in predictive tasks. Furthermore, it serves as a functional bridge between traditional data quality standards and the evolving needs of ML systems. Figure 2 displays these categories and their corresponding dimensions, offering a systematic approach to structuring data quality dimensions.

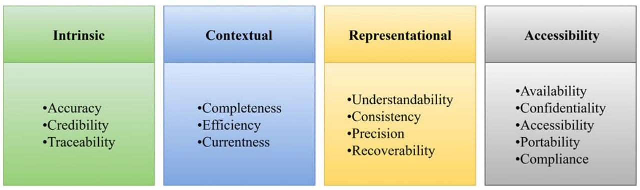

FIGURE 2. Categories of data quality dimensions.

B. ANALYTIC HIERARCHY PROCESS (AHP) FRAMEWORK DEVELOPMENT AND DIMENSION EVALUATION

In these stages, we employed the AHP to prioritize and rank the importance of data quality dimensions. The primary objective was to establish a validated and expert-driven ranking that reflects the dimensions' relative importance in ML-driven healthcare applications. Figure 3 depicts the AHP methodological steps, which are elaborated in the following subsections.

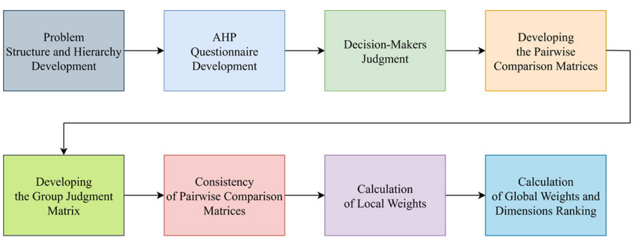

FIGURE 3. Steps of the AHP methodology.

1) HIERARCHY DEVELOPMENT AND PROBLEM STRUCTURE

This step involves constructing an appropriate hierarchical structure for the AHP, which includes defining the objectives, criteria, and sub-criteria. In this study, AHP methodology was employed to prioritize and rank the data quality dimensions. Consequently, the hierarchy created features, such as the overarching objective, various categories (criteria), and specific dimensions (sub-criteria).

The primary goal of employing AHP is to prioritize and rank the data quality dimensions relevant to ML-driven healthcare systems. Within this structure, the top level represents the main goal, the second level includes the categories, and the third level lists the individual data quality dimensions. The hierarchical arrangement used in this study is illustrated in Figure 4.

TABLE 2. The scale of importance for pair-wise comparisons [87].

Linguistic Term	Value
Equally importance	1
Equally to moderately importance	2
Moderately importance	3
Moderately to strong importance	4
Strongly importance	5
Strongly to very strongly importance	6
Very strongly importance	7
Very strongly to extremely important	8
Extremely importance	9

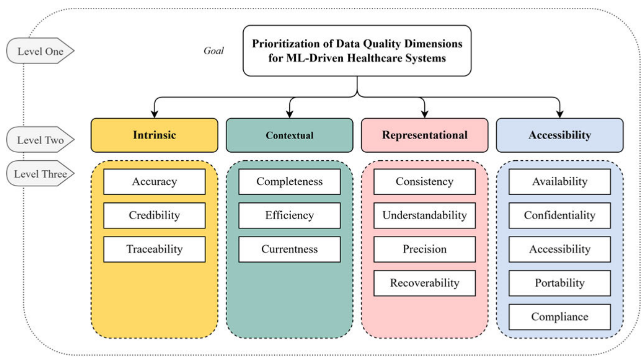

FIGURE 4. Structure of the AHP hierarchy for prioritizing data quality dimensions in ML-driven healthcare systems.

2) AHP QUESTIONNAIRE DEVELOPMENT

In this step, academic experts conducted pairwise comparisons between categories and dimensions of ML applications in healthcare settings. To assign the pair-wise comparison values, a nine-point scale ranging from "1. This is equally important" to "9. This is extremely important," as shown in Table 2.

This study developed a questionnaire utilizing Saaty's nine-point scale [87] to facilitate pairwise comparison assessments. The questionnaire was divided into five sections: the first addressed pairwise comparisons between categories, whereas the remaining four focused on pairwise comparisons and evaluations among different dimensions. Appendix presents the AHP questionnaire used in the study.

TABLE 3. Expert panel profile and affiliation.

No.	Role/Title	Affiliation	Specialization / Research Interests
1	Professor	Department of Information Technology, King Abdulaziz University, Saudi Arabia	Data Quality, Machine Learning, Health Informatics
2	Associate Professor	Faculty of Computer and Information Technology, Jordan University of Science and Technology, Jordan	Artificial Intelligence, Data Mining, Big Data, Healthcare Data Analytics
3	Senior Lecturer	Department of Computer Engineering, University of Engineering and Technology, Pakistan	Deep Learning, Medical Image Processing, Data Quality
4	Professor	Faculty of Computer and Information Science, University of Ljubljana, Slovenia	Data Management, Big Data Analytics, Health Informatics, AI
5	Associate Professor	Department of Informatics, University of Évora, Portugal	Decision Support Systems, Healthcare Information Systems, ML
6	Associate Professor	Department of Electrical Engineering, IT and Cybernetics, University of South-Eastern Norway	Data Governance, Machine Learning, Data Security
7	Professor	Department of Information Systems, University of the Western Cape, South Africa	Big Data, Machine Learning in Healthcare Systems, Knowledge Discovery
8	Senior Lecturer	Department of Information Technology, Kwame Nkrumah University of Science and Technology, Ghana	Data Analysis, Artificial Intelligence, Health Informatics
9	Associate Professor	Faculty of Information Technology, Al-Zaytoonah University, Jordan	Artificial Intelligence, Data Mining, Data Quality, Big Data, Drug Data
10	Professor	Faculty of Information Technology, University of Babylon, Iraq	Medical Informatics, Data Quality, Cloud Computing in Healthcare
11	Senior Lecturer	Department of Health Informatics, Jordan University of Science and Technology, Jordan	Health Information Systems, Data Quality in Electronic Health Records
12	Senior Lecturer	Department of Information Technology, University of Information Technology and Communications, Iraq	Machine Learning, Health Informatics, Cloud/IoT-based Healthcare
13	Professor	Department of Information Systems, Palestine Technical University, Palestine	Information Technology, Technology Adoption, Data Quality in E-Health
14	Associate Professor	King Abdullah II School of Information Technology, University of Jordan, Jordan	Business Models, Big Data, Data Science, Artificial Intelligence, Health Informatics

3) DECISION-MAKERS JUDGMENT

The questionnaire was distributed to experts from diverse academic institutions to collect their perspectives on the importance of data quality dimensions in ML applications and healthcare. Among the recipients, 14 experts with experience in ML, data science, and healthcare systems provided comprehensive responses. This collective expertise ensures a well-rounded evaluation of data quality dimensions, considering all relevant aspects of the data quality challenges in ML-driven healthcare systems. Given that AHP is not a statistical technique, it does not require a statistically significant sample size, as adopted from [88]. AHP is designed to concentrate on the decision-making process rather than the demographic representativeness of the respondents [89]. Shrestha et al. [90] noted that AHP does not require large sample sizes, because it targets knowledgeable individuals in the relevant area. Table 3 outlines the detailed professional profiles, institutional affiliations, and areas of specialization of the experts who participated in the evaluation. Their diverse geographical representation and multidisciplinary backgrounds, including data quality, ML, and health informatics, enhance the robustness and generalizability of the findings.

4) DEVELOPING THE PAIRWISE COMPARISON MATRICES

Using the expert feedback gathered through the questionnaire, pairwise comparison matrices were developed for the categories and dimensions. In these matrices, comparisons between two categories or dimensions are represented using integer values according to Saaty's nine-point scale (see Table 2). When a category or dimension in a row i is deemed more significant than the one in column j, a corresponding integer value a_{ij} is assigned based on the level of importance. Its reciprocal, $a_{ji} = \frac{1}{a_{ij}}$, is placed in the symmetric position, reflecting inverse importance. When two categories or dimensions have equal importance, both a_{ij} and a_{ji} are assigned a value of 1. Moreover, all diagonal elements (a_{ii}) are also set to 1, as each category or dimension is equally important when compared to itself. Thus, each pairwise comparison matrix is a positive reciprocal matrix. The general structure of a pairwise comparison matrix $A = [a_{ij}]$ is represented as follows:

$$A = [a_{ij}] = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1m} \\ a_{21} & 1 & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & 1 \end{bmatrix} = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1m} \\ 1/a_{12} & 1 & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ 1/a_{1m} & 1/a_{2m} & \cdots & 1 \end{bmatrix}$$

where:

A is the pairwise comparison matrix.

m represents the number of categories or dimensions being compared.

$\boldsymbol{a_{ij}}$ is the relative importance value of dimension or category i overdimension or category j, as judged by experts.

TABLE 4. Random indexes for random matrices [87].

n	3	4	5	6	7	8	9	10	11	12	13
RI(n)	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49	1.51	1.54	1.56

$\boldsymbol{a_{ji}}$ is the reciprocal of a_{ij} and represents the inverse importance.

This reciprocal relationship is formally expressed as:

$$a_{ji} = {}^{1}/_{a_{ij}};\ i, j = 1, 2, \cdots m. \tag{1}$$

5) CONSISTENCY OF PAIRWISE COMPARISON MATRICES

Evaluating the consistency of expert judgments is critical to ensuring reliable prioritization. Saaty [87] developed a consistency measure known as the Consistency Index (CI), which quantifies the logical coherence of pairwise comparison matrices larger than 2 × 2.

The CI is calculated using the following equation:

$$CI = \frac{\lambda_{max} - n}{n - 1} \tag{2}$$

where:

$\boldsymbol{\lambda_{\max}}$ is the largest eigenvalue of the pairwise comparison matrix, representing the degree of internal consistency.

$\boldsymbol{n}$ denotes the size of the comparison matrix (i.e., the number of categories or dimensions evaluated).

To standardize consistency evaluation, the Consistency Ratio (CR) is subsequently calculated by comparing the computed CI to a corresponding Random Index (RI). The RI is an average CI derived from randomly generated reciprocal matrices of similar size, defined as:

$$CR = \frac{CI}{RI\ (n)} \tag{3}$$

where:

CR is the Consistency Ratio, which indicates the overall consistency of the judgments.

RI (***n***) is the Random Index, varying according to matrix size, as listed in Table 4.

According to Saaty [87], a matrix with a Consistency Ratio of 0.10 or less is considered acceptable. Matrices with CR > 0.10 indicate inconsistent judgments, necessitating a reevaluation by experts. In this study, the consistency of expert-generated matrices was thoroughly checked, yielding CR values ranging from 0.00040 to 0.0102. This demonstrates a high degree of internal coherence and confirms the reliability and credibility of the expert judgments.

6) DEVELOPING THE GROUP JUDGMENT MATRICES

Combining consensus judgments from the individual evaluations provided by experts requires a combination of pairwise comparison matrices. AHP offers two principal methods for deriving a group matrix: the first involves aggregating individual judgments, and the second focuses on aggregating individual priorities [91]. In this study, we opted for the first method, which compiles a pairwise comparison matrix for each expert to construct a group-judgment matrix. This aggregation was achieved by using the geometric mean method [87]. The geometric mean aggregation method reduces potential biases by incorporating every expert's input equally. This aggregation is mathematically represented by Equation (4):

$$[a_{ij}] = \left(\prod_{k=1}^{N} a_{ij}^{(k)}\right)^{\frac{1}{N}} \tag{4}$$

where:

$\boldsymbol{a_{ij}}$ is the aggregated group judgment for the comparison between elements i and j.

$\boldsymbol{a_{ij}^{(k)}}$ represents the individual judgment of expert k regarding the importance of element i over element j.

$\boldsymbol{N}$ denotes the total number of experts who provided judgments.

7) CALCULATION OF LOCAL WEIGHTS

Following the construction of the group judgment matrices, the next step involved determining local weights for categories and dimensions. Local weights quantify the relative importance of each category or dimension within its immediate grouping. The local weights were derived through a normalization process, forming a normalized matrix N, ensuring values' comparability within each matrix column. The normalization of a matrix N was performed using Equation (5):

$$N = \left[N_{ij}\right]; \left[N_{ij}\right] = \frac{a_{ij}}{\sum_{i=1}^{m} a_{ij}} \tag{5}$$

where:

$\boldsymbol{N}$ is the normalized matrix.

$\boldsymbol{N_{ij}}$ represents the normalized value corresponding to the pairwise comparison between element i and element j.

$\boldsymbol{a_{ij}}$ denotes the aggregated judgment value from the group judgment matrix.

m indicates the number of elements (categories or dimensions) being evaluated.

The local weights were subsequently calculated by taking the arithmetic average of the normalized values in each matrix row. This method reflects the proportionate contribution of each element within its category or dimension set. Equation (6) illustrates the calculation of these local weights:

$$w_i = \frac{\sum_{j=1}^{m} N_{ij}}{m} \tag{6}$$

where:

$\boldsymbol{w_i}$ is the local weight of element i, representing its relative priority within its category.

$\boldsymbol{N_{ij}}$ are elements of the normalized matrix N.

$\boldsymbol{m}$ again, refers to the number of elements within the comparison matrix.

8) CALCULATION OF GLOBAL WEIGHTS AND DIMENSIONS RANKING

Global weights signify the overall significance of each dimension within the entire hierarchy. Although the local weight of each category naturally reflects its global weight, the global weights of individual dimensions are obtained by combining their local weights with those of their parent categories. This hierarchical combination is formally expressed in Equation (7):

$$GWD = LWD \times LWC \tag{7}$$

where:

GWD is the global weight of a given dimension, representing its priority across the full hierarchy.

LWD denotes the local weight of the dimension, as computed previously.

LWC signifies the local weight of the category to which the dimension belongs.

Computed global weights enable ranking the dimensions comprehensively, identifying the most critical dimensions for targeted improvements in ML-driven healthcare data quality.

C. EVALUATING THE STABILITY OF AHP RESULTS USING SENSITIVITY ANALYSIS

The reliability of rankings derived from AHP is paramount for the strength of the prioritization framework presented in this research. A consistency check was conducted first on the pairwise comparison matrices utilizing Saaty's consistency ratio (CR). This process guaranteed that all expert evaluations remained internally consistent and logically sound, with CR values staying below the recommended threshold of 0.10, thus validating the credibility of the expert assessments.

However, consistency alone does not entirely guarantee stability under realistic variations. Therefore, to further reinforce the validity and stability of these results, we conducted an additional sensitivity analysis using the OAT perturbation method. Sensitivity analysis assesses how slight modifications in expert input or judgment weights might influence the final dimension rankings, thereby offering essential validation for the strength and applicability of the findings.

This test provides empirical support for the stability of dimension prioritization by evaluating whether realistic fluctuations of this scale could meaningfully alter the rankings. This step is essential for confirming that prioritized dimensions remain stable, reliable, and practically significant, even amid expected and reasonable variations in expert evaluations, thereby enhancing insights and further solidifying the practical value and reliability of the proposed prioritization framework. Figure 5 illustrates the methodological workflow of the sensitivity analysis in this study, highlighting the key stages in assessing the AHP outcomes from establishing the baseline to perturbation scenarios, recalculating global weights, re-ranking dimensions, and the process concludes with a final evaluation of sensitivity.

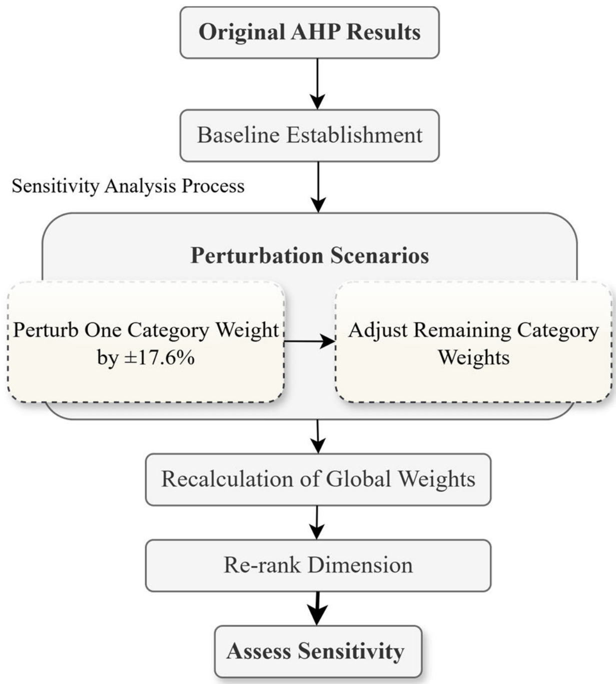

FIGURE 5. Workflow of the one-at-a-time sensitivity analysis.

1) RATIONALE FOR SENSITIVITY ANALYSIS

The sensitivity analysis function evaluates MCDM methods by examining the reliability of results derived from techniques such as AHP.

This study's importance lies in prioritizing data quality in ML-driven healthcare applications. This focus is essential for directing resource allocation, influencing research pathways, and setting future goals to refine and improve predictive models used in healthcare data analysis. Even minor changes in the weights assigned to various categories can lead to significant shifts in the rankings of these dimensions. Consequently, conducting a thorough sensitivity analysis is of paramount importance. It guarantees that established priorities remain consistent, trustworthy, and credible, regardless of reasonable variations in expert judgments. Given the present emphasis on ML within healthcare research, this process addresses a significant gap in the field, as reliable prioritization can influence model development and enhance the effectiveness of clinical decision-making.

2) SENSITIVITY ANALYSIS DESIGN AND TECHNIQUE

This analysis applies a ±17.6% perturbation, derived from the standard deviation of aggregated expert judgments at the category level. This particular value effectively encapsulates the variability and uncertainty present in expert inputs. Given that previous studies recommend perturbations ranging from 5% to 20% for sensitivity testing [80], [81], [92], [93], [94], a perturbation of 17.6% strikes an appropriate balance between practical relevance and methodological rigor, thereby enhancing the ecological validity of stability checks.

This methodology varies one category at a time and rescales the others to keep the total at 1.0.

3) PERTURBATION AND COMPUTATIONAL PROCEDURES

In each scenario, we changed one category by ±17.6%. The perturbed weight of a designated category, denoted as $W_m^{perturbed}$ perturbed, is computed utilizing Equation (8) in the following manner:

$$W_m^{perturbed} = W_m^{original} \mp (0.176 \times W_m^{oriiginal}) \tag{8}$$

where:

$\boldsymbol{W_m^{perturbed}}$ is the adjusted (perturbed) weight of the selected category m.

$\boldsymbol{W_m^{original}}$ is the original AHP-derived weight of the category m.

±0.176 (17.6%) is the perturbation factor based on the standard deviation of expert judgments.

Subsequently, the weights of the remaining categories were recalculated proportionally to ensure that the total weight remains at 1.0. This recalculation employed Equation (9):

$$W_i^{adjusted} = \left(1 - W_m^{perturbed}\right) \times \frac{W_i^{original}}{1 - W_m^{original}} \tag{9}$$

where:

$\boldsymbol{W_i^{adjusted}}$ is the recalculated weight of an unaffected category i.

$\boldsymbol{W_i^{original}}$ is the original weight of the unaffected category i.

$\boldsymbol{W_m^{original}}$ and $\boldsymbol{W_m^{perturbed}}$ are defined as in Equation (8).

Subsequent to the modification of category-level weights, the global weights for each dimension were re-evaluated utilizing Equation (10), thereby preserving the structural coherence of the AHP hierarchy:

$$GW_{dimension} = LW_{dimension} \times CW_{category}^{perturbed} \tag{10}$$

where:

$\boldsymbol{GW_{dimension}}$ is the newly recalculated global weight of each dimension.

$\boldsymbol{LW_{dimension}}$ is the dimension's local weight within its respective category.

$\boldsymbol{CW_{category}^{perturbed}}$ is the perturbed or adjusted category-level weight obtained from Equations (8) and (9).

4) IMPLEMENTATION AND ANALYSIS STEPS

The sensitivity analysis in this study was conducted utilizing Python programming to ensure accuracy, transparency, and reproducibility. The following computational steps succinctly summarize the procedure for conducting the sensitivity analysis:

- Baseline Establishment: Set the original AHP global weights as the basis for comparison.
- Perturbation Scenarios: Adjust each category's weight independently by ±17.6% (see Equation 8).
- Weight Normalization: Adjust the weights of unaffected categories proportionally (see Equation 9) to maintain a total weight of 1.0.
- Global Weights Recalculation: Integrate the perturbed category weights into the local dimension weights (see Equation 10) to recalculate the global dimension weights.
- Ranking Comparison: Re-rank the dimensions based on the recalculated global weights and compare these rankings to the original rankings, thereby assessing sensitivity and robustness.

The incorporation of consistency checks and OAT sensitivity analysis ensures the methodological integrity of the AHP-based prioritization framework proposed in this study. The consistency ratio confirms the logical coherence of expert judgments. In contrast, the sensitivity analysis, utilizing a justified perturbation of ±17.6% derived from the variability in expert inputs, validates the stability and resilience of the resultant rankings. This dual-validation strategy reinforces confidence in the prioritization of data quality dimensions. It enhances the solidity and applicability of the proposed framework in real-world ML-driven healthcare environments. Having established a definite methodological foundation, the subsequent section presents and discusses the empirical results of this study.

VI. DISCUSSION AND RESULTS

This study handles the importance of prioritizing data quality dimensions in ML-driven healthcare systems by utilizing AHP, a structured decision-making approach, to evaluate and rank the various dimensions of data quality. Furthermore, the study includes a comprehensive sensitivity analysis to examine how variations in prioritization influence the overall decision-making process and outcomes, thereby ensuring a robust and reliable assessment of the importance of data quality in this context.

A. AHP-DERIVED PRIORITIZATION OF DATA QUALITY DIMENSIONS

The study established a hierarchical model structured across three separate levels aimed at prioritizing the data quality dimensions paramount to ML-driven healthcare systems. The top level defined the overarching objective of prioritizing data quality dimensions, the second level comprised dimension categories (Intrinsic, Contextual, Representational, Accessibility), and the third level examined individual data quality dimensions.

To gather comprehensive and diverse expert perspectives, a structured questionnaire was disseminated to specialists from various academic institutions, thereby ensuring broad representation across disciplines relevant to ML, healthcare informatics, and data science. Fourteen specialists provided complete evaluations of each category and dimension related to the study's primary objective. These responses enabled the formulation of pairwise comparison matrices in

TABLE 5. Weights and ranks for data quality categories and dimensions.

Category	Category Weight %	Dimension	Local Weight %	Global Weight %	Ranking
INT	20.8%	ACCY	74.7%	15.5%	2
		CRD	12.1%	2.5%	9
		TRC	13.1%	2.7%	8
CTX	38.7%	CMP	54.1%	20.9%	1
		EFF	9.9%	3.8%	6
		CUR	36%	13.9%	4
REP	33%	UND	7.6%	2.5%	10
		CON	43.2%	14.3%	3
		PRC	40.8%	13.5%	5
		RCV	8.5%	2.8%	7
ACC	7.5%	AVL	20.5%	1.5%	13
		CONF	25.3%	1.9%	11
		ACCS	18.4%	1.4%	14
		PORT	13.4%	1.0%	15
		COMP	0.224%	1.7%	12

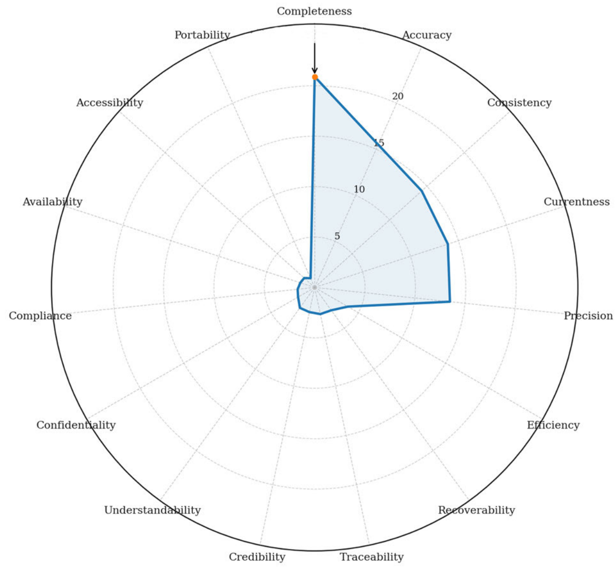

FIGURE 6. Radar plot showing global weights of data quality dimensions in ML-driven healthcare systems.

TABLE 6. Consistency test and group judgment matrix for categories.

	INT	CTX	REP	ACC	Consistency Test
INT	1	0.471	0.570	3.467	λ_{max}= 4.0276 n= 4
CTX	2.119	1	1.219	4.342	CI= 0.0092
REP	1.753	0.820	1	4.211	CR= 0.0102
ACC	0.288	0.230	0.237	1	

accordance with the standard AHP procedure as illustrated in Equation (1). Subsequently, the consistency ratio (CR) for these matrices was rigorously evaluated utilizing Equations (2) and (3), thus ensuring logical coherence in the judgments rendered by the experts.

Upon verifying the internal consistency of each expert's responses, Equation (4) was employed to generate aggregated group judgement matrices for all categories and dimensions. All resulting matrices exhibited satisfactory consistency levels (CR values below the 0.10 threshold), thus confirming the reliability and internal coherence of the aggregated expert assessments. In order to calculate accurate local and global weights, these group judgement matrices were normalized using Equation (5), while Equations (6) and (7) were applied to determine their respective local and global weights.

Figure 6 offers a quick overview of the weights across the fifteen dimensions, emphasizing the dominance of Completeness, Accuracy, Consistency, Currentness, and Precision. Table 5 summarizes the calculated weights and final rankings, illustrating the relative significance of each data quality category and dimension based on experts' cumulative assessments. Moreover, Tables 6 through 9 include comprehensive results of the consistency evaluations and group judgment matrices for each hierarchical category and dimension.

The analysis first identified the Contextual category as the most influential, representing about 38.7% of the overall category weights. This stresses a consensus among experts regarding the paramount role of contextual dimensions in impacting the performance of ML-driven healthcare systems. The Representational category followed closely as the second most influential, with a total weight of 33.0%. Intrinsic ranked third at 20.8%, with Accuracy as its key dimension, reflecting the urgent need for error-free data. Finally, Accessibility had the lowest aggregate weight at 7.5%, indicating that dimensions within this category (Confidentiality, Compliance, Portability, Accessibility, and Availability) are strategically important but have a less immediate impact on ML performance. However, these dimensions are crucial for maintaining compliance and ethical standards, thereby fostering robust and equitable ML implementations in the long term.

As shown in Table 6, the group judgment matrix for these categories satisfied the consistency requirements, with a Consistency Ratio (CR) of 0.0102, which is well below the accepted threshold of 0.10. This low CR confirms the internal consistency and reliability of the aggregated expert judgments.

To improve interpretability, the category-level pairwise comparisons are visualized as a heatmap in Figure 7.

According to the results, the weights within the contextual category indicated that completeness had the highest priority, with a local weight of 54.1%. This was followed by currentness, which had a local weight of 36%, and efficiency, which had a local weight of 9.9%.

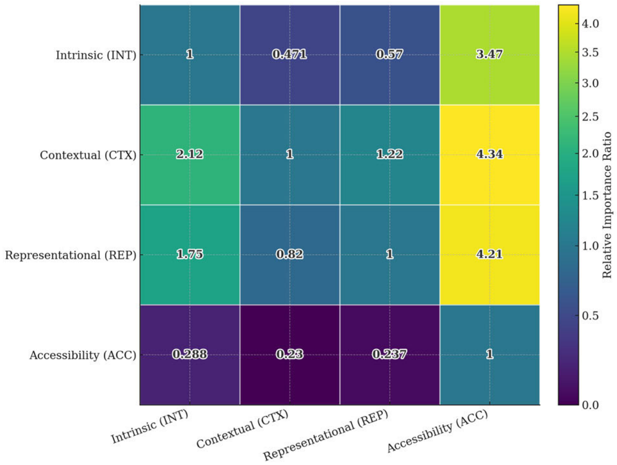

FIGURE 7. Figure 6. Pairwise comparison matrix (Categories) heatmap.

TABLE 7. Consistency test and group judgment matrix for contextual dimensions.

	CMP	EFF	CUR	Consistency Test
CMP	1	5.037	1.634	λ_{max}= 3.0071
EFF	0.198	1	0.252	n=3
CUR	0.611	3.965	1	CI= 0.00353 CR= 0.00609

Table 7 provides further evidence supporting the consistency of the dimension evaluations within the Contextual category, showing a consistency ratio (CR) of 0.00609. This CR value is significantly below the standard threshold of 0.1, reaffirming strong coherence and logical consistency among expert judgments in this category.

These results highlight that experts agree that completeness is essential for covering all necessary aspects needed to achieve reliable and effective ML predictions. The substantial emphasis on currency underlines the need for continuously updating and maintaining data relevance, which is crucial in fast-paced clinical environments. While the lower focus on efficiency is recognized as beneficial, it suggests that efficiency depends on the context and is fairly less critical for immediate model performance, thus reflecting the experts' agreement on prioritizing fundamentally impactful dimensions.

In the representational category, the findings indicated that consistency had the highest priority with a local weight of 43.2%. This was followed by precision, which had a local weight of 40.8%. recoverability and understandability with local weights of 8.5% and 7.6%, respectively, received less focus than consistency and precision.

Table 8 demonstrates that the group judgment matrix for the representational category complies with consistency standards, achieving a consistency ratio (CR = 0.00040, which is below 0.1).

These results highlight the pivotal role of consistency, emphasizing its importance in ensuring that data remains uniform and reliable across various systems and contexts. Precision is similarly valued, emphasizing the need for accurate and error-free data. This prioritization suggests that ensuring data uniformity and accuracy is essential for practical ML applications, with recoverability and understandability as supportive dimensions.

TABLE 8. Consistency test and group judgment matrix for representational dimensions.

	UND	CON	PRC	RCV	Consistency Test
UND	1	0.181	0.190	0.844	λ_{max}= 4.0011
CON	5.506	1	1.081	5.179	n= 4 CI= 0.00036
PRC	5.253	0.924	1	5.030	CR= 0.00040
RCV	1.184	0.193	0.198	1	

TABLE 9. Consistency test and group judgment matrix for intrinsic dimensions.

	ACCY	CRD	TRC	Consistency Test
ACCY	1	6.791	5.146	λ_{max} = 3.0099
CRD	0.147	1	1.020	n=3
TRC	0.194	0.979	1	CI= 0.00493 CR= 0.00850

The results for the intrinsic category revealed that accuracy was of utmost importance and priority, with a local weight of 74.7% (as shown in Table 5). Credibility and traceability followed with lower local weights of 12.1% and 13.1%, respectively. The group judgment matrix pertaining to the intrinsic dimensions, as illustrated in Table 9, has satisfied the AHP consistency check, resulting in a consistency ratio (CR) of 0.00850, which is beneath the 0.1 threshold. This notably low CR affirms the consensus among experts and supports the reliability of the pairwise comparisons conducted within this category. The focus on accuracy aligns with clinical expectations, where error-free data is essential for generating reliable predictions, guiding interventions, and ensuring patient safety. While credibility and traceability are valued, they have been given a lower priority, suggesting that although the trustworthiness of sources and the ability to audit data changes are important, they are secondary to the primary requirement of data accuracy.

In the accessibility category, the expert panel designated confidentiality as the most significant dimension, obtaining a local weight of 25.3% (as illustrated in Table 5). Compliance was ranked next with a weight of 22.4%, followed by accessibility with 18.4%, availability at 20.5%, and portability at 13.4%. These findings reflect a clear consensus among experts regarding prioritizing data protection and legal adherence over infrastructural or convenience-related considerations. The group judgment matrix of the Accessibility category (refer to Table 10) has demonstrated full adherence to AHP consistency standards, culminating in an exceptionally low consistency ratio (CR = 0.00062). This consistency reinforces the reliability of expert judgments and the internal coherence within this category, which, although

TABLE 10. Consistency test and group judgment matrix for accessibility dimensions.

	AVL	CONF	ACCS	PORT	COMP	Consistency Test
AVL	1	0.865	1.077	1.467	0.924	λ_{max}=5.0028
CONF	1.154	1	1.455	1.951	1.104	n=5
ACCS	0.928	0.686	1	1.345	0.848	CI= 0.00069
PORT	0.681	0.512	0.743	1	0.589	CR= 0.00062
COMP	1.082	0.905	1.178	1.695	1	

lower-weighted, remains of ethical and operational significance.

The paramount importance of confidentiality accentuates its essential function in safeguarding sensitive patient data, particularly within the framework of ML systems that utilize vast datasets, which present potential privacy risks. This emphasis is consistent with international regulations such as the Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GDPR), which mandate rigorous protections in managing health data. Furthermore, the considerable emphasis placed on compliance highlights the necessity of ensuring that ML-driven systems operate within institutional and legal parameters, thereby mitigating risk and enhancing ethical accountability.

Despite being ranked lower, the remaining dimensions, accessibility, availability, and portability, contribute to essential system-level functions such as ensuring authorized data access, minimizing service disruptions, and facilitating data integration across platforms. Their lower priority in this study does not imply irrelevance; instead, it reflects their more supportive and indirect influence on the immediate performance of ML models, as perceived by experts.

The results indicate that, although accessibility infrastructure is paramount for preserving operational continuity, protecting privacy and assuring regulatory compliance are imperative in data governance for ML healthcare systems. This emphasis affords healthcare practitioners and system designers essential insights for the strategic allocation of resources aimed at enhancing data quality, which, in turn, fosters the ethical and secure implementation of AI.

Subsequent to computing local weights across all four categories, we evaluated and ranked the data quality dimensions on a global scale. As shown earlier in Table 5 and illustrated in Figure 8, global weights were determined by multiplying the local weight of each dimension by the corresponding weight of its category. This final ranking of the AHP results synthesizes expert focus on dimension-level attributes and their categorical importance.

AHP facilitated the systematic and rigorous prioritization of fifteen critical data quality dimensions categorized under four categories. Figure 9 below illustrates a heatmap resembling a confusion matrix that visually summarizes the global weights and rankings derived from the AHP analysis.

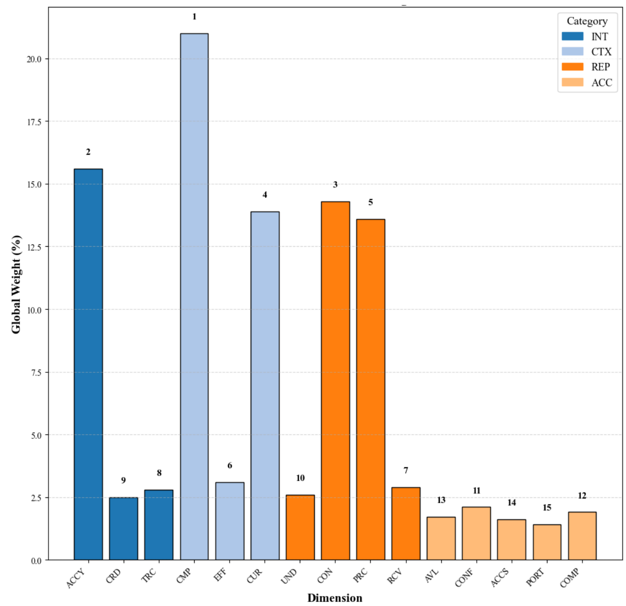

FIGURE 8. Global weights and rankings of data quality dimensions for ML-driven healthcare systems.

These rankings are based on comprehensive expert judgements and reflect the strategic importance of each dimension in enhancing ML-driven healthcare systems. The highest-ranked dimension was Completeness (20.9%) from the Contextual category, emphasizing its vital role in capturing all essential data elements necessary for ML performance. Close behind are Accuracy (15.5%) from the Intrinsic category and Consistency (14.3%) from the Representational category, highlighting the required need for precise and uniformly structured data.

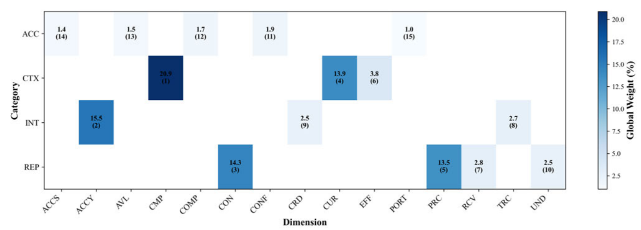

FIGURE 9. AHP results confusion matrix-like heatmap.

The dominance of Completeness, Accuracy, Consistency, Currentness, and Precision at the top of the rankings reflects their explicit and measurable influence on the technical performance and clinical reliability of ML models. These dimensions directly address the stringent data requirements of healthcare predictive modeling, where gaps, inaccuracies, inconsistent formats, outdated records, or insufficient granularity can immediately undermine diagnostic accuracy, risk stratification, and treatment planning. Their high weights indicate that experts intentionally prioritized dimensions capable of reducing algorithmic bias, improving generalizability, and improving performance exactness.

In contrast, dimensions within the Accessibility category, such as Confidentiality (1.9%) and Compliance (1.7%), received lower global weights, reflecting their indirect yet fundamental roles in ensuring ethical, compliant, and secure ML deployment. Besides, Portability, while operationally important, was de-emphasized because it affects system implementation more than predictive performance accuracy. This ranking pattern shows a pragmatic prioritization strategy where experts focused on dimensions that have the most immediate and significant impact on algorithmic outputs in real-world clinical settings.

B. EVALUATING THE STABILITY OF AHP PRIORITIZATION USING SENSITIVITY ANALYSIS

To verify the robustness and generalizability of the AHP-derived rankings, a sensitivity analysis was performed utilizing the OAT perturbation method. An adjustment of ±17.6% was applied to the weights of each top-level category, a rate ascertained from the standard deviation of expert inputs. This perturbation is consistent with practices established in the MCDM literature and reflects realistic variability in expert judgment.

The principal objective of this analysis was to assess whether minor fluctuations in category weights would substantially disrupt the global prioritization of dimensions. Considering the hierarchical nature of the AHP, modifications at the category level have the potential to propagate throughout the system, particularly influencing dimensions situated near ranking boundaries. Consequently, evaluating the sensitivity of the prioritization model is imperative to guarantee its practical reliability. Although sensitivity analysis is widely used in AHP studies, our unique contribution lies in its domain-specific adaptation and data-driven perturbation design. First, the analysis is applied to a full set of ISO/IEC 25012 data quality dimensions restructured for ML-driven healthcare, which is rarely explored in previous AHP research. Second, the perturbation factor (±17.6%) is not chosen at random but is based on the standard deviation of aggregated expert judgments at the category level, ensuring the test reflects realistic variability in the target domain. Third, the analysis is performed at categories and then spread to dimension levels, allowing detection of stability patterns throughout the entire AHP hierarchy. This integrated, context-aware approach offers a more rigorous and practically relevant stability assessment than traditional AHP sensitivity tests.

The results demonstrate that the top five dimensions exhibited remarkable stability throughout all perturbation scenarios, with none of the rankings falling below fifth place. This substantiates that their prioritization is merely a product of a singular weighting configuration and indicates a persistent expert consensus. These dimensions maintained their priority rankings across all eight perturbation scenarios. This is evidenced in Figure 10, where, even amid substantial category-level weight fluctuations, no dimension fell below the fifth position, and Completeness consistently upheld the highest rank in all instances, except one (Contextual −17.6%), during which it momentarily descended to the second position. This consistency level reinforces these dimensions' significance in facilitating effective ML outcomes within healthcare applications.

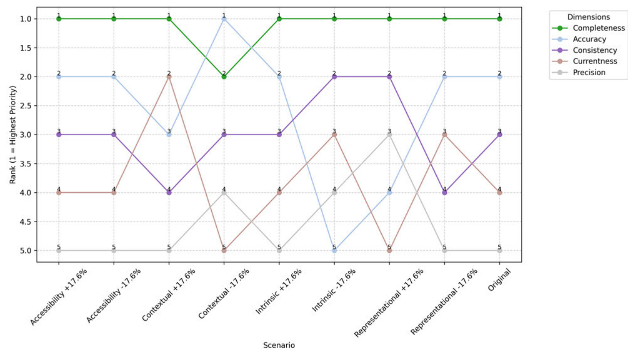

FIGURE 10. Sensitivity analysis results for top-ranked dimensions.

Accuracy, while generally stable in the second rank, showed a drop to the fourth position in scenarios where the weight of the Intrinsic category was reduced, highlighting its reliance on the specific weight of that category. Consistency and Currentness demonstrated occasional fluctuations between the third and fourth ranks, yet they remained among the highest priorities throughout all simulations, reinforcing their mutual significance. Precision consistently occupied the fifth position in most scenarios, evidencing a strong consensus among experts regarding its ranking.

As illustrated in Figure 11, the middle-ranked dimensions, Efficiency, Recoverability, Traceability, Credibility, and Understandability, exhibited moderate sensitivity. Efficiency, positioned sixth, proved to be the most stable within this classification, remaining unchanged across all perturbations. Recoverability and Traceability alternated between the seventh and eighth positions, contingent upon shifts within the Intrinsic and Contextual categories. Credibility and Understandability exhibited greater volatility, occasionally relegating to the tenth position in response to heightened emphasis on certain categories, indicating their secondary, context-specific relevance. Although not consistently ranked at the top, this cohort possesses contextual significance for various ML applications, including auditability, real-time recovery, and interpretability within clinical support systems.

Figure 12 presents the outcomes regarding the lowest-ranked dimensions: Confidentiality, Compliance, Availability, Accessibility, and Portability. These dimensions exhibited complete stability across all perturbation scenarios, remaining unchanged from their initial rankings. This observation highlights the minimal weight assigned to the Accessibility category and the prevailing consensus among experts regarding their indirect impact on the immediate performance of ML models. Although these dimensions do not directly influence predictive accuracy, they are vital for the legal, ethical, and operational frameworks, facilitating interoperability,

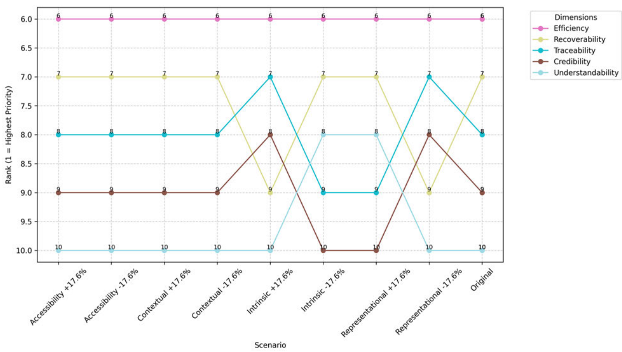

FIGURE 11. Sensitivity analysis results for middle-ranked dimensions.

ensuring regulatory compliance, and providing secure access to ML-driven healthcare systems.

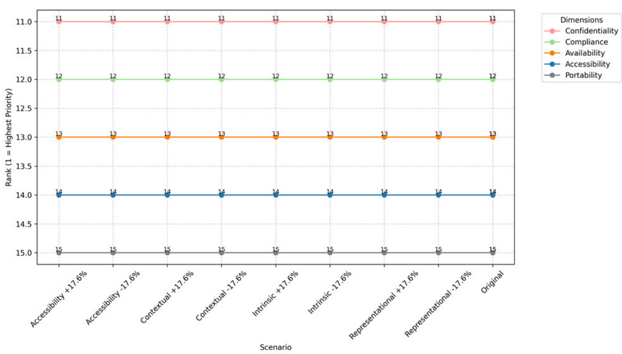

FIGURE 12. Sensitivity analysis results for lowest-ranked dimensions.

The combined visualization in Figure 13 shows a clear priority stratification: strong consensus and high stability among top-ranked dimensions; moderate variability among mid-level dimensions; and complete stability but lower criticality in the lowest-ranked group.

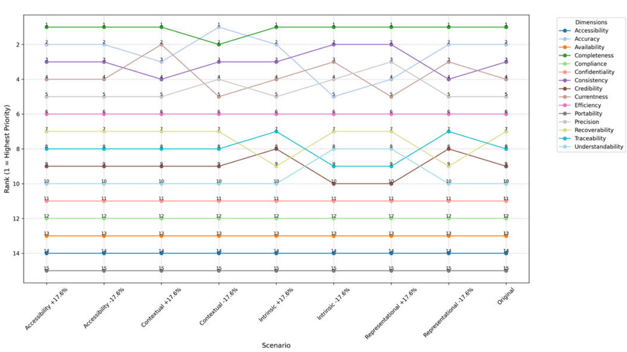

FIGURE 13. Sensitivity analysis of all dimensions.

The sensitivity analysis results classified the fifteen data quality dimensions into three levels based on their global weights derived from the AHP. These levels are categorized as top-ranked, middle-ranked, and lowest-ranked, as illustrated in Figures 10, 11, and 12. The following will discuss these dimensions obtained from AHP at each of these levels.

Top-Ranked Dimensions Discussion (1–5)

1) **Completeness** (Rank 1, Global Weight: 20.9%) achieved the highest global weight, emphasizing its paramount significance for effective ML implementation in the healthcare sector. Situated within the Contextual category, which itself received the highest weight, this dimension highlights the necessity for healthcare datasets to encompass all pertinent patient variables. Incomplete records can compromise ML model performance, potentially resulting in erroneous clinical decisions and missed diagnoses.
2) Accuracy (Rank 2, Global Weight: 15.5%) ranked second, emphasizes the criticality of aligning healthcare data with precise real-world observations. High-quality, error-free data ensures reliable ML outputs and cultivates trust in clinical decision-making. In contrast, inaccuracies can lead to errors in diagnostic predictions and therapeutic choices, adversely impacting patient outcomes.
3) **Consistency** (Rank 3, Global Weight: 14.3%) highlights the necessity for uniform data representation across healthcare platforms. Discrepancies in data, such as varying coding standards or inconsistent measurement techniques, may compromise the accuracy and interpretability of ML models, negatively affecting clinical decisions derived from integrated datasets. Sensitivity analysis revealed moderate rank shifts between Consistency and Currentness, primarily influenced by modifications in the Contextual and Representational categories. Nonetheless, Consistency was consistently prioritized above fourth place, indicating overall stability.
4) **Currentness** (Rank 4, Global Weight: 13.9%) is rated fourth; it highlights the critical need for timely and updated data within healthcare ML systems. Outdated information can undermine clinical efficacy, particularly in acute care scenarios such as emergency medicine or intensive care units, where real-time decision-making is essential. The sensitivity analysis indicated that Currentness occasionally exchanged rankings with Consistency, illustrating their comparable significance. Nevertheless, Currentness consistently remained within the top-level ranks (third or fourth), indicating its sustained importance across diverse contexts.
5) **Precision** (Rank 5, Global Weight: 13.5%) reflects the importance of capturing detailed, granular clinical data. Highly precise data enhances ML models' ability to discern subtle variations in patient conditions, thereby refining risk predictions and treatment guidelines. Remarkably, Precision exhibited rank stability throughout the majority of sensitive analysis scenarios, only slightly altering its fifth rank under all perturbation

conditions. Even when the Representational category, its parent category, was adjusted, Precision's relative ranking remained unaffected. This stability emphasizes consensus and directs efforts towards improving data precision, which is especially vital in healthcare applications requiring detailed physiological measurements and precise medication dosages.

Middle-Ranked Dimensions Discussion (6–10)

6) **Efficiency** (Rank 6, Global Weight: 3.8%) occupies the sixth position, embodying the promptness and computational simplicity associated with data retrieval and processing within ML systems. In critical healthcare environments such as intensive care units or emergency departments, access to dependable and swift data is of utmost importance. Although not among the top-ranked dimensions, the positioning of Efficiency signifies its contextual significance in systems where minimal latency and optimized performance are essential. Sensitivity analysis demonstrated that Efficiency was uniquely stable, maintaining its sixth rank across all perturbation scenarios within a $\pm 17.6\%$ category-level variation. This consistent placement indicates a strong consensus among experts regarding its vital role in maintaining operational agility. Such stability affirms the strategic importance of Efficiency for real-time health monitoring and early-warning systems, where the timeliness of data ingestion has a direct impact on patient outcomes.
7) **Recoverability** (Rank 7, Global Weight: 2.8%) denotes the capacity to restore data integrity following loss or corruption, thereby ensuring system resilience and continuity. Its prioritization highlights the necessity of establishing backup and recovery protocols, particularly within healthcare infrastructures susceptible to hardware failures, cybersecurity threats, or data management errors. The sensitivity analysis revealed moderate positional fluctuations, with Recoverability occasionally shifting between seventh and eighth ranks, depending on the emphasis placed on the Intrinsic or Contextual categories. This variation indicates a situational but consistent acknowledgment of Recoverability's significance in safeguarding the reliability of data pipelines, which is essential for the performance of ML systems.
8) **Traceability** (Rank 8, Global Weight: 2.7%) pertains to the capacity to trace data lineage from its origin through transformation to final application. Within healthcare ML systems, this dimension facilitates auditability, ensures compliance with legal mandates (e.g., HIPAA), and enhances the interpretability of predictions, particularly in explainable artificial intelligence (XAI) models. Sensitivity analyses indicated that Traceability occasionally ascended to the seventh position, especially under reductions in weight within the Contextual category. This adaptability signifies its growing significance in scenarios where algorithmic transparency is imperative. Such versatility corresponds with the increasing demands for ethical ML, thereby reinforcing Traceability's importance in substantiating clinical decisions and fostering stakeholder trust.
9) **Credibility** (Rank 9, Global Weight: 2.5%) measures the perceived reliability of data as assessed by domain experts and end users. In clinical settings, the acceptance of ML-generated recommendations relies not only on model performance but also on the trustworthiness of the underlying data. During sensitivity analysis, Credibility sometimes fell to tenth place, particularly when the weights of Intrinsic or Representational categories increased. These fluctuations indicate that, while valuable, Credibility is regarded as a supporting dimension rather than a primary driver of ML effectiveness. However, maintaining transparent documentation and validation processes is essential for building trust in AI-enabled healthcare delivery.
10) **Understandability** (Rank 10, Global Weight: 2.5%) pertains to the ease with which clinicians, data scientists, and IT practitioners can interpret data content. Its significance is paramount for interface design, model debugging, and patient communication, particularly within decision support tools. Sensitivity tests indicated that Understandability was the variable element among the dimensions. This variability reflects its dependence on the context of application, which is crucial in systems involving human-in-the-loop interactions, yet secondary in fully automated pipelines. Nevertheless, its role in interpretability is consistent with regulatory requirements for transparency and highlights its importance in the ethical implementation of ML in healthcare.

Lowest-Ranked Dimensions Discussion (11–15)

11) **Confidentiality** (Rank 11, Global Weight: 1.9%) involves protecting patient data from unauthorized access. While both ethically and legally essential, experts perceive this aspect as having a less direct impact on the operational performance of ML models. In healthcare ML systems, confidentiality is predominantly upheld through institutional data governance policies and encryption protocols that operate independently of model training pipelines. Sensitivity analysis revealed notable stability for Confidentiality, as there was no variation in its ranking across all perturbation scenarios. This observation suggests a consistent expert consensus regarding its supportive rather than central role in data-driven modelling. These findings highlight its strategic significance in system-level design, rather than in enhancing algorithmic accuracy.
12) **Compliance** (Rank 12, Global Weight: 1.7%) involves aligning data use with regulations. Its lower rank indicates that experts see it more as a regulatory requirement than a dimension that directly enhances

ML performance. While it doesn't affect prediction accuracy or learning efficiency, ensuring compliance is essential for deployment readiness and legal sustainability in healthcare ML.

13) **Availability** (Rank 13, Global Weight: 1.5%) reflects how easily and consistently data can be accessed. In healthcare ML, it influences the timeliness of model training and updates, particularly with large EHR or cloud datasets. Its consistent ranking suggests that, while fundamental, availability seldom affects core model performance metrics such as precision, recall, or F1 score.
14) **Accessibility** (Rank 14, Global Weight: 1.4%) measures how easily stakeholders, including clinicians, researchers, and administrators, can efficiently access and interpret data. Despite its critical organizational significance, its ranking reflects a limited influence on the technical aspects of model training and validation.
15) **Portability** (Rank 15, Global Weight: 1.0%), identified as the dimension with the lowest ranking, pertains to the transfer of datasets between various systems or platforms. In the domain of ML, particularly within federated or multi-institutional environments, it influences model retraining, data harmonization, and integration processes. Consistent evaluations indicate that it presently holds limited immediate priority, reflecting the prevailing expert consensus that, although significant for the exchange of global health data, its impact on the fundamental performance of ML within single-institution frameworks is minimal. While future advancements may enhance its significance, its current limited operational relevance substantiates its designated position.

These results highlight the varied effects of different data quality dimensions on the performance of ML-driven healthcare systems. The rankings and their underlying reasons reflect the complex relationship between data quality characteristics and the operational effectiveness of machine learning models. This detailed understanding of all fifteen dimensions supports focused resource allocation and guides future research aimed at integrating ethical compliance and infrastructure robustness into healthcare model development processes. Building upon these findings, this study introduces an innovative prioritization framework that uniquely integrates ISO/IEC 25012 with a restructured categorization designed to address the specific needs of ML applications. Whereas prior MCDM studies often depended on generalized or loosely adapted criteria in healthcare applications or on the separate selection of ML algorithms or other applications, our approach is specifically developed to encompass the multidimensional complexity of healthcare data quality through structured expert input, while explicitly considering the requirements of ML systems.

Furthermore, this study's sensitivity analysis deviates from conventional approaches that employ arbitrary incremental perturbations such as ±5%, ±10%, or ±15% to evaluate stability. Such uniform adjustments can introduce artificial bias and may not accurately reflect real-world uncertainty. Instead, the ±17.6% perturbation utilized in this investigation was empirically derived from the standard deviation of aggregated expert assessments at the category level. This approach, driven by variance, enhances realism and effectively captures the influence of uncertainty on prioritization outcomes, considering the hierarchical structure of the AHP, wherein variations propagate throughout the model and impact dimension-level rankings.

This integration of ISO-based reclassification, AHP prioritization, and statistically grounded sensitivity testing signifies a significant progression from previous endeavors. It provides a more resilient, context-aware, and scalable framework for aligning data quality strategies with the performance improvement needs of ML-driven healthcare systems.

VII. CONTRIBUTIONS AND FUTURE DIRECTIONS

This study contributes to the advancement of prioritizing essential data quality dimensions for effective ML-driven healthcare applications. By utilizing the AHP to evaluate and classify critical data quality dimensions, the research enhances our understanding of their direct impacts on the performance of ML models. This foundational work addresses significant gaps in the existing literature, primarily by providing a structured and validated methodology. In contrast to previous approaches that broadly address data quality without detailed methodological rigor or comprehensive dimension categorization, this study integrates expert-driven assessment and systematic validation through sensitivity analysis, offering clear theoretical and methodological contributions. The forthcoming subsections will elaborate on the theoretical, practical implications, practical implementation, limitations, and future directions of the findings.

A. THEORETICAL IMPLICATIONS

This research provides significant theoretical contributions by identifying and prioritizing the dimensions of data quality that are most critical for ML applications within healthcare systems. Utilizing AHP, this offers a quantitative evaluation of essential data quality dimensions, each assessed within a structured hierarchy explicitly developed for healthcare ML contexts.

A significant methodological contribution lies in the rigorous validation of prioritization outcomes through sensitivity analysis, employing the OAT perturbation method. By introducing controlled perturbations derived from expert judgment variability (±17.6%), the study confirms the stability and reliability of the prioritization results, thereby enhancing the credibility and theoretical soundness of the AHP methodology in healthcare ML contexts.

Ethical considerations are crucial in healthcare applications driven by ML, especially when these systems depend on data of varying quality. This study emphasizes the importance of prioritizing data quality dimensions to ensure fairness and transparency in clinical decision-making. For example, dimensions such as accuracy, completeness, and traceability help mitigate the risks of model bias, which can disproportionately impact underrepresented or vulnerable patient groups. By ensuring high data quality, we can reduce algorithmic discrimination and enhance the representativeness and integrity of the input data.

Moreover, the diversity of stakeholders was an intentional aspect of the expert-driven AHP process used in this study. The panel included specialists from various academic and healthcare institutions, which allowed for a comprehensive understanding of data quality priorities. However, future expansions of this work should involve a wider range of stakeholders, such as clinical practitioners, more IT practitioners, hospital administrators, and representatives from patient advocacy groups. This inclusivity would enhance the framework's applicability and ensure that ethical concerns are addressed across different healthcare settings. As ML systems advance to support essential functions like diagnosis, triage, and treatment recommendations, implementing ethical safeguards related to data quality becomes critical. This study lays the groundwork for incorporating fairness-aware data governance into ML workflows by pinpointing which dimensions of data quality most affect algorithmic reliability and trustworthiness.

Additionally, this research advances theoretical knowledge by identifying and addressing existing research gaps as highlighted by recent foundational studies. While prior research has often lacked structured prioritization frameworks, the structured approach of this study directly fills this methodological void.

B. PRACTICAL IMPLICATIONS

This study offers valuable insights for enhancing ML outcomes within real-world healthcare environments by identifying the most critical dimensions of data quality. Completeness is recognized as the highest priority, emphasizing the necessity for comprehensive and continuous data capture to ensure that ML models are developed using a representative and unbiased patient population. Accuracy and consistency are also accorded with significant importance, indicating the need for reliable, error-free, and standardized data across various sources to support credible model outputs. These findings highlight the significance of data governance practices, encompassing validation pipelines, standardization protocols, and audit trails to preserve high-quality datasets.

In practical terms, healthcare institutions engaged in the development of ML tools should adequately invest in integrated data infrastructures that facilitate comprehensive, accurate, and harmonized data collection across various departments and systems. For instance, the implementation of real-time data validation within ICU monitoring systems, or the automated flagging of inconsistent entries in EHRs, can substantially enhance ML readiness.

Although currentness has emerged as significant, its comparatively lower weight implies that timeliness must be judiciously balanced with fundamental quality metrics. This indicates that, although real-time updates hold considerable value, they should not jeopardize the accuracy or completeness of the data. Consequently, healthcare analytics teams ought to concentrate on ensuring the freshness of data in high-impact use cases, such as early-warning systems, while simultaneously upholding the overall integrity of the data.

In contrast to prior studies that have approached data quality in abstract terms or within broader healthcare contexts, this research explicitly concentrates on ML-driven systems. This study presents a practical and ranked roadmap designed to inform data management decisions. By delivering a validated prioritization via the AHP and further substantiating its robustness through sensitivity analysis, it establishes a replicable and empirically grounded framework that can be adapted by hospitals, health technology companies, and policymakers who aim to implement effective ML-driven interventions.

C. PRACTICAL IMPLEMENTATION OF THE FRAMEWORK IN ML-DRIVEN HEALTHCARE

The research outlines five fundamental data quality dimensions: Completeness, Accuracy, Consistency, Currentness, and Precision, which can serve as direct guides in developing and deploying ML-driven healthcare systems.

- Project Planning: Employ these top-ranked dimensions to define data quality requirements at the outset, such as establishing minimum thresholds for completeness and implementing accuracy validation procedures, thereby mitigating the risk of suboptimal model performance. Additionally, identify detailed data issues associated with each respective dimension.
- Data Curation: Direct data engineers in the processes of cleaning, transformation, and integration workflows by emphasizing curation rules pertinent to the prioritized dimensions, while considering the operational constraints of each (e.g., harmonizing units to ensure consistency, verifying timestamps to maintain currentness).
- Quality Assurance Pipelines: Establish continuous quality checks during both model development and deployment stages, utilizing dashboards and alert systems to monitor parameters such as data completeness and accuracy.

Focusing on the most critical dimensions, ML healthcare teams can effectively mitigate data-related risks, thereby enhancing the reliability, safety, and overall trustworthiness of deployed models.

D. LIMITATIONS AND FUTURE WORK

Despite the strength of its findings, this study acknowledges various limitations and outlines directions for future research.

First, the application of the AHP presupposes independence among dimensions, a condition that may not entirely represent the intricate interactions inherent in real-world data

TABLE 11. Questionnaire used for AHP-based prioritization of data quality dimensions.

Q	Category	Scale																	Category
1	Intrinsic	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Contextual
2		9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Representational
3		9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Accessibility
4	Contextual	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Representational
5		9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Accessibility
6	Representational	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Accessibility
Q	**Intrinsic**	**Scale**																	**Dimension**
1	Accuracy	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Credibility
2		9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Traceability
3	Credibility	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Traceability
Q	**Contextual**	**Scale**																	**Dimension**
1	Completeness	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Efficiency
2		9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Currentness
3	Efficiency	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Currentness
Q	**Representational**	**Scale**																	**Dimension**
1	Understandability	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Consistency
2		9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Precision
3		9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Recoverability
4	Consistency	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Precision
5		9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Recoverability
6	Precision	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Recoverability
Q	**Accessibility**	**Scale**																	**Dimension**
1	Availability	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Confidentiality
2		9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Accessibility
3		9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Portability
4		9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Compliance
5	Confidentiality	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Accessibility
6		9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Portability
7		9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Compliance
8	Accessibility	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Portability
9		9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Compliance
10	Portability	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Compliance

ecosystems. For example, accuracy and consistency may mutually reinforce one another; neglecting such relationships could result in oversimplified conclusions. Future research endeavors should investigate advanced MCDM techniques, such as the Analytic Network Process (ANP) or fuzzy cognitive mapping approaches, which can provide a more nuanced understanding of how dimensions interact and influence the performance of ML models.

Second, while the study adopts ISO/IEC 25012 as a guiding standard, it recognizes that this standard was not originally designed for ML-centric healthcare data, which often involves heterogeneous formats, unstructured data, and streaming inputs. To address this, future frameworks should consider domain-specific expansions of existing standards or develop bespoke data quality taxonomies for ML use cases.

Third, a significant contribution of this research is its establishment of a foundation for the empirical validation of the proposed framework by identifying the most critical dimensions. While the current study relies on expert-driven AHP analysis and is supported by sensitivity testing, it has not yet been empirically validated using actual healthcare datasets. Future research should apply the prioritization framework to real clinical data repositories, such as ICU records or diagnostic datasets, to evaluate how improvements in the top-ranked dimensions (e.g., Completeness, Accuracy, Consistency) impact ML performance metrics like AUC, sensitivity, and specificity. Moreover, simulation studies could investigate how enhancements in lower-ranked dimensions affect model interpretability and reliability in high-stakes healthcare settings. By combining theoretical prioritization with empirical findings, this approach will enhance the framework's generalizability and real-world applicability.

Another significant limitation pertains to the dependence on expert evaluations, which may introduce subjective bias resulting from the composition of the expert panel, their professional backgrounds, or domain-specific perspectives. Such biases could impact pairwise comparison judgments and, consequently, the prioritization outcomes. While the panel in this study was deliberately diverse in expertise and affiliation, future research could further alleviate this risk by expanding the number of participants, incorporating stakeholders from diverse geographic regions, and employing iterative consensus-building methods such as the Delphi technique. Furthermore, hybrid approaches that combine expert judgments with empirical weighting based on real-world healthcare data could improve objectivity and reproducibility.

Finally, although expert perspectives were included, the sample could benefit from greater diversity to include clinicians, data engineers, and patient advocates. This change would strengthen the relevance of each dimension within real clinical workflows and ensure that prioritization considers ethical and usability factors, such as fairness and interpretability. Future research should involve this prioritization framework across different healthcare sectors, including critical systems, oncology, cardiology, emergency care, and others, to test how well the findings generalize and to adapt the framework to specific context needs. Further, longitudinal studies could examine how changes in certain data quality metrics over time influence ML outcomes, helping to create feedback loops for continuous quality improvement.

VIII. CONCLUSION

This study utilized the AHP to evaluate and prioritize the dimensions of data quality that are crucial for the performance of ML systems in healthcare. The importance of dimensions such as completeness, consistency, currentness, accuracy, and precision is vital for ensuring the effective performance of ML models. This study presents a practical framework aimed at improving data quality, which ultimately enhances the reliability and predictive capabilities of ML applications in healthcare. Additionally, the integration of sensitivity analysis has confirmed the robustness of these prioritized dimensions, increasing confidence in their stability under various assumptions.

This research makes significant theoretical and practical contributions by providing a structured and expert-driven prioritization of data quality dimensions specifically designed for ML in healthcare. This area has previously lacked empirical frameworks. It also addresses the demand for practical relevance by offering clear, actionable insights that healthcare providers and data engineers can implement to enhance the effectiveness of their systems. Further, the research considers ethical issues and the diversity of stakeholders, emphasizing the necessity for fair, transparent, and inclusive data quality strategies.

Future research should further develop this work through empirical validation using real-world datasets, refining the framework to adapt to changing data standards, and exploring the interdependencies among dimensions to enhance its practical application.

APPENDIX

See Table 11.

ACKNOWLEDGMENT

The authors wish to convey their sincere appreciation to Universiti Putra Malaysia for the institutional support provided throughout this research process.

Furthermore, they extend their profound gratitude to the domain experts and practitioners who participated in the pairwise comparison procedure. Their valuable insights and contributions were vital in the development of the prioritization framework.

REFERENCES

[1] R. S. Geiger, K. Yu, Y. Yang, M. Dai, J. Qiu, R. Tang, and J. Huang, "Garbage in, garbage out? Do machine learning application papers in social computing report where human-labeled training data comes from?" in *Proc. Conf. Fairness, Accountability, Transparency*, 2020, pp. 325–336.

[2] M. A. Gianfrancesco and N. D. Goldstein, "A narrative review on the validity of electronic health record-based research in epidemiology," *BMC Med. Res. Methodol.*, vol. 21, no. 1, p. 234, Dec. 2021.

[3] T. Sarwar, S. Seifollahi, J. Chan, X. Zhang, V. Aksakalli, I. Hudson, K. Verspoor, and L. Cavedon, "The secondary use of electronic health records for data mining: Data characteristics and challenges," *ACM Comput. Surv.*, vol. 55, no. 2, pp. 1–40, Feb. 2023.

[4] A. E. Lewis, N. Weiskopf, Z. B. Abrams, R. Foraker, A. M. Lai, P. R. O. Payne, and A. Gupta, "Electronic health record data quality assessment and tools: A systematic review," *J. Amer. Med. Inform. Assoc.*, vol. 30, no. 10, pp. 1730–1740, Sep. 2023.

[5] R. Bey, A. Cohen, V. Trebossen, B. Dura, P.-A. Geoffroy, C. Jean, B. Landman, T. Petit-Jean, G. Chatellier, K. Sallah, X. Tannier, A. Bourmaud, and R. Delorme, "Natural language processing of multi-hospital electronic health records for public health surveillance of suicidality," *npj Mental Health Res.*, vol. 3, no. 1, p. 6, Feb. 2024.

[6] S. M. Meystre, C. Lovis, T. Bürkle, G. Tognola, A. Budrionis, and C. U. Lehmann, "Clinical data reuse or secondary use: Current status and potential future progress," *Yearbook Med. Informat.*, vol. 26, no. 1, pp. 38–52, 2017.

[7] T. Pham, T. Tran, D. Phung, and S. Venkatesh, "DeepCare: A deep dynamic memory model for predictive medicine," in *Proc. 20th Pacific-Asia Conf. Adv. Knowl. Discovery Data Mining*, Auckland, New Zealand. Cham, Switzerland: Springer, Apr. 2016, pp. 30–41.

[8] B. Smith, S. Van Steelandt, and A. Khojandi, "Evaluating the impact of health care data completeness for deep generative models," *Methods Inf. Med.*, vol. 62, no. 2, pp. 031–039, May 2023.

[9] S. Moazemi, S. Vahdati, J. Li, S. Kalkhoff, L. J. V. Castano, B. Dewitz, R. Bibo, P. Sabouniaghdam, M. S. Tootooni, R. A. Bundschuh, A. Lichtenberg, H. Aubin, and F. Schmid, "Artificial intelligence for clinical decision support for monitoring patients in cardiovascular ICUs: A systematic review," *Frontiers Med.*, vol. 10, Mar. 2023, Art. no. 1109411.

[10] A. C. Liu, K. Patel, R. D. Vunikili, K. W. Johnson, F. Abdu, S. K. Belman, B. S. Glicksberg, P. Tandale, R. Fontanez, O. K. Mathew, A. Kasarskis, P. Mukherjee, L. Subramanian, J. T. Dudley, and K. Shameer, "Sepsis in the era of data-driven medicine: Personalizing risks, diagnoses, treatments and prognoses," *Briefings Bioinf.*, vol. 21, no. 4, pp. 1182–1195, Jul. 2020.

[11] J. Venugopalan, N. Chanani, K. Maher, and M. D. Wang, "Novel data imputation for multiple types of missing data in intensive care units," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 3, pp. 1243–1250, May 2019.

[12] Y. Gong, G. Liu, Y. Xue, R. Li, and L. Meng, "A survey on dataset quality in machine learning," *Inf. Softw. Technol.*, vol. 162, Oct. 2023, Art. no. 107268.

[13] R. Syed, R. Eden, T. Makasi, I. Chukwudi, A. Mamudu, M. Kamalpour, D. K. Geeganage, S. Sadeghianasl, S. J. J. Leemans, K. Goel, R. Andrews, M. T. Wynn, A. T. Hofstede, and T. Myers, "Digital health data quality issues: Systematic review," *J. Med. Internet Res.*, vol. 25, Mar. 2023, Art. no. e42615.

[14] C. Liu, A. Talaei-Khoei, V. C. Storey, and G. Peng, "A review of the state of the art of data quality in healthcare," *J. Global Inf. Manage.*, vol. 31, no. 1, pp. 1–18, Jan. 2023.

[15] M. G. Kahn, T. J. Callahan, J. Barnard, A. Bauck, J. S. Brown, B. N. Davidson, H. Estiri, C. Goerg, E. Holve, S. Johnson, S. Liaw, M. Hamilton-Lopez, D. Meeker, T. C. Ong, P. Ryan, N. Shang, N. G. Weiskopf, C. Weng, M. Zozus, and L. M. Schilling, "A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data," *Egems*, vol. 4, no. 1, p. 18, 2016.

[16] S. T. Liaw, A. Rahimi, P. Ray, J. Taggart, S. Dennis, S. de Lusignan, B. Jalaludin, A. E. T. Yeo, and A. Talaei-Khoei, "Towards an ontology for data quality in integrated chronic disease management: A realist review of the literature," *Int. J. Med. Informat.*, vol. 82, no. 1, pp. 10–24, Jan. 2013.

[17] R. Y. Wang and D. M. Strong, "Beyond accuracy: What data quality means to data consumers," *J. Manage. Inf. Syst.*, vol. 12, no. 4, pp. 5–33, Mar. 1996.

[18] N. G. Weiskopf and C. Weng, "Methods and dimensions of electronic health record data quality assessment: Enabling reuse for clinical research," *J. Amer. Med. Inform. Assoc.*, vol. 20, no. 1, pp. 144–151, Jan. 2013.

[19] *Software Engineering-Software Product Quality Requirements and Evaluation (SQuaRE)—Data Quality Model*, Standard IEC 25012, International Organization for Standardization, 2000.

[20] M. Priestley, F. O'donnell, and E. Simperl, "A survey of data quality requirements that matter in ML development pipelines," *J. Data Inf. Qual.*, vol. 15, no. 2, pp. 1–39, Jun. 2023.

[21] J. Więckowski and W. Sałabun, "Comparative sensitivity analysis in composite material selection: Evaluating OAT and COMSAM methods in multi-criteria decision-making," *Spectr. Mech. Eng. Oper. Res.*, vol. 2, no. 1, pp. 1–12, Jan. 2025, doi: 10.31181/smeor21202524.

[22] L. L. Pipino, Y. W. Lee, and R. Y. Wang, "Data quality assessment," *Commun. ACM*, vol. 45, no. 4, pp. 211–218, 2002.

[23] C. Batini, C. Cappiello, C. Francalanci, and A. Maurino, "Methodologies for data quality assessment and improvement," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 1–52, Jul. 2009.

[24] W. Elouataoui, "AI-driven frameworks for enhancing data quality in big data ecosystems: Error_detection, correction, and metadata integration," 2024, *arXiv:2405.03870*.

[25] A. Aldoseri, K. N. Al-Khalifa, and A. M. Hamouda, "Re-thinking data strategy and integration for artificial intelligence: Concepts, opportunities, and challenges," *Appl. Sci.*, vol. 13, no. 12, p. 7082, Jun. 2023.

[26] Y. Jaradat, M. Masoud, A. Manasrah, M. Alia, and I. Jannoud, "Review of data imputation techniques in time series data: Comparative analysis," *The Eurasia Proceedings of Science Technology Engineering and Mathematics*, vol. 27, pp. 122–129, Jun. 2024. Accessed: Jun. 7, 2025. [Online]. Available: http://www.epstem.net/tr/pub/issue/86233/1518433

[27] K. Dzobo, S. Adotey, N. E. Thomford, and W. Dzobo, "Integrating artificial and human intelligence: A partnership for responsible innovation in biomedical engineering and medicine," *OMICS, J. Integrative Biol.*, vol. 24, no. 5, pp. 247–263, May 2020.

[28] H. Habehh and S. Gohel, "Machine learning in healthcare," *Curr Genomics*, vol. 22, no. 4, pp. 291–300, 2021.

[29] A. Al-Hgaish, W. ALzyadat, M. Al-Fayoumi, A. Alhroob, and A. Thunibat, "Preserve quality medical drug data toward meaningful data lake by cluster," *Int. J. Recent Technol. Eng.*, vol. 8, no. 3, pp. 270–277, Sep. 2019.

[30] B. A. M. Al-Hgaish, M. Al-Fayoumi, and D. Al-Zyadat. (2019). *An Approach Preserve Quality Medical Drug Data (Semi-Structure) Toward Meaningful Data Lake by Cluster*. Accessed: Jun. 7, 2017. [Online]. Available: https://iu.edu.jo/files/thesis/2019/68187.pdf

[31] W. Alzyadat, M. Muhairat, A. Alhroob, and T. Rawashdeh, "A recruitment big data approach to interplay of the target drugs," *Int. J. Adv. Soft Comput. Appl.*, vol. 14, no. 1, pp. 02–13, Apr. 2022.

[32] B. Saini, S. Srivastava, and A. Bajpai, "Deep CNN model for nanotoxicity classification using microscopic images," *Int. J. Adv. Soft Comput. Appl.*, vol. 12, no. 2, p. 22, 2020.

[33] Y. Akhtar, M. Ali, F. A. Almaliki, and R. A. Almarzouki, "A novel IoT-based approach using fractional fuzzy Hamacher aggregation operators application in revolutionizing healthcare selection," *Sci. Rep.*, vol. 15, no. 1, p. 2712, Jan. 2025, doi: 10.1038/s41598-024-83805-6.

[34] O. Refai Momani, M. Belal Al-Zoubi, and M. Al Tawil, "Multi-class lung disease classification using convolutional neural networks," *Int. J. Adv. Soft Comput. Appl.*, vol. 17, no. 1, pp. 192–216, Mar. 2025, doi: 10.15849/ijasca.250330.11.

[35] A. Al-Yousef, M. Al-Shannag, and S. M. Alkhushayni, "The impact of using BI-RADS with voting classifier fusion for early detection of breast cancer," *Int. J. Adv. Soft Comput. Appl.*, vol. 17, no. 1, pp. 133–146, Mar. 2025, doi: 10.15849/ijasca.250330.08.

[36] M. Javaid, A. Haleem, R. Pratap Singh, R. Suman, and S. Rab, "Significance of machine learning in healthcare: Features, pillars and applications," *Int. J. Intell. Netw.*, vol. 3, pp. 58–73, Jan. 2022.

[37] F. Magrabi, E. Ammenwerth, J. B. McNair, N. F. D. Keizer, H. Hyppönen, P. Nykänen, M. Rigby, P. Scott, T. Vehko, Z. S.-Y. Wong, and A. Georgiou, "Artificial intelligence in clinical decision support: Challenges for evaluating AI and practical implications," *Yearb Med. Inf.*, vol. 28, no. 1, pp. 128–134, 2019.

[38] S. M. D. A. C. Jayatilake and G. U. Ganegoda, "Involvement of machine learning tools in healthcare decision making," *J. Healthcare Eng.*, vol. 2021, pp. 1–20, Jan. 2021.

[39] M. Siddiq, "Use of Machine Learning to predict patient developing a disease or condition for early diagnose," *Int. J. Multidisciplinary Sci. Arts*, vol. 1, no. 1, pp. 13–23, 2022.

[40] W. Li, F. Jia, and Q. Hu, "Automatic segmentation of liver tumor in CT images with deep convolutional neural networks," *J. Comput. Commun.*, vol. 3, no. 11, pp. 146–151, 2015.

[41] R. Rayan, "Artificial intelligence perspective on healthcare," in *Proc. ICEAT*, 2020, pp. 1–13.

[42] I. Jannoud, M. Z. Masoud, Y. Jaradat, A. Manaserah, and D. Zaidan, "A multi-layered hybrid machine learning algorithm (MLHA) for type II diabetes classification," *Proc. Comput. Sci.*, vol. 237, pp. 445–452, Jan. 2024.

[43] A. R. O'Brien, L. O. W. Wilson, G. Burgio, and D. C. Bauer, "Unlocking HDR-mediated nucleotide editing by identifying high-efficiency target sites using machine learning," *Sci. Rep.*, vol. 9, no. 1, p. 2788, Feb. 2019.

[44] X. Pan, J. Zuallaert, X. Wang, H.-B. Shen, E. P. Campos, D. O. Marushchak, and W. De Neve, "ToxDL: Deep learning using primary structure and domain embeddings for assessing protein toxicity," *Bioinformatics*, vol. 36, no. 21, pp. 5159–5168, Jan. 2021.

[45] J. Bajwa, U. Munir, A. Nori, and B. Williams, "Artificial intelligence in healthcare: Transforming the practice of medicine," *Future Healthcare J.*, vol. 8, no. 2, pp. e188–e194, Jul. 2021.

[46] C. L. Andaur Navarro, J. A. A. Damen, T. Takada, S. W. J. Nijman, P. Dhiman, J. Ma, G. S. Collins, R. Bajpai, R. D. Riley, K. G. M. Moons, and L. Hooft, "Risk of bias in studies on prediction models developed using supervised machine learning techniques: Systematic review," *BMJ*, vol. 375, p. 2281, Oct. 2021.

[47] M. Gianfrancesco, S. Tamang, J. Yazdany, and G. Schmajuk, "Potential biases in machine learning algorithms using electronic health record data," *JAMA Internal Med.*, vol. 178, no. 11, p. 1544, 2018.

[48] D. S. Char, N. H. Shah, and D. Magnus, "Implementing machine learning in health care—Addressing ethical challenges," *New England J. Med.*, vol. 378, no. 11, pp. 981–983, Mar. 2018.

[49] R. Challen, J. C. Denny, M. Pitt, L. Gompels, T. L. Edwards, and K. Tsaneva-Atanasova, "Artificial intelligence, bias and clinical safety," *BMJ Quality Safty*, vol. 28, no. 3, pp. 231–237, 2019.

[50] A. Al-Hgaish, R. Atan, Y. Razali, and M. H. Osman, "Enhancing performance of machine learning models in healthcare: An analytical framework for assessing and improving data quality," in *Proc. Int. Conf. Adv. Inf. Commun. Technol. Comput.*, 2024, pp. 137–153.

[51] A. Al-Hgaish, R. Atan, R. Yaakob, and M. H. Osman, "An analytical framework for assessing and improving data quality?" in *Advances in Information Communication Technology and Computing*. Singapore: Springer, 2025. [Online]. Available: https://books.google.com/books?hl=ar&lr=&id=JJssEQAAQBAJ&oi=fnd&pg=PA137&dq=An+Analytical+Framework+for+Assessing+and+Improving+Data+Quality&ots=fWMDy8aV2k&sig=w8hwOax8VCV56LqaMowwzteYWWg

[52] S. Ajami and T. BagheriTadi, "Barriers for adopting electronic health records (EHRs) by physicians," *Acta Inf. Medica*, vol. 21, no. 2, p. 129, 2013.

[53] N. Menachemi, "Benefits and drawbacks of electronic health record systems," *Risk Manage. Healthcare Policy*, vol. 4, pp. 47–55, May 2011.

[54] G. Hripcsak and D. J. Albers, "Next-generation phenotyping of electronic health records," *J. Amer. Med. Inform. Assoc.*, vol. 20, no. 1, pp. 117–121, Jan. 2013.

[55] P. C. Austin, I. R. White, D. S. Lee, and S. van Buuren, "Missing data in clinical research: A tutorial on multiple imputation," *Can. J. Cardiol.*, vol. 37, no. 9, pp. 1322–1331, Sep. 2021.

[56] I. A. Gheyas and L. S. Smith, "A neural network-based framework for the reconstruction of incomplete data sets," *Neurocomputing*, vol. 73, nos. 16–18, pp. 3039–3065, Oct. 2010.

[57] J. Yang, R. El-Bouri, O. O'Donoghue, A. S. Lachapelle, A. A. S. Soltan, D. W. Eyre, L. Lu, and D. A. Clifton, "Deep reinforcement learning for multi-class imbalanced training: Applications in healthcare," *Mach. Learn.*, vol. 113, no. 5, pp. 2655–2674, May 2024.

[58] E. Šabić, D. Keeley, B. Henderson, and S. Nannemann, "Healthcare and anomaly detection: Using machine learning to predict anomalies in heart rate data," *AI Soc.*, vol. 36, no. 1, pp. 149–158, Mar. 2021.

[59] C. Peng and P. Goswami, "Meaningful integration of data from heterogeneous health services and home environment based on ontology," *Sensors*, vol. 19, no. 8, p. 1747, Apr. 2019.

[60] *Improving Data Quality: A Guide for Developing Countries. 2003*, World Health Organization, Geneva, Switzerland, 2019.

[61] K. Lucyk, K. Tang, and H. Quan, "Barriers to data quality resulting from the process of coding health information to administrative data: A qualitative study," *BMC Health Services Res.*, vol. 17, no. 1, pp. 1–10, Dec. 2017.

[62] W. V. de Souza, "Notificação compulsória de doenças e agravos no brasil: Um breve histórico sobre a criação do sistema de informação de agravos de notificação—SINAN," in *A Experiência Brasileira em Sistemas de Informação em Saúde*. Brasília, Brazil: Editora MS (Ministério da Saúde), vol. 39, 2009.

[63] E. Ndabarora, J. A. Chipps, and L. Uys, "Systematic review of health data quality management and best practices at community and district levels in LMIC," *Inf. Develop.*, vol. 30, no. 2, pp. 103–120, May 2014.

[64] N. Puttkammer, J. G. Baseman, E. B. Devine, J. S. Valles, N. Hyppolite, F. Garilus, J. G. Honoré, A. I. Matheson, S. Zeliadt, K. Yuhas, K. Sherr, J. R. Cadet, G. Zamor, E. Pierre, and S. Barnhart, "An assessment of data quality in a multi-site electronic medical record system in Haiti," *Int. J. Med. Informat.*, vol. 86, pp. 104–116, Feb. 2016.

[65] Z. Wang, M. Penning, and M. Zozus, "Analysis of anesthesia screens for rule-based data quality assessment opportunities," *Stud. Health Technol. Inform.*, vol. 257, p. 473, Jan. 2019.

[66] M. V. Lucadou, T. Ganslandt, H. Prokosch, and D. Toddenroth, "Feasibility analysis of conducting observational studies with the electronic health record," *BMC Med. Inform. Decis. Making*, vol. 19, no. 1, pp. 1–14, 2019.

[67] M. Isgut, L. Gloster, K. Choi, J. Venugopalan, and M. D. Wang, "Systematic review of advanced AI methods for improving healthcare data quality in post COVID-19 era," *IEEE Rev. Biomed. Eng.*, vol. 16, pp. 53–69, 2023.

[68] A. Vrudhula, A. C. Kwan, D. Ouyang, and S. Cheng, "Machine learning and bias in medical imaging: Opportunities and challenges," *Circulat., Cardiovascular Imag.*, vol. 17, no. 2, Feb. 2024, Art. no. 015495.

[69] T. L. Saaty, "The analytic hierarchy process (AHP)," *J. Oper. Res. Soc.*, vol. 41, no. 11, pp. 1073–1076, 1980.

[70] A. Mumtaz, "Prioritizing and overcoming barriers to e-health use among elderly people: Implementation of the analytical hierarchical process (AHP)," *J. Healthcare Eng.*, vol. 2022, pp. 1–11, Jan. 2022.

[71] D. H. D. Citrawati, M. N. Widyawati, and S. Suryono, "The role of the analytic hierarchy process (AHP) algorithm in health care services," *J. Phys., Conf. Ser.*, vol. 1524, no. 1, Apr. 2020, Art. no. 012116.

[72] M. J. Liberatore and R. L. Nydick, "The analytic hierarchy process in medical and health care decision making: A literature review," *Eur. J. Oper. Res.*, vol. 189, no. 1, pp. 194–207, Aug. 2008.

[73] S. Khanmohammadi and M. Rezaeiahari, "AHP based classification algorithm selection for clinical decision support system development," *Proc. Comput. Sci.*, vol. 36, pp. 328–334, Jan. 2014.

[74] B. R. D. Oliveira, C. C. E. D. Abreu, M. A. Q. Duarte, and J. V. Filho, "Geometrical features for premature ventricular contraction recognition with analytic hierarchy process based machine learning algorithms selection," *Comput. Methods Programs Biomed.*, vol. 169, pp. 59–69, Feb. 2019.

[75] S. Kubler, J. Robert, S. Neumaier, J. Umbrich, and Y. Le Traon, "Comparison of metadata quality in open data portals using the analytic hierarchy process," *Government Inf. Quart.*, vol. 35, no. 1, pp. 13–29, Jan. 2018.

[76] D. Khosroanjom, M. Ahmadzade, A. Niknafs, and R. K. Mavi, "Using fuzzy AHP for evaluating the dimensions of data quality," *Int. J. Bus. Inf. Syst.*, vol. 8, no. 3, pp. 269–285, 2011.

[77] M. I. Merhi, "Evaluating the critical success factors of data intelligence implementation in the public sector using analytical hierarchy process," *Technol. Forecasting Social Change*, vol. 173, Dec. 2021, Art. no. 121180.

[78] A. M. Talib, A. M. Al-Hgaish, R. B. Atan, A. Alshammari, F. O. Alomary, R. Yaakob, A. Alsahli, and M. H. Osman, "Evaluating critical success factors in AI-driven drug discovery using AHP: A strategic framework for optimization," *IEEE Access*, vol. 13, pp. 42045–42063, 2025, doi: 10.1109/ACCESS.2025.3546925.

[79] T. Liu, L. Cao, H. R. Wang, Y. J. Ma, X. Y. Lu, P. J. Li, and H. B. Wang, "Development and application of a WebGIS-based prediction system for multi-criteria decision analysis of porcine pasteurellosis," *Sci. Rep.*, vol. 14, no. 1, p. 21082, Sep. 2024.

[80] F. Weng, J. Tian, X. Fang, S. Cheng, Z. Pan, and Y. Liu, "Research on establishment decision of medical equipment measurement standard based on GDM-AHP," *Sci. Rep.*, vol. 15, no. 1, p. 9309, Mar. 2025, doi: 10.1038/S41598-025-94546-5.

[81] Y. Chen, J. Yu, and S. Khan, "Spatial sensitivity analysis of multi-criteria weights in GIS-based land suitability evaluation," *Environ. Model. Softw.*, vol. 25, no. 12, pp. 1582–1591, Dec. 2010, doi: 10.1016/j.envsoft.2010.06.001.

[82] T. Alballa, M. Rahim, N. M. Aloraini, and H. A. E.-W. Khalifa, "An extension of the CODAS method using p,q-quasirung orthopair fuzzy information: Application in location selection for retail store," *Int. J. Fuzzy Syst.*, vol. 27, pp. 1–17, Dec. 2024, doi: 10.1007/S40815-024-01870-5.

[83] M. Rahim, Y. Akhtar, M.-S. Yang, H. E. M. Ali, and A. A. Elhag, "Improved COPRAS method with unknown weights under p, q-quasirung orthopair fuzzy environment: Application to green supplier selection," *IEEE Access*, vol. 12, pp. 69783–69795, 2024, doi: 10.1109/ACCESS.2024.3400016.

[84] E. Adar, E. K. Delice, and T. Adar, "Prioritizing of industrial wastewater management processes using an integrated AHP–CoCoSo model: Comparative and sensitivity analyses," *Int. J. Environ. Sci. Technol.*, vol. 19, no. 6, pp. 4691–4712, Jun. 2022.

[85] J. Więckowski and W. Sałabun, "A new sensitivity analysis method for decision-making with multiple parameters modification," *Inf. Sci.*, vol. 678, Sep. 2024, Art. no. 120902.

[86] T. M. Mekonnen, A. B. Mitiku, and A. T. Woldemichael, "Flood hazard zoning of upper awash river basin, ethiopia, using the analytical hierarchy process (AHP) as compared to sensitivity analysis," *Sci. World J.*, vol. 2023, pp. 1–15, Apr. 2023.

[87] T. L. Saaty, "What is the analytic hierarchy process?" in *Mathematical Models for Decision Support*. Cham, Switzerland: Springer, 1988, pp. 109–121.

[88] A. Dias and P. G. Ioannou, "Company and project evaluation model for privately promoted infrastructure projects," *J. Construct. Eng. Manage.*, vol. 122, no. 1, pp. 71–82, Mar. 1996.

[89] J. M. Duke and R. Aull-Hyde, "Identifying public preferences for land preservation using the analytic hierarchy process," *Ecological Econ.*, vol. 42, nos. 1–2, pp. 131–145, Aug. 2002.

[90] R. K. Shrestha, J. R. R. Alavalapati, and R. S. Kalmbacher, "Exploring the potential for silvopasture adoption in south-central florida: An application of SWOT–AHP method," *Agricult. Syst.*, vol. 81, no. 3, pp. 185–199, Sep. 2004.

[91] E. Forman and K. Peniwati, "Aggregating individual judgments and priorities with the analytic hierarchy process," *Eur. J. Oper. Res.*, vol. 108, no. 1, pp. 165–169, Jul. 1998.

[92] Y. Chen, J. Yu, and S. Khan, "The spatial framework for weight sensitivity analysis in AHP-based multi-criteria decision making," *Environ. Model. Softw.*, vol. 48, pp. 129–140, Oct. 2013.

[93] T. L. Saaty and L. G. Vargas, "Sensitivity analysis in the analytic hierarchy process," in *Decision Making With the Analytic Network Process: Economic, Political, Social and Technological Applications With Benefits, Opportunities, Costs and Risks*. Oxford, U.K.: Elsevier, 2013, pp. 345–360.

[94] P. Banerjee, M. K. Ghose, and R. Pradhan, "AHP-based spatial air quality impact assessment model of vehicular traffic change due to highway broadening in Sikkim Himalaya," *Ann. GIS*, vol. 24, no. 4, pp. 287–302, Oct. 2018.

AREEN METIB AL-HGAISH (Member, IEEE) received the B.Sc. degree in software engineering, the B.Sc. degree in communications and computer engineering from Al-Zaytoonah University of Jordan, and the M.Sc. degree in software engineering.

She is currently pursuing the Ph.D. degree in software engineering with the University of Putra Malaysia (UPM). Her research interests include data quality management, machine learning optimization, and the development of analytical frameworks to enhance the performance of artificial intelligence systems. She particularly focused on improving healthcare applications by ensuring the integrity of medical data and exploring innovative AI-driven solutions. She has published research on data quality, machine learning, and AI applications in healthcare and has actively participated in academic collaborations and conferences. Her contributions reflect a dedication to addressing critical challenges in data-driven technologies and advancing the intersection of data quality and artificial intelligence.

RODZIAH BINTI ATAN received the Ph.D. degree in software engineering from the University of Putra Malaysia, in 2005.

She is currently with the Faculty of Computer Science and Information Technology, University of Putra Malaysia. She is also an academician responsible for teaching and supervising undergraduate and postgraduate students. Successfully, nine M.Sc. and 11 Ph.D. students graduated as primary supervisors. She specializes in software engineering, process modeling, and cloud computing services. She collaborates with industry and other institutions of higher learning locally and internationally. She has given multiple keynote speeches at international conferences and colloquiums, been appointed as a panel assessor, subject matter expert, local and international journals, and conference reviewer, and was appointed to a position as the Head of the Halal Policy and Management, for seven years, from 2016 to 2023.

RAZALI YAAKOB received the bachelor's and M.Sc. degrees in computer science from the Universiti Putra Malaysia (UPM), in 1996 and 1999, respectively, and the Ph.D. degree from the University of Nottingham, U.K., in 2008. He is currently an Associate Professor and a Lecturer with the Faculty of Computer Science and Information Technology, UPM.

He is the Dean of the Faculty. His research interests include artificial neural networks, pattern recognition, and evolutionary gaming computation. He is a member of the Faculty of Intelligent Computing Groups.

MOHD HAFEEZ OSMAN (Member, IEEE) received the B.Sc. degree in computer science and the M.Sc. degree in software engineering from the Universiti Teknologi Malaysia, Johor Bahru, Malaysia, in 2001 and 2002, respectively, and the Ph.D. degree in computer software engineering from Leiden University, Leiden, The Netherlands, in 2015.

His research interests include software engineering and data mining.