Expert Systems With Applications 246 (2024) 123066

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Evaluation of trustworthy artificial intelligent healthcare applications using multi-criteria decision-making approach

M.A. Alsalem [a], A.H. Alamoodi [i,j], O.S. Albahri [b,e], A.S. Albahri [d,e], Luis Martínez [f], R. Yera [f,*], Ali M. Duhaim [g], Iman Mohamad Sharaf [h]

[a] Department of Management Information Systems, College of Administration and Economics, University of Mosul, Mosul, Iraq
[b] Victorian Institute of Technology, Australia
[c] Computer Techniques Engineering Department, Mazaya University College, Nasiriyah, Iraq
[d] Technical College, Imam Ja'afar Al-Sadiq University, Baghdad, Iraq
[e] Iraqi Commission for Computers and Informatics (ICCI), Baghdad, Iraq
[f] Department of Computer Sciences, University of Jaén, Jaén 23071, Spain
[g] Ministry of Education, Thi-Qar Education Directorate, Iraq
[h] Department of Basic Sciences, Higher Technological Institute, Tenth of Ramadan City, Egypt
[i] Hourani Center for Applied Scientific Research, Al-Ahliyya Amman University, Amman19328, Jordan
[j] Institute of Informatics and Computing in Energy, Department of Computing, College of Computing and Informatics, Universiti Tenaga Nasional, Kajang 43000, Malaysia

ARTICLE INFO

Keywords:
Trustworthy
Artificial intelligence
q-ROF2TL-FWZIC
q-ROF2TL-CODAS
Multi-attribute decision-making

ABSTRACT

The purpose of this paper is to propose a novel hybrid framework for evaluating and benchmarking trustworthy artificial intelligence (AI) applications in healthcare by using multi-criteria decision-making (MCDM) techniques under a new fuzzy environment. To develop such a framework, a new decision matrix has been built, and then integrated with q-ROF2TL-FWZIC (q-Rung Orthopair Fuzzy 2-Tuple Linguistic Fuzzy-Weighted Zero-Inconsistency) and q-ROF2TL-CODAS (q-Rung Orthopair Fuzzy 2-Tuple Linguistic Combinative Distance-Based Assessment). In this integration, q-ROF2TL-FWZIC is utilized for assigning the weights of evaluation attributes of trustworthy AI, while q-ROF2TL-CODAS is employed for benchmarking trustworthy AI applications. Findings show that the q-ROF2TL-FWZIC method effectively weights the evaluation attributes. The transparency attribute receives the highest importance weight (*0.173566825*), whereas the human agency and oversight criterion has the lowest weight (*0.105741901*). The remaining attributes are distributed in between. Moreover, alternative_4 receives the highest rank order (score of 7.370410417), while alternative_13 receives the lowest rank order (score of −4.759794397). To evaluate the validity of the proposed framework, systematic ranking and sensitivity analysis assessments were employed.

1. Introduction

Recently, there has been a noticeable interest in using mobile applications extensively (Albahri et al., 2022), and the popularity of these apps is dramatically increasing. As people use it daily worldwide, statistics indicate that applications users reached in the year 2019 almost 2.5 billion (Holl & Elberzhager, 2019). Due to their growing use, apps have a very diverse range in their content (Ouhbi et al., 2015), commerce (Tang, 2019), medicine (Lakhan et al., 2023), education (Albrecht et al., 2018), and more. The majority of these apps are available at no cost, and the market was larger than $90 billion in 2018 (Freier, 2018), Furthermore, it is anticipated that this growth will persist in the coming years (Cheney & Thompson, 2018). These apps have been integrated and infused with some of the most advanced and recent technologies. Artificial intelligence (AI), considered amongst the most well-known and huge advanced technologies, has been integrated into various application domains for multiple purposes, such as ensuring better customer service, interactivity, 24/7 support, and reducing errors. AI has been integrated into various domains, and many of its users are deeply invested and accustomed to such technology; nevertheless,

* Corresponding author.
E-mail addresses: mohamed_asem@uomosul.edu.iq (M.A. Alsalem), alamoodi.abdullah91@gmail.com (A.H. Alamoodi), osamahsh89@gmail.com (O.S. Albahri), ahmed.bahri1978@gmail.com (A.S. Albahri), ryera@ujaen.es (R. Yera), Ali_Duhaim@utq.edu.iq (A.M. Duhaim), iman_sharaf@hotmail.com (I.M. Sharaf).

https://doi.org/10.1016/j.eswa.2023.123066
Received 30 September 2023; Received in revised form 19 December 2023; Accepted 25 December 2023
Available online 13 January 2024

some areas where AI has been integrated have shown tremendous benefits in comparison with others, including medicine. According to (Albahri et al., 2023), the medical sector is one of the highly promising fields interested in the utilization of AI apps, where numerous AI-based medical applications have been developed and presented considerable success (Albahri, Zaidan, Zaidan, et al., 2022; Yu et al., 2018). The integration of AI has significantly aided clinicians in developing diagnostic tools, offering explanations for clinical reasoning, and selecting appropriate treatments (Rong et al., 2020). Applications of this nature have played a crucial role in identifying numerous risks associated with diagnosing health aspects in clinical practices across hospitals worldwide (Albahri et al., 2023). That can be attributed to its ability to extract novel and significant insights from the vast amount of data generated daily during the delivery of healthcare (Administration, 2022; Alamoodi et al., 2020). This technology also holds the potential to significantly reduce costs associated with complications in chronic diseases (Darras et al., 2019). Given its widely recognized benefits and potential, many organizations worldwide have begun advocating for its utilization in their respective applications. One notable example is the U.S. Food and Drug Administration (FDA), which has authorized the use of AI in the medical field to improve healthcare services and mitigate health risks (Yu et al., 2018). This bolsters efforts by the FDA and the European Union (EU) to develop AI based apps for diagnosing diseases that take into consideration ethical concerns as safety, accountability, transparency, and privacy (Administration, 2022). As per the statistics of International Data Corporation, expected the spending on AI applications will raise to 97.9 billion dollars (USD) in 2023 (Markus et al., 2021). With enormous AI applications related to healthcare and their impact, we must ensure that all these apps we rely on are reliable and trustworthy. At the same time, despite the great benefits offered to the healthcare sector by using AI-based apps, they face a critical challenge related to trustworthiness (Kaur et al., 2021; Thiebes et al., 2021). Trust in applications-based AI in healthcare will raise the providing outstanding medical services and ensuring good care to patients (Leung et al., 2022; Markus et al., 2021).

In contrast, the lack of trustworthiness in AI poses a critical barrier to the widespread deployment of modern AI-based applications. There is a myriad of challenges stemming from the growing ethical and legal concerns surrounding these applications, which become more pronounced considering that medical and clinical decisions directly impact people's well-being. Additionally, the brittle trustworthiness of AI exacerbates issues with clinical decision-making and undermines accountability for errors (Fan et al., 2020; Topol, 2019). Even worse, these are only a few instances; if AI is used more widely, these risks will almost likely grow. This could have serious repercussions for society as a whole. However, because AI is now a necessary business capacity, companies cannot neglect AI's risks through avoiding AI totally. Instead of that, they ought to learn to recognize and successfully manage AI risks efficiently (Saif & Ammanath, 2020). Therefore, the concept of *trustworthiness AI* is still unclear. However, the literature places a strong emphasis on three main concepts of trustworthy AI: morality, robustness, and legitimacy. The ethical aspect denotes adherence to moral principles and values, the robustness aspect alludes to concerns about security and safety and the legal aspect indicates compliance with laws and regulations (Kale et al., 2022; Kaur et al., 2022; Liang et al., 2022). Additionally, putting fundamental concepts of trustworthy AI into practice is an open-challenge. Trust is "*The degree to which a user or other stakeholder has confidence that a product or system will behave as intended*" (Mattioli et al., 2023). The term "trustworthy AI" is described of AI distinguished that ethically adherent, lawful, and technically robust. It is predicated on the notion that trust of AI can be built at every stage of AI's lifecycle, from design to use, in order for AI to realize its full potential. To better address the challenges related to AI trustworthiness, recently, many significant efforts have been achieved by the government and several professional and scientific communities to make AI trustworthy (Kaur et al., 2022). In order to regulate and promote the development and operation of AI systems, the EU proposed ethical principles and guidelines for trustworthy AI.[1] The principles and guidelines state that trustworthy AI must adhere to all applicable rules and regulations, be ethical and uphold moral principles and values, and be resilient from a technical standpoint while taking into account its social environment (Hasani et al., 2022; Smuha, 2019). A framework for AI's accountability and responsible usage was released by the U.S. Government Accountability Office (GAO) (Kaur et al., 2022). In the same context, the National Institute of Standards and Technology (NIST) released a framework to gauge and enhance user trust in AI applications. According to the proposed framework, the trustworthiness characteristics of the AI system encompass reliability, accuracy, explainability, objectivity, resiliency, accountability, privacy, security, and safety (Stanton & Jensen, 2021). Additionally, the International Organization for Standardization (ISO) has proposed a framework for trustworthiness in AI, encompassing characteristics such as fairness, transparency, accountability, and controllability (Kaur et al., 2022). The study by Thiebes et al. (2021) presents a proposed framework comprising five foundational principles: (1) explainability, (2) autonomy non-maleficence, (3) non-maleficence, (4) justice, and (5) beneficence. Deloitte[2] has proposed a trustworthy AI framework that introduces six key dimensions, namely, not biased, fair, accountable and responsible, explainable and transparent, reliable and robust, secure, privacy, which have to be considered in all phases of building AI systems (Saif & Ammanath, 2020). According to statistics, 86 % of users will trust and remain loyal to companies that use ethical AI principles (Barometer, 2019). With the availability of huge amounts of data and increased processing power, AI systems have been used in many high-stake applications. So, it becomes important to make these systems reliable and trustworthy (B. Li et al., 2023). The current studies in the field of AI hint at the need for considering "*Trust" as a design principle rather than an option*. Moreover, designing AI-based critical systems of healthcare and biomedicine requires proving their trustworthiness (Gille et al., 2020; Procter et al., 2023). Hence, various stakeholders, including regulators, developers, customers, reinsurance companies, and end-users, need to evaluate these dimensions for different purposes (Mattioli et al., 2023). However, the pivotal challenge lies in evaluating these systems, which becomes particularly evident due to the absence of a comprehensive set of trustworthy attributes for AI applications. Other contributing factors include the complexity of the subject, which incorporates qualitative concepts. Moreover, the difficulty arises from selecting relevant attributes, compounded by issues related to the contextual application, modeled based on various elements such as Operational Design Domain, intended domain of use, nature and roles of stakeholders, among others (Adedjouma et al., 2022; Mattioli et al., 2023).

To the best of our knowledge, the study by Albahri, et al. (Albahri et al., 2023) is considered one of the most recent and noticeable attempts to assess trustworthy applications in healthcare. However, its assessment approach was subjective, and this type of evaluation suffers from several drawbacks that could render its results inconsistent and unreliable in some cases. Moreover, the subjective assessment relied on individual opinions, thereby potentially suffering from biases and personal interpretations. This could lead to unfair evaluations, as comparisons and measurements become challenging in such an assessment environment. In certain instances, it lacks a clear and quantifiable measurement scale, making precise comparisons and rankings of entities challenging. Additionally, it may suffer from inconsistency over time. The lack of consistency compromises the validity and reliability of subjective evaluations as indicators of performance or quality. On the other hand, the significance of the trustworthiness attributes of AI applications is not equal. Moreover, benchmarking trustworthy AI applications to select the best ones poses a challenge, compounded by the difficulty of

[1] European Union Ethics guidelines for trustworthy AI.

[2] Trustworthy Artificial Intelligence (AI)™ | Deloitte US.

benchmarking these applications according to multiple attributes simultaneously (Mattioli et al., 2023). Therefore, there is a need to develop a systematic and user-friendly evaluation and benchmarking approach that can be easily adopted by practitioners from diverse backgrounds to assess and benchmark trustworthy AI applications.

Therefore, it is essential to develop a comprehensive platform that encompasses all the previous aspects of trustworthiness in the evaluation and benchmarking of AI applications. This platform must be flexible and capable of addressing the issues. Consequently, the incorporation of explicit and structured methods in decision-making processes that involve multiple criteria can significantly enhance the overall quality of decision-making (Khatari et al., 2019; Rahman et al., 2023), and a group of techniques grouped together under the term Multiple Criteria Decision-Making (MCDM), which can be considered useful for such purposes (Ali et al., 2023; Alsalem, Mohammed et al., 2022; Tzeng & Huang, 2011). MCDM is a branch of operational research that takes into consideration multiple criteria while making decisions under conditions that arise in a variety of real-world scenarios across various disciplines (Rahman et al., 2023; Tešić and Marinković, 2023). Many methods can be employed to handle MCDM for addressing real-world problems (Almahdi et al., 2019; Alsalem, Alamoodi et al., 2022; Khatari et al., 2019; Sahoo & Goswami, 2023). These methods not only assist decision-makers to organize problems but also enable them to analyze, rank, sort and assign scores to different alternatives (Albahri et al., 2023; Alsalem et al., 2022).

MCDM techniques can solve selection problems in a wide domain of management, the study of (Albahri et al., 2023) used MCDM for develop data-driven strategy for evaluating the organizational culture, engineering, Sharaf et al. (2024) propose a Architecture for selection 5G-radio access network, Vahdani et al., (2013) proposed new Modification of TOPSIS and used it for a robot selection problem., management (Enaizan et al., 2020), energy (Albahri et al., 2023), medical (Arshad et al., 2024; Limboo & Dutta, 2022), sports science (Hsu et al., 2023), communication (Al-Samarraay et al., 2024) (Ranjan et al., 2023), transportation (Moslem et al., 2023), etc. The MCDM methods have been widely employed to resolve the issues related to evaluating AI applications in healthcare (Alsalem et al., 2019; Saleh & Salaheldin, 2022; Zaidan et al., 2020). Despite previous efforts, there has been no study to date addressing the challenges of evaluating and benchmarking trustworthy AI applications (Amri and Abed, 2023; Obaid, 2023). This study aims to fill this research gap by proposing a new framework based on MCDM for the evaluation and benchmarking of trustworthy AI applications in healthcare. To deal with decision-making situations in uncertainty, the theory of Fuzzy number sets can be used (Ghoushchi & Sarvi, 2023; Nezhad et al., 2023; Ranjan et al., 2023; Saqlain & Saeed, 2024). To build the framework with incomplete weight information, this study proposes the use of q-ROF2TL-CODAS (Q-Rung Orthopair Fuzzy 2-Tuple Linguistic Combinative Distance-Based Assessment); the benefits of utilizing this method lies in its capability not only in covering the uncertainty of human cognition but also it provides DMs with a larger space to represent their decisions (Naz et al., 2022). In addition, another method, namely the fuzzy-weighted zero-inconsistency (FWZIC), has been actively employed to assign weights to attributes with zero inconsistencies. In this research, FWZIC has been extended under a q-rung orthopair fuzzy 2-tuple linguistic sets (q-ROF2TL) environment, and it is termed (q-ROF2TL-FWZIC). The outcome will result in a homogeneous fuzzy environment between q-ROF2TL-CODAS and q-ROF2TL-FWZIC to efficiently address ambiguity, uncertainty, and fuzziness issues in benchmarking the trustworthy AI applications in healthcare. The 2-tuple fuzzy linguistic representation model (Herrera & Martínez, 2000) is one of the most important approaches for linguistic decision-making problemsthat can represent accurately linguistic computing results. Based on intuitionistic (IFSs) and Pythagorean fuzzy sets (PFSs), Yager (2017) proposed the q-rung orthopair fuzzy sets (q-ROFSs). The sum of the q^{th} power of the orthopair, i.e., the membership grade, and the non-membership grade, is bounded by one. Increasing the rung increases the space of acceptable orthopairs which allows the expression of a wider range of fuzzy information (Yager, 2016). Later, Wei et al. (2018) defined q-rung orthopair fuzzy 2-tuple linguistic sets (q-ROF2TLSs). All these features for the selected fuzzy environment rendered it suitable for integration with proposed methods in this research (Wei et al., 2018). The primary contributions of this study are as follows:

(1) Developing a new decision matrix for trustworthy AI applications in healthcare, considering the intersection of various AI applications as alternatives and incorporating components of trustworthy AI as evaluation attributes.
(2) Formulating a novel extension of FWZIC, termed q-ROF2TL-FWZIC (q-Rung Orthopair Fuzzy 2-Tuple Linguistic Fuzzy-Weighted Zero-Inconsistency), designed to weigh the evaluation attributes involved in assessing trustworthy AI applications.
(3) Ranking the healthcare trustworthy AI applications using q-ROF2TL-CODAS.
(4) To validate the proposed benchmarking framework, it has been evaluated using several scenarios based on sensitivity analysis.

The paper is structured as follows: Section 2 reviews the methodology, including the phases of the proposed framework. Section 3 presents the discussion of results, while Section 4 reviews the evaluation results of the proposed framework. Research implications are presented in Section 5, and Section 6 concludes the study.

2. Methodology

This section outlines the methodology employed in this study, consisting of two phases as depicted in Fig. 1. The first phase involves the development of the Decision Matrix (DM), wherein evaluation attributes are identified and overlaid with alternatives (AI healthcare applications). In the second phase, the q-ROF2TL-FWZIC method is developed and applied to calculate the weights of evaluation attributes for trustworthy AI. Subsequently, the q-ROF2TL-CODAS method is employed to benchmark the trustworthy AI healthcare applications.

2.1. Phase 1: Decision matrix development

This section describes the proposed decision matrix for benchmarking trustworthy AI healthcare applications. The two predefined components (alternatives and criterion) are crossover in the proposed decision matrix. It includes seven attributes of evaluation (*Components of Trustworthy AI*) and 50 alternatives (*AI Healthcare Applications*); It allows decision-makers to assess each alternative against the evaluation criteria, taking into account multiple viewpoints, the decision matrix is shown in Table 1.

As shown in Table 1, this matrix is used for evaluation and benchmarking 50 healthcare apps according to 7 evaluation attributes, the evaluation data of the 50 healthcare apps according to 7 evaluation attributes will be adopted by the study (A. S. Albahri et al., 2023), where, the mentioned study achieved subjectively evaluation the healthcare apps. The evaluation data provided by (A. S. Albahri et al., 2023) were a linguistic terms ("Very High", "High", "Medium", "Very Low", "Low"), and they will be transformed into their equivalent numerical to be suitable to use in MCDM method.

2.2. Phase 2: Development phase

This section presents the integrated fuzzy MCDM methods for the evaluation and benchmarking of trustworthy AI healthcare applications that are applied for developing the decision matrix presented in phase 1. This process begins by assigning weights to the seven attributes constructed based on q-ROF2TL-FWZIC, while the q-ROF2TL-CODAS method is used for benchmarking trustworthy AI healthcare applications. Details of this integration are presented in the following sections.

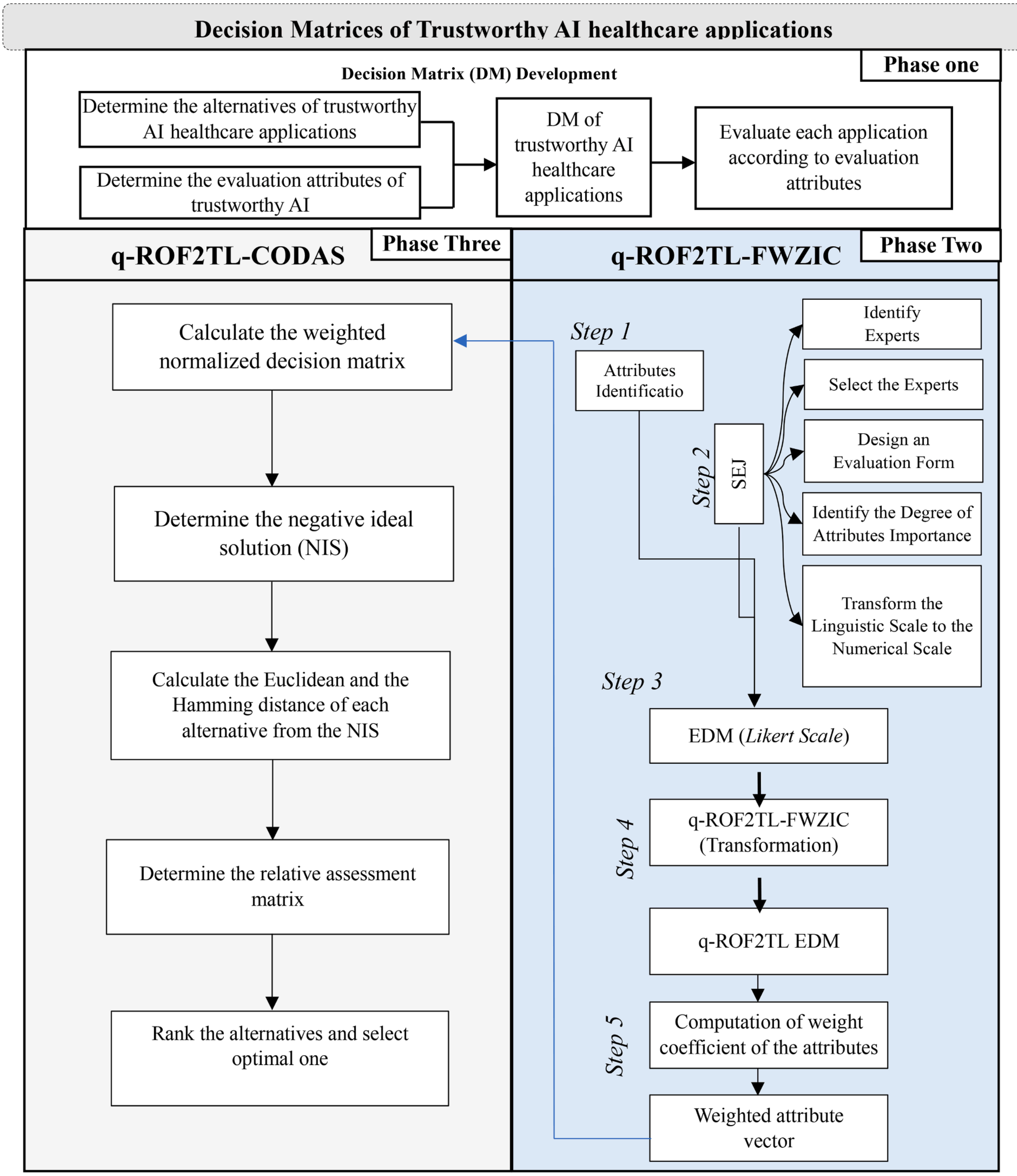

Fig. 1. Phases of the Proposed Framework.

2.2.1. q-Rung Orthopair Fuzzy 2-Tuple Linguistic Fuzzy-Weighted Zero-Inconsistency (q-ROF2TL-FWZIC)

The proposed extension of FWZIC utilized in this research is named q-ROF2TL-FWZIC. This method effectively calculates the weights of trustworthy AI attributes by taking into account the relative importance of each attribute and considering the degree of uncertainty or incompleteness in the available information. It involves five steps for calculating the weight of evaluation attributes, including human agency, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal and environmental well-being, and accountability (which requires auditability). The following steps are demonstrated as follows:

Step 1: Defining Evaluation Attributes: The evaluation attributes of trustworthy AI healthcare applications in the proposed decision matrix are human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal and environmental well-being, and accountability. Obtaining the subjective preferences for each attribute is discussed in the next step.

Step 2: Structured Expert Judgment (SEJ): The selection of specialised experts in the domain is needed to evaluate the relative significance of the attributes of trustworthy AI healthcare applications. In this context, experts have been defined as highly educated people with significant healthcare application experience based on AI and professional academic knowledge (i.e., *academic degrees*). Accordingly, specialists are gathered into a pool to determine the most appropriate ones. The next

Table 1
Decision Matrix for Evaluating Trustworthy AI Healthcare Applications.

	Evaluation Attributes (Criteria)						
Alternatives	C1	C2	C3	C4	C5	C6	C7
App 1	A_1/C_1	A_1/C_2	A_1/C_3	A_1/C_4	A_1/C_5	A_1/C_6	A_1/C_7
App 2	A_2/C_1	A_2/C_2	A_2/C_3	A_2/C_4	A_2/C_5	A_2/C_6	A_2/C_7
App 3	A_3/C_1	A_3/C_2	A_3/C_3	A_3/C_4	A_3/C_5	A_3/C_6	A_3/C_7
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
App 50	A_{50}/C_1	A_{50}/C_2	A_{50}/C_3	A_{50}/C_4	A_{50}/C_5	A_{50}/C_6	A_{50}/C_7

C = Evaluation attributes, **A** = Alternative, **C1**= "**Human agency and oversight** which includes the protection of fundamental rights, the involvement of human decision-making, and the need for human oversight ", **C2**= "**Technical robustness and safety:** which encompasses protection against security threats and the need for backup plans in case of system failure, as well as accuracy, reproducibility, and reliability ", **C3**= "**Privacy and data governance**: which includes ensuring the quality and integrity of data, respect for privacy, and enabling access to data ", **C4**= "**Transparency**: which entails the ability to trace and explain how decisions are made and to clearly communicate outcomes to stakeholders. ", **C5**=" **Diversity, non-discrimination and fairness**: which includes avoiding unfair bias, ensuring accessibility and universal design, and promoting stakeholder participation ", **C6**= "**Societal and environmental well-being**: which includes considerations of sustainability and environmental impact, social impact, and democracy ", **C7**= "**Accountability**: which requires auditability, minimisation and reporting of negative impact, trade-offs, and providing redress when necessary ".

Table 2
Linguistic terms with their equivalent numerical and q-ROF2TLSs.

Linguistic scoring scale	Numerical scoring scale	q-ROF2TLSs
Not important	1	$<(S_1,0)(0.8,0.3)>$
Slightly important	2	$<(S_2,0)(0.7,0.4)>$
Moderately important	3	$<(S_3,0)(0.6,0.5)>$
Important	4	$<(S_4,0)(0.7,0.4)>$
Extremely important	5	$<(S_5,0)(0.8,0.3)>$

sub-step is to create an evaluation form to obtain data from experts. The experts provide their preferences about each attribute's importance according to a 5-point Likert scale. The 5-point Likert scale has been used because it is minimally biased and highly reliable (Matell & Jacoby, 1971). Lastly, the used linguistic scale of preferences of experts will be converted to the numerical scale based on Table 2.

Step 3: Building Expert's decision matrix: It includes the alternatives and attributes as presented in Table 3.

As shown in Table 3 EDM, includes the significance score appointed by each expert to each criterion of the decision. where each expert crosse with every criterion. The Experts appoint the importance level (IL) to each criterion according to their judgment. IL(E_i/C_j) indicates the importance level appointed by (E_i) expert to (C_j) criterion **(Attribute).** These importance levels used to calculate the scores of weights for criteria **(Attributes).**

Step 4: Applying q-ROF2TLSs: This step entails fuzzifying the expert's decision matrix data using q-ROF2TLSs. The data (i.e., numerical scoring scale) are converted into q-ROF2TLS expert's decision matrix. The q-ROF2TLSs used in this process are shown in Table 2. Ambiguous and vague information is frequently problematic in MCDA, as assigning a precise preference to any attribute is difficult. The q-ROF2TLSs' advantage is that they confront uncertain, vague, and inconsistent information to estimate the relative importance of the attributes. The q-ROF2TLSs are defined, and the necessary operations are presented as follows.

A linguistic term set (LTS) $S=\{S_0,S_1,S_2,\cdots,S_g\}$ It is a predefined odd set of linguistic variables. For example, when $g=\ 2$, $S=\{S_0,S_1,S_2\}=\{low, medium, high\}$. When the indices of some labels in S are aggregated using a symbolic method, and the result is the actual number

Table 3
EDM.

	Experts				
Attributes	E_1	E_2	E_3	…	E_k
C_1	IL (E_1/C_1)	IL (E_2/C_1)	IL (E_3/C_1)	…	IL (E_k/C_1)
C_2	IL (E_1/C_2)	IL (E_2/C_2)	IL (E_3/C_2)	…	IL (E_k/C_2)
⋮	⋮	⋮	⋮	⋮	⋮
C_n	IL (E_1/C_n)	IL(E_2/C_n)	IL(E_3/C_n)	…	IL(E_k/C_n)

IL represents the importance level.

$\beta\in[0,g]$ which is not an integer, β is divided into two numerical values,α that represent the "difference in information" between β and the index of the linguistic term S_i Which is the closest to β in S. Then, a 2- tuple representation model (2TRM) is given by (S_i,α).

Definition (Herrera & Martínez, 2000): For a linguistic term set S and a real number $\beta\in[0,g]$ representing the result of aggregating linguistic symbols, the function Δ used to obtain the 2TRM equivalent to β is defined as

$$\Delta:[0,g]\rightarrow S\times[-0.5,0.5) \tag{1}$$

$$\Delta(\beta)=(S_i,\alpha),\begin{cases}S_i, i=round(\beta),\\ \alpha=\beta-i,\alpha\in[-0.5,0.5).\end{cases} \tag{2}$$

Definition (Herrera & Martínez, 2000): For a linguistic term set S and a 2-tuple (S,α) an inverse operation that converts the 2-tuple to the equivalent real number $\beta\in[0,g]$ is defined by

$$\Delta^{-1}:S\times[-0.5,0.5)\rightarrow[0,g], \tag{3}$$

$$\Delta^{-1}(S_i,\alpha)=i+\alpha=\beta \tag{4}$$

Definition (Z. Li et al., 2022**):** For a non-empty set X, a q-rung orthopair fuzzy set (q-ROFS) is denoted as

$$Q=\{\langle x,(\mu(x),\upsilon(x))\rangle|x\in X\} \tag{5}$$

where $(x):X\rightarrow[0,1]$,$\upsilon(x):X\rightarrow[0,1]$ are the membership and non-membership grades of each element x in $\widetilde{Q}$, respectively which satisfy

$$0\leq(\mu(x))^q+(\upsilon(x))^2\leq1,q\geq \text{for all } x\in X1 \tag{6}$$

Definition (Abbas et al., 2023**):** For a non-empty set X and a LTS S, a q-rung orthopair fuzzy 2-tuple linguistic set (q-ROF2TLS) is denoted by

$$Q=\{(S_{\theta(x)},\alpha),(\mu(x),\upsilon(x))|x\in X\} \tag{7}$$

where

$$S_{\theta(x)}\in S,\alpha\in[-0.5,0.5),0\leq\mu(x)\leq1,\ 0\leq\upsilon(x)\leq1,$$

$$0\leq\mu^q(x)+\upsilon^q(x)\leq1(q\geq1),$$

$\mu(x)$ and$\upsilon(x)$ Represent the degree of membership and degree of non-membership of the element x to the linguistic term set. $(S_{\theta(x)},\alpha)$.

Definition (Li et al., 2022): For a q-ROF2TLS $\{(S_a,\alpha_a),(\mu_a,\upsilon_a)\}$ the score function is computed by

$$ScoreQ=\frac{\Delta^{-1}(S_a,\alpha_a).(1+\mu_a^q+\upsilon_a^q)}{2(g+1)} \tag{8}$$

Definition (Li et al., 2022): The multiplication of a q-ROF2TLS $\{(S_a,\alpha_a),(\mu_a,\upsilon_a)\}$ by scalar number is defined as follows:

$$\lambda Q=\langle\Delta(\lambda\Delta^{-1}(S_a,\alpha_a)),\left(\left(1-(1-\mu_a^q)^\lambda\right)^{1/q},\upsilon_a^q\right)\rangle,\lambda>0. \tag{9}$$

Definition (Li et al., 2022): The weighting averaging aggregation operator of a family of q-ROF2TLSs $Q = \{Q_1, Q2, \cdots, Q_n\}$

$$q-\mathrm{ROF2TLWA}\{Q, Q2, \cdots, Q_n\} = \left\{\Delta\left(\sum_{i=1}^{n} w_i \Delta^{-1}(S_i, \alpha_i)\right), \left(\left(1-\prod_{i=1}^{n}(1-\mu_i^q)\right)^{\frac{1}{q}}, \prod_{i=1}^{n} \upsilon_i^q\right)\right\},$$

$$w_i > 0, \sum_{i=1}^{n} w_i = 1. \tag{10}$$

Step 5: Computing the Final Weight Value: Based on the fuzzified data of the attributes in the previous step, the final values of the weight coefficients of the attributes $(w1, w2, \cdots, wn)^T$ Are calculated in this step as follows. The ratio of fuzzified data is computed using Eq. (8) and (9). Eq. (11) symbolises the process, as shown in Table 4.

$$R(E_k/C_j) = \frac{IL(E_k/C_j)}{\sum_{j=1}^{n} IL(E_k/C_{kj})}, \text{for } k = 1,2,3,..K \text{ and } j = 1,2,3,..n \tag{11}$$

where $R(E_k/C_j)$ represents the q-ROF2TL of $R(E_k/C_j)$.

- The mean values are calculated to obtain the final fuzzy values of the weight coefficients of the attributes $\left(\widetilde{w}_1, \widetilde{w}_2, \cdots, \widetilde{w}_n\right)^T$. Using *Eq. (10)*, each value of the q-ROF2TL-EDM is then computed. Eq. (12) depicts the actual procedure of this step symbolically.

$$W_j = \frac{\sum_{k=1}^{K} R(E_k/C_j)}{K}, \text{for } k = 1,2,3,..K \text{ and } j = 1,2,3,..n. \tag{12}$$

- The final weight is obtained after the defuzzification using Eq. (8). Each attribute's weight importance should be assigned whilst considering the entire weight sum to rescale and apply.

2.2.2. *q-Rung Orthopair Fuzzy 2-Tuple Linguistic Combinative Distance-Based Assessment (q-ROF2TL-CODAS)*

In this subsection, the CODAS method (Keshavarz Ghorabaee et al., 2016) integration with q-ROF2TLSs is presented. This method calculates the preference of alternatives using two different distance measures. The largest and the most critical measurement is the Euclidean distance (ED) between the alternatives and the negative-ideal solution (NIS), and the second measure is the Hamming distance (HD). The alternative, which has a greater distance from the NIS, is preferable. The ED and HD measures are used for the relative assessment (RA) of alternatives to construct the RA-based matrix to fuse the information. The technique of implementing the q-ROF2TL-CODAS approach is described below:

Table 4
q-ROF2TL-EDM.

Attributes	Experts E_1	E_2		E_k
C_1	$\frac{\widetilde{IL}(E_1/C_1)}{\sum_{j=1}^{n} IL(E_1/C_{1j})}$	$\frac{\widetilde{IL}(E_2/C_1)}{\sum_{j=1}^{n} IL(E_2/C_{2j})}$	$\cdots$	$\frac{\widetilde{IL}(E_k/C_1)}{\sum_{j=1}^{n} IL(E_k/C_{kj})}$
C_2	$\frac{\widetilde{IL}(E_1/C_2)}{\sum_{j=1}^{n} IL(E_1/C_{1j})}$	$\frac{\widetilde{IL}(E_2/C_2)}{\sum_{j=1}^{n} IL(E_2/C_{2j})}$	$\cdots$	$\frac{\widetilde{IL}(E_k/C_2)}{\sum_{j=1}^{n} IL(E_k/C_{kj})}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
C_n	$\frac{\widetilde{IL}(E_1/C_n)}{\sum_{j=1}^{n} IL(E_1/C_{1j})}$	$\frac{\widetilde{IL}(E_2/C_n)}{\sum_{j=1}^{n} IL(E_2/C_{2j})}$	$\cdots$	$\frac{\widetilde{IL}(E_k/C_n)}{\sum_{j=1}^{n} IL(E_k/C_{kj})}$

- ***Step 1.*** Switch the linguistic information in each expert decision matrix into q-ROF2TLSs.

$$EM^k = \begin{bmatrix} \eta_{11}^k & \eta_{12}^k & \cdots & \eta_{1n}^k \\ \eta_{21}^k & \eta_{22}^k & \cdots & \eta_{2n}^k \\ \cdots & \cdots & \vdots & \cdots \\ \eta_{m1}^k & \eta_{m2}^k & \cdots & \eta_{mn}^k \end{bmatrix} \rightarrow \widetilde{EM}^k = \begin{bmatrix} \widetilde{\eta}_{11}^k & \widetilde{\eta}_{12}^k & \cdots & \widetilde{\eta}_{1n}^k \\ \widetilde{\eta}_{21}^k & \widetilde{\eta}_{22}^k & \cdots & \widetilde{\eta}_{2n}^k \\ \cdots & \cdots & \vdots & \cdots \\ \widetilde{\eta}_{m1}^k & \widetilde{\eta}_{m2}^k & \cdots & \widetilde{\eta}_{mn}^k \end{bmatrix}, \widetilde{\eta}_{ij}^k$$

$$= \{(S_{ij}, \alpha_{ij}), (\mu_{ij}, \upsilon_{ij})\}^k, S_{ij} \in \{S_0, \cdots, S_g\}$$

where $\widetilde{\eta}_{ij}^k$ is the assessment of the k^{th} expert of the i^{th} alternative $(i = 1,2,3,..m)$ for the j^{th} attribute $(j = 1,2,3,..n)$ expressed as a q-ROF2TLS to emulate the doubtfulness and uncertainty associated with human evaluations.

- ***Step 2.*** By utilizing Eq. (13), independent panel evaluations can be combined to form the fused q-ROF2TLSs matrix

$$\widetilde{F} = \begin{bmatrix} \widetilde{\eta}_{11} & \widetilde{\eta}_{12} & \cdots & \widetilde{\eta}_{1n} \\ \widetilde{\eta}_{21} & \widetilde{\eta}_{22} & \cdots & \widetilde{\eta}_{2n} \\ \cdots & \cdots & \vdots & \cdots \\ \widetilde{\eta}_{m1} & \widetilde{\eta}_{m2} & \cdots & \widetilde{\eta}_{mn} \end{bmatrix} \tag{13}$$

- ***Step 3.*** Calculate the weighted q-ROF2TLSs matrix using *Eq. (14)* as follows:

$$Fw = w_j \eta_{ij}, \tag{14}$$

where w_j is the attribute weight, and $0 \le w_j \le 1, \sum_{J=1}^{n} w_j = 1$.

- ***Step 4.*** Find the NIS using the q-ROF2TLSs score function (1). If the score function is similar, the accuracy function is used to rank the q-ROF2TLSs:

$$\widetilde{NIS} = \left[\widetilde{NIS}_J\right]_{1\times n;} \tag{15}$$

$$\widetilde{NIS}_J = \min_i \widetilde{\eta}_{ij} \tag{16}$$

- ***Step 5.*** Calculate the weighted *ED* and *HD* as follows.

Before proceeding with this step, the employed distance measures are first defined.

Definition (Z. Li et al., 2022): For two q-ROF2TLS $Q_a = \{(S_a, \alpha_a), (\mu_a, \upsilon_a)\}$ and $Q_b = \{(S_b, \alpha_a), (\mu_b, \upsilon_b)\}$ the normalized Hamming distance between them is measured by

$$d_H(Q_a, Qb) = \frac{1}{2(g+1)}\left[\left|\left(1+\mu_a^q+\upsilon_a^q\right) \bullet \Delta^{-1}(S_a, \alpha_a) - \left(1+\mu_b^q+\upsilon_b^q\right) \bullet \Delta^{-1}(S_b, \alpha_b)\right|\right]. \tag{17}$$

Definition: The Euclidean distance between two q-ROF2TLS $\underset{a}{\breve{Q}} = \{(S_a, \alpha_a), (\mu_a, \upsilon_a)\}$ and $Qb = \{(S_b, \alpha_a), (\mu_b, \upsilon_b)\}$ is measured by

$$d_E(Qb,Qb) = \left(\left(\frac{1}{(g+1)} \bullet \left|\left(\Delta^{-1}(S_a,\alpha_a) \bullet \mu_a\right)^q - \left(\Delta^{-1}(S_b,\alpha_b) \bullet \mu_b\right)^q\right|\right)^q + \left(\frac{1}{(g+1)} \bullet \left|\left(\Delta^{-1}(S_a,\alpha_a) \bullet \upsilon_a\right)^q - \left(\Delta^{-1}(S_b,\alpha_b) \bullet \upsilon_b\right)^q\right|\right)^q\right)^{1/q} \tag{18}$$

Proposition: Let d_E Be the Euclidean distance between two q-ROF2TLS $\underset{a}{\breve{Q}}$ and $\underset{b}{\breve{Q}}$, then the following properties hold

i. $d_E\left(\underset{a}{\breve{Q}}, \underset{b}{\breve{Q}}\right) \geqslant 0$;

ii. $d_E\left(\underset{a}{\breve{Q}}, \underset{b}{\breve{Q}}\right) = d_E\left(\underset{b}{\breve{Q}}, \underset{a}{\breve{Q}}\right)$;

iii. $d_E\left(\underset{a}{\breve{Q}}, \underset{b}{\breve{Q}}\right) = 0 \text{ if } \underset{a}{\breve{Q}} = \underset{b}{\breve{Q}}$;

iv. If $\underset{a}{\breve{Q}} \geqslant \underset{b}{\breve{Q}} \geqslant \underset{c}{\breve{Q}}$, then $d_E\left(\underset{a}{\breve{Q}}, \underset{c}{\breve{Q}}\right) \geqslant d_E\left(\underset{a}{\breve{Q}}, \underset{b}{\breve{Q}}\right)$;

Proof. The first three properties are trivial. We need to prove the fourth property only.

If $Qa \geq Qb \geq Qc$, we have $\Delta^{-1}(S_a,\alpha_a) \geq \Delta^{-1}(S_b,\alpha_b) \geq \Delta^{-1}(S_c,\alpha_c)$, $\mu_a \geq \mu_b \geq \mu_c$, and

$\upsilon_a \leq \upsilon_b \leq \upsilon_c$. Therefore

$$\left|\left(\Delta^{-1}(S_a,\alpha_a) \bullet \mu_a\right)^q - \left(\Delta^{-1}(S_c,\alpha_c) \bullet \mu_c\right)^q\right| \geq \left|\left(\Delta^{-1}(S_a,\alpha_a) \bullet \mu_a\right)^q - \left(\Delta^{-1}(S_b,\alpha_b) \bullet \mu_b\right)^q\right|,$$

and

$$\left|\left(\Delta^{-1}(S_a,\alpha_a) \bullet \upsilon_a\right)^q - \left(\Delta^{-1}(S_c,\alpha_c) \bullet \upsilon_c\right)^q\right| \geq \left|\left(\Delta^{-1}(S_a,\alpha_a) \bullet \upsilon_a\right)^q - \left(\Delta^{-1}(S_b,\alpha_b) \bullet \upsilon_b\right)^q\right|.$$

Hence,

$$d_E(Q_a,Q_b) = \left(\left(\frac{1}{(g+1)} \bullet \left|\left(\Delta^{-1}(S_a,\alpha_a) \bullet \mu_a\right)^q - \left(\Delta^{-1}(S_b,\alpha_b) \bullet \mu_b\right)^q\right|\right)^q + \left(\frac{1}{(g+1)} \bullet \left|\left(\Delta^{-1}(S_a,\alpha_a) \bullet \upsilon_a\right)^q - \left(\Delta^{-1}(S_b,\alpha_b) \bullet \upsilon_b\right)^q\right|\right)^q\right)^{1/q}$$
$$\leq \left(\left(\frac{1}{(g+1)} \bullet \left|\left(\Delta^{-1}(S_a,\alpha_a) \bullet \mu_a\right)^q - \left(\Delta^{-1}(S_c,\alpha_c) \bullet \mu_c\right)^q\right|\right)^q + \left(\frac{1}{(g+1)} \bullet \left|\left(\Delta^{-1}(S_a,\alpha_a) \bullet \upsilon_a\right)^q - \left(\Delta^{-1}(S_c,\alpha_c) \bullet \upsilon_c\right)^q\right|\right)^q\right)^{\frac{1}{q}} = d_E(Q_a,Qc).$$

Eq.s (19) and (20) are utilized to find the distance between each alternative and the NIS.

$$ED_i = \sum_{j=1}^{n} ED\left(\widetilde{\eta}_{ij}, \widetilde{NIS_j}\right); \tag{19}$$

$$= \sum_{j=1}^{n} HD\left(\widetilde{\eta}_{ij}, \widetilde{NIS_j}\right). \tag{20}$$

- ***Step 6***. Using the following Eqs., build the relative assessment matrix RA:

$RA = [h_{i\ell}]_{m\times m;}$

$$h_{i\ell} = (ED_i - ED_\ell) + (\phi(ED_i - ED_\ell) \times (HD_i - HD_\ell)) \tag{21}$$

where $\ell \in \{1,2,3,\ldots,m\}$ Moreover, ϕ denotes a threshold function that could be designed by:

$$\phi(x) = \begin{cases} One \;\; if|x| \leq \vartheta \\ 0 \;\; if|x| > \vartheta \end{cases} \tag{22}$$

Each pair of alternatives is compared using the ED. If the alternatives are not comparable by this distance measure, then the HD is employed. In other words, the HD is used only when the ED between the two alternatives is very close. The main purpose of the threshold function is to recognize the equality of the ED. For two alternatives, if the difference between the ED is less than a threshold value, then the HD is used for comparison. The threshold value is usually assigned a value $\vartheta \in [0.01, 0.05]$ specified by DMs, $\vartheta = 0.2$ is commonly used.

- ***Step 7***. Derive the average solution (£$_i$) by using.

$$£_i = \sum_{\ell=1}^{m} h_{i\ell} \tag{23}$$

- ***Step 8***. Based on computing outcomes of £$_i$, all the alternatives can be ranked.

The alternative having greater distances from the negative-ideal solution is preferable. Hence, the best option is the one with the highest evaluation score.

3. Results and discussion

This section presents and discusses the results of the proposed evaluation and benchmarking framework for trustworthy AI applications in the healthcare domain. Section 3.1 describes the evaluation results of DM. Section 3.2 describes the results of the weight determination of trustworthy AI evaluation attributes using q-ROF2TL-FWZIC.Then, Section 3.3 presents the benchmarking results of benchmarking and ranking trustworthy AI healthcare applications using q-ROF2TL-CODAS.

3.1. DM evaluation results

This section presents the results of DM's evaluation. The evaluation results are presented according to the following Table 5., The DM includes the 50 alternatives (A1 to A50) and seven criteria (C1 to C7) as mentioned in section 2.1., The numerical values in the cells of the matrix represent the evaluation score of each alternative for each criterion.

3.2. Weighting results using q-ROF2TL-FWZIC

This section presents the weights results of the evaluation attributes using the q-ROF2TL-FWZIC method developed in Section 2.2.1. The first

Table 5
Evaluation Decision Matrix Results.

	Evaluation Attributes						
Alternative	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**
A 1 (Ho & Caals, 2021)	1	1	3	1	2	3	1
A 2 (El-Sappagh et al., 2021)	1	1	1	5	1	5	1
A 3 (Rostami & Oussalah, 2022)	1	1	1	5	1	4	1
A 4 (Rahman et al., 2021)	3	3	4	1	4	5	1
A 5 (Rahman et al., 2021)	1	1	1	5	4	5	1
A 6 (Müller et al., 2020)	1	1	1	4	1	1	1
A 7 (Lucieri et al., 2020)	1	1	1	4	1	2	1
A 8 (Karim et al., 2022)	1	1	1	1	1	5	1
A 9 (Larasati et al., 2021)	1	1	2	4	1	1	1
A 10 (Rathi et al., 2021)	1	1	1	1	1	4	1
A 11 (Du et al., 2022)	1	1	1	5	1	4	1
A 12 (Zerka et al., 2020)	1	1	3	3	4	3	1
A 13 (Wang et al., 2022)	1	1	1	1	1	3	1
A 14 (Delacroix & Wagner, 2021)	1	1	1	1	1	3	5
A 15 (Chou et al., 2022)	1	1	3	3	1	1	1
A 16 (Shi et al., 2019)	1	1	3	3	1	1	1
A 17 (Abou-Nassar et al., 2020)	1	1	1	3	4	4	1
A 18 (Hussein et al., 2012)	1	1	1	1	1	4	1
A 19 (Guiñazú et al., 2020)	1	3	3	1	1	3	1
A 20 (Rehman et al., 2021)	1	1	1	3	3	3	1
A 21 (Collins et al., 2022)	3	1	3	4	4	3	1
A 22 (Kerasidou, 2021)	1	1	4	5	1	2	1
A 23 (Nicora et al., 2022)	1	1	1	1	1	5	1
A 24 (Zarour et al., 2020)	1	1	1	1	3	5	1
A 25 (Sachan et al., 2021)	1	1	1	4	1	4	1
A 26 (Lucieri et al., 2022)	1	1	1	4	1	3	1
A 27 (Müller et al., 2022)	1	1	1	4	1	1	1
A 28 (Arrieta et al., 2020)	3	1	3	4	3	3	1
A 29 (Deperlioglu et al., 2022)	1	1	1	4	1	4	1
A 30 (Setchi et al., 2020)	1	1	1	4	2	2	1
A 31 (Angerschmid et al., 2022)	3	1	5	4	1	1	1
A32 (Sheikh et al., 2021)	1	3	1	1	4	4	1
A33 (Barclay & Abramson, 2021)	1	1	1	4	4	1	1
A34 (Pal, 2020)	1	1	1	4	1	1	1
A 35 (Holzinger et al., 2022)	1	1	1	4	3	4	1
A 36 (Martínez-Agüero et al., 2022)	1	1	1	4	1	4	1
A 37 (Saheb et al., 2021)	1	3	3	1	1	3	1
A 38 (Balagurunathan et al., 2021)	1	1	1	3	1	4	3
A 39 (Bania & Halder, 2021)	1	1	3	1	1	3	1
A 40 (Saba et al., 2020)	1	1	1	1	3	3	1
A 41 (Washington et al., 2020)	1	1	1	1	3	3	1
A 42 (Leal et al., 2021)	1	1	1	4	4	3	1
A 43 (Rieke et al., 2020)	1	1	1	1	3	3	1
A 44 (Wenzel & Wiegand, 2020)	1	1	1	1	3	3	1
A 45 (Oprescu et al., 2022)	1	1	3	1	4	4	1
A 46 (Holzinger et al., 2021)	1	1	1	4	1	1	1
A 47 (Séroussi et al., 2020)	1	1	1	4	3	1	1
A 47 (González-Gonzalo et al., 2022)	3	1	3	3	4	4	1
A 49 (Yang et al., 2022)	1	1	1	5	1	1	1
A 50 (Abdar et al., 2023)	1	1	1	1	1	4	1

C = Evaluation attributes, **A** = Alternative (AI Healthcare Applications), **C1**= "Human agency and oversight", **C2**= "Technical robustness and safety", **C3**= "Privacy and data governance", **C4**= "Transparency", **C5**=" Diversity, non-discrimination and fairness", **C6**= "Societal and environmental well-being", **C7**= "Accountability".

Table 6
EDM.

Expert	C1	C2	C3	C4	C5	C6	C7
E1	2	4	4	5	3	2	3
E2	3	4	5	4	5	3	4
E3	2	3	4	4	4	3	5

step in q-ROF2TL-FWZIC is the evaluation attributes identification (i.e., ***C1, C2***…***C7***), and then the opinions of 3 specialist experts are collected regarding the significance level of each evaluation attribute by evaluation form developed for this purpose, the collected preferences of experts are converted into an equivalent scoring scale as appeared in Table 2. After that, the EDM is built, as mentioned in the third step, as shown in Table 6. The values in Table 6 represent the significance level for each evaluation attribute according to the expert preference. According to Step 4, the EDM converted into q-ROF2TLS-EDM, as presented in Table 7, where the crisp values transformed to equivalent fuzzy numbers based on Table 2. Finally, the final weight is obtained after the defuzzification using Eq. (8). Table 8 presents the computed final weights of the seven evaluation attributes of trustworthy AI.

The weighting results of the seven evaluation attributes based on the extended q-ROF2TL-FWZIC are shown in Table 8. The weight results are ordered from the most significant importance weight to the smallest one, where the most excellent importance weight (*0.173566825*) is that of **C4**, followed by **C3** with an importance weight of (*0.16933294*) and **C7** with an importance weight of (*0.154864455*). **C1** received the lowest importance weight (*0.105741901*), followed by **C6** with an importance weight of (*0.107250068*). The weights of the remaining attributes are distributed in between.

3.3. Benchmarking results using q-ROF2TL-CODAS

This subsection presents the results of applying q-ROF2TL-CODAS for benchmarking and ranking trustworthy AI applications. According to the weights of the criteria presented in Table 8, the weighted normalized decision matrix can be calculated; after that, the negative-ideal solution is determined by using Eqs. (15), 16), and then Euclidean distance ED (Ed - Ei) and Taxicab Distances (TDs-Ti) are calculated by using Eqs. (19), 20) as presented in Table 9, and then building the relative assessment matrix RA by using Eqs. (21),22), finally using Eq. (23) to find the assessment scores (H) of alternatives as presented in Table 10.

As indicated in Table 10, A4 attains the highest assessment score, signifying its status as the top-performing option, followed by A5. Conversely, A13 receives the lowest assessment score, trailed by A50, while the remaining alternatives fall in between.

3.4. Framework evaluation

The following subsections describe the evaluation and testing processes for the results of the proposed benchmarking framework, which integrates q-ROF2TL-FWZIC with q-ROF2TL-CODAS. These processes involve two assessment stages.

3.5. Systematic ranking

This section conducted a systematic ranking assessment to evaluate the results of the ranking of trustworthy AI applications. This process is well-established and has been performed in previous MCDM studies (Albahri et al., 2021; Qader et al., 2017). A validation of the benchmarking results for trustworthy AI applications is conducted, it involving the procedures outlined as follows:

- The decision matrix is sorted according to trustworthy AI applications' final rank results.
- The trustworthy AI applications are divided into five groups.
- The mean and standard deviation (STD) for each group is calculated afterwards.

Once the mean and STD have been computed for each of the five groups, it is necessary to compare these results. This step ensures the credibility of the systematic ranking and follows certain conditions based on the q-ROF2TL-CODAS method during the comparison process. In particular, the group with the lowest mean value signifies its validity.

Table 7
Q-rof2tls-edm.

Expert	C1	C2	C3	C4	C5	C6	C7
E1	(S2,0)(0.7,0.4)	(S4,0)(0.7,0.4)	(S4,0)(0.7,0.4)	(S5,0)(0.8,0.3)	(S3,0)(0.6,0.5)	(S2,0)(0.7,0.4)	(S3,0)(0.6,0.5)
E2	(S3,0)(0.6,0.5)	(S4,0)(0.7,0.4)	(S5,0)(0.8,0.3)	(S4,0)(0.7,0.4)	(S5,0)(0.8,0.3)	(S3,0)(0.6,0.5)	(S4,0)(0.7,0.4)
E3	(S2,0)(0.7,0.4)	(S3,0)(0.6,0.5)	(S4,0)(0.7,0.4)	(S4,0)(0.7,0.4)	(S4,0)(0.7,0.4)	(S3,0)(0.6,0.5)	(S5,0)(0.8,0.3)

Table 8
Final Weights.

Evaluation Attributes	Weight	Fuzzy Weight
C4	0.173566825	(S3,-0.0190)(0.6259,0.6290)
C3	0.16933294	(S3,-0.0524)(0.6185,0.6351)
C7	0.154864455	(S3,-0.1990)(0.6018,0.6554)
C5	0.152337199	(S3,-0.2191)(0.5967,0.6592)
C2	0.136906613	(S3,-0.3392)(0.5602,06829)
C6	0.107250068	(S2,0.1939)(0.5352,0.7036)
C1	0.105741901	(S2,0.0405)(0.5637,0.6799)

In the comparison process, and to further ensure the result validity according to the q-ROF2TL-CODAS philosophy, the following conditions need to be met:

- The mean of the first group is assumed to be the highest when checking result validity.
- The mean of the first group must be higher than the mean of the second group.
- The mean of the second group must be lower than the mean of the first group and higher than the mean of the third group.
- This pattern continues for subsequent groups, where the mean of each group should be lower than the previous group and higher than the subsequent group.

By fulfilling these conditions, the validity of the systematic ranking is ensured by the q-ROF2TL-CODAS approach. The validation results for the group outcomes obtained through the proposed methodology are displayed in Table 11.

The group results were validated to assess the consistency and reliability of the rankings. Table 11 presents the validation results for the groups G1, G2, G3, G4 and G5, along with their corresponding mean and standard deviation values. The mean values in each group represent the average performance or score of the alternatives within that group, while the standard deviation indicates the variability or dispersion of the scores within the group. These validation results serve as an indicator of the consistency and agreement among the group rankings. The ranking results of all groups were found to be consistent with the q-ROF2TL-CODAS philosophy comparison conditions. This was determined by comparing the mean values of the first group in each pre-processing approach to the mean results of the second group. In all ranks, the mean value of the first group was observed to be larger than the mean value of the corresponding second group. This consistent pattern was applied across all ranks, indicating a valid ranking. Additionally, the STD was used to measure the variability around the mean. Most data points were observed to fall within one STD of the mean. This suggests that the data points were relatively close to the mean value, indicating a certain level of consistency and stability in the ranking results. In sum, the mean and STD values based on the statistical validation results indicated that the groups ranking based on q-ROF2TL-CODAS results of the trustworthy AI applications were valid and systematically ranked. The second critical evaluation analysis is presented in the next section.

4. Sensitivity analysis

Sensitivity analysis estimates the influence of the most important criteria in terms of its relative weight on the alternative benchmarking results. Many related literatures such (Alamoodi et al., 2022; Alsalem et al., 2021; Albahri et al., 2022) have used sensitivity analysis as an assessment method to measure the sensitivity of the criteria' weights and analyze its change. Based on the q-ROF2TL-FWZIC weights for the evaluation attributes of trustworthy AI (C1 to C7), a sensitivity analysis can be conducted to assess the impact of variations in weights on the overall rankings of the alternatives. This analysis helps in understanding

Table 9
Calculation of (EDs) Ei, (TDs) Ti, and Assessment Score.

Alternative	Ei	Ti	Assessment Score	Alternative	Ei	Ti	Assessment Score
A1	0.06962731	0.02265817	−2.667248699	**A26**	0.1080356	0.03257909	−0.787359214
A2	0.1862112	0.10260363	3.403739112	**A27**	0.08044423	0.02614479	−2.09156601
A3	0.16457754	0.0786079	2.251138712	**A28**	0.21631611	0.05828629	4.651165884
A4	0.27027846	0.09852186	7.370410417	**A29**	0.13476546	0.04297967	0.595714777
A5	0.26349483	0.12758648	7.118423217	**A30**	0.09506699	0.04008404	−1.211881859
A6	0.08044423	0.02614479	−2.09156601	**A31**	0.20464648	0.08356291	4.125198092
A7	0.08647992	0.03175642	−1.694974769	**A32**	0.1492397	0.04583455	1.256467817
A8	0.07595488	0.0408306	−2.022317118	**A33**	0.15772787	0.05112764	1.69644532
A9	0.08766158	0.03299244	−1.612407348	**A34**	0.08044423	0.02614479	−2.09156601
A10	0.05432123	0.01683487	−3.423666997	**A35**	0.17514817	0.05264197	2.48428733
A11	0.16457754	0.0786079	2.251138712	**A36**	0.13476546	0.04297967	0.595714777
A12	0.18050763	0.04944051	2.72193293	**A37**	0.07867507	0.01834738	−2.335972143
A13	0.02759137	0.00643429	−4.759794397	**A38**	0.11368288	0.03087004	−0.516958476
A14	0.07810101	0.03212074	−2.08920799	**A39**	0.06104024	0.01433055	−3.24372205
A15	0.07563262	0.01802336	−2.494574988	**A40**	0.06797407	0.01609659	−2.894265255
A16	0.07563262	0.01802336	−2.494574988	**A41**	0.06797407	0.01609659	−2.894265255
A17	0.17378862	0.05194483	2.409338642	**A42**	0.18531923	0.05756194	3.043849142
A18	0.05432123	0.01683487	−3.423666997	**A43**	0.06797407	0.01609659	−2.894265255
A19	0.07867507	0.01834738	−2.335972143	**A44**	0.06797407	0.01609659	−2.894265255
A20	0.11015782	0.0262237	−0.725735642	**A45**	0.16505374	0.04971398	2.014903318
A21	0.25321704	0.07360684	6.448211604	**A46**	0.08044423	0.02614479	−2.09156601
A22	0.18124817	0.0879276	3.105343987	**A47**	0.12082693	0.03580709	−0.142437076
A23	0.07595488	0.0408306	−2.022317118	**A48**	0.24168643	0.06798972	5.939341384
A24	0.11633759	0.0504929	−0.234732043	**A49**	0.11025631	0.06177303	−0.471965842
A25	0.13476546	0.04297967	0.595714777	**A50**	0.05432123	0.01683487	−3.423666997

Table 10
Calculation of Assessment Score of alternatives.

Alternative	Assessment Score	Alternative	Assessment Score
A1	−2.667248699	**A26**	−0.787359214
A2	3.403739112	**A27**	−2.09156601
A3	2.251138712	**A28**	4.651165884
A4	7.370410417	**A29**	0.595714777
A5	7.118423217	**A30**	−1.211881859
A6	−2.09156601	**A31**	4.125198092
A7	−1.694974769	**A32**	1.256467817
A8	−2.022317118	**A33**	1.69644532
A9	−1.612407348	**A34**	−2.09156601
A10	−3.423666997	**A35**	2.48428733
A11	2.251138712	**A36**	0.595714777
A12	2.72193293	**A37**	−2.335972143
A13	−4.759794397	**A38**	−0.516958476
A14	−2.08920799	**A39**	−3.24372205
A15	−2.494574988	**A40**	−2.894265255
A16	−2.494574988	**A41**	−2.894265255
A17	2.409338642	**A42**	3.043849142
A18	−3.423666997	**A43**	−2.894265255
A19	−2.335972143	**A44**	−2.894265255
A20	−0.725735642	**A45**	2.014903318
A21	6.448211604	**A46**	−2.09156601
A22	3.105343987	**A47**	−0.142437076
A23	−2.022317118	**A48**	5.939341384
A24	−0.234732043	**A49**	−0.471965842
A25	0.595714777	**A50**	−3.423666997

Table 11
Validation Results.

Alternatives ID	Groups	Mean ± STD
A4, A5, A21, A48, A28, A31, A2, A22, A42, A12	G1	2.4286 ± 1.0081
A35, A17, A3, A11, A45, A33, A32, A25, A29, A36	G2	2.0000 ± 0.7361
A47, A24, A49, A38, A20, A26, A30, A9, A7, A8	G3	1.7143 ± 0.6860
A23, A14, A6, A27, A34, A46, A19, A37, A15, A16	G4	1.6286 ± 0.9454
A1, A40A41, A43, A44, A39, A10, A18, A50, A13	G5	1.5143 ± 0.3315

the robustness and stability of the decision-making process. Based on the q-ROF2TL-FWZIC weights for the evaluation attributes of trustworthy AI (C1 to C7), and to assess the impact of variations in these weights on the overall rankings of the alternatives. Amongst the ways on which sensitivity analysis is conducted, an example can be done by decreasing weight by 50 %. This technique has been used in sensitivity analysis of decision-making processes, particularly in MCDM. It involves decreasing the weight by 50 % assigned to each evaluation attribute and distributing equally the remaining weight to the other evaluation attributes and observing the resulting changes in the q-ROF2TL-CODAS decision outcomes. In that regard, decision-makers can assess the sensitivity of the decision to changes in the importance of those evaluation attributes. This provides insights into the relative importance of evaluation attributes and allows decision-makers to make informed decisions based on different weight scenarios. By performing weight decreasing and analyzing the resulting in q-ROF2TL-CODAS rank, decision-makers can gain a better understanding of the trade-offs between different evaluation attributes and their impact on the final decision of the trustworthy AI applications. This information can assist in identifying critical evaluation attributes, exploring potential biases or uncertainties in the decision process, and enhancing the overall robustness of the decision-making process. Table 12 reports the seven scenarios of new weights for evaluation attributes generated based on the philosophy of sensitivity analysis.

The extracted weights in Table 12 are employed to assess their sensitivity and their impact on the benchmarking of trustworthy AI applications. After applying these weights to the q-ROF2TL-CODAS method, Table 13 illustrates the variations in ranks, comparing the original q-ROF2TL-CODAS rank with the sensitivity analysis-based ranks, as visualized in Fig. 2.

Fig. 2 shows ranks variation by different sensitivity analysis weights with the q-ROF2TL-CODAS method. By examining how the rankings change across different scenarios, decision-makers can better understand the sensitivity of the outcomes to variations in weights and assumptions. Based on the provided Table 13 and Fig. 2, it appears that the ranking results are relatively stable for certain trustworthy AI applications, while others show some variability across scenarios. Some discussion aspects are apparent, and following is a discussion of their results as follows.

- **Proposed Methodology**: The rankings obtained from the proposed methodology represent the baseline scenario where the original weights are used. It serves as a reference point for comparing the rankings from the sensitivity analysis scenarios.
- **1st Scenario:** In this scenario, the rankings of some alternative's changes compared to the proposed framework. Alternative A10 moved from the 47th rank to the 46th rank, indicating an increase in its importance, while alternatives A14 moved from the 32nd rank to the 49th rank, indicating a decrease in its importance. Alternatives A15, A16, and A18 get higher significance in this scenario, where shifted from 39, 40, and 48 to 38, 39, and 47 respectively.
- **2nd Scenario:** In this scenario, most of the alternatives have change their order compared to the proposed framework. Alternative A21 was 3rd rank in proposed framework be 1st rank in this scenario, that means its significance be increased, while the significance of the alternatives; A2, A3, A4, A5, A8 are decreased in this scenario compared to the proposed framework, where were in orders 7th, 13th, 1st, 2nd, and 30th respectively in the proposed framework and be 11th, 14th, 2nd, 3rd, and 45th in this scenario.
- **3rd Scenario**: In this scenario, 20 % of the alternatives (*A1, A4, A8, A13, A15, A16, A20, A21, A22, A23, A35*) have same orders in rank compared to the proposed framework, remining alternatives get different orders in rank compared to the proposed framework.
- **4th Scenario**: This scenario has high variations from the proposed framework, where only 8 % of the alternatives (*A4, A20, A21, A33*) got same orders compared to the proposed framework, while remining alternatives get different orders.
- **5th Scenario:** In this scenario, 26 % of the alternatives (*A6, A13, A20, A25, A27, A28, A29, A30*) got same orders in rank compared to the proposed framework, while remining alternatives get different orders.

Table 12
Generated Weights.

Original weights	C1	C2	C3	C4	C5	C6	C7
	0.105741901	**0.136906613**	**0.16933294**	**0.173566825**	**0.152337199**	**0.107250068**	**0.154864455**
1st Scenario	0.05287095	0.145718438	0.178144765	0.18237865	0.161149024	0.116061893	0.16367628
2st Scenario	0.117150785	0.068453306	0.180741824	0.18497571	0.163746083	0.118658953	0.166273339
3st Scenario	0.119852979	0.151017691	0.08466647	0.187677904	0.166448277	0.121361147	0.168975533
4st Scenario	0.120205803	0.151370515	0.183796842	0.086783413	0.166801101	0.12171397	0.169328357
5st Scenario	0.118436667	0.149601379	0.182027706	0.186261592	0.076168599	0.119944835	0.167559221
6st Scenario	0.114679406	0.145844118	0.178270445	0.182504331	0.161274705	0.053625034	0.16380196
7st Scenario	0.118647272	0.149811984	0.182238311	0.186472196	0.16524257	0.12015544	0.077432227

Table 13
Ranks by Different Sensitivity Weights Scenarios.

Alternatives	Trustworthy AI applications	Ranking Scenarios							
		Proposed framework	**1st Scenario**	**2nd Scenario**	**3rd Scenario**	**4rd Scenario**	**5rd Scenario**	**6rd Scenario**	**7rd Scenario**
A1	App (Ho & Caals, 2021)	41	40	39	41	29	48	39	41
A2	App (El-Sappagh et al., 2021)	7	7	11	5	14	6	7	5
A3	App (Rostami & Oussalah, 2022)	13	13	14	9	17	10	13	13
A4	App (Rahman et al., 2021)	1	1	2	1	1	3	2	2
A5	App (Rahman et al., 2021)	2	2	3	2	4	1	1	1
A6	App (Müller et al., 2020)	33	32	28	32	47	33	33	33
A7	App (Lucieri et al., 2020)	29	29	27	28	46	28	29	29
A8	App (Karim et al., 2022)	30	30	45	30	21	31	31	31
A 9	App (Larasati et al., 2021)	28	28	24	27	44	30	28	28
A10	App (Rathi et al., 2021)	47	46	47	43	36	41	47	47
A11	App (Du et al., 2022)	14	14	15	10	18	11	14	14
A12	App (Zerka et al., 2020)	10	10	9	16	8	13	10	10
A13	App (Wang et al., 2022)	50	50	50	50	45	50	50	50
A14	App (Delacroix & Wagner, 2021)	32	49	35	29	19	29	30	30
A15	App (Chou et al., 2022)	39	38	32	39	40	46	37	39
A16	App (Shi et al., 2019)	40	39	33	40	41	47	38	40
A17	App (Abou-Nassar et al., 2020)	12	12	13	17	10	9	12	12
A18	App (Hussein et al., 2012)	48	47	48	44	37	42	48	48
A19	App (Guiñazú et al., 2020)	37	36	37	36	23	44	44	37
A20	App (Rehman et al., 2021)	25	24	25	38	25	25	25	25
A21	App (Collins et al., 2022)	3	3	1	4	3	2	3	3
A22	App (Kerasidou, 2021)	8	8	7	8	15	17	8	7
A23	App (Nicora et al., 2022)	31	31	46	31	22	32	32	32
A24	App (Zarour et al., 2020)	22	22	36	24	12	21	22	22
A25	App (Sachan et al., 2021)	18	18	20	12	26	18	18	18
A26	App (Lucieri et al., 2022)	26	25	26	21	39	26	26	26
A27	App (H. Müller et al., 2022)	34	33	29	33	48	34	34	34
A28	App (Arrieta et al., 2020)	5	5	6	7	6	5	5	6
A29	App (Deperlioglu et al., 2022)	19	19	21	13	27	19	19	19
A30	App (Setchi et al., 2020)	27	27	23	25	42	27	27	27
A31	App (Angerschmid et al., 2022)	6	6	4	3	9	15	6	9
A32	App (Sheikh et al., 2021)	17	17	19	23	7	14	17	17
A33	App (Barclay & Abramson, 2021)	16	16	10	22	16	12	16	16
A34	App (Pal, 2020)	35	34	30	34	49	35	35	35
A35	App (Holzinger et al., 2022)	11	11	12	11	13	8	11	11
A36	App (Martínez-Agüero et al., 2022)	20	20	22	14	28	20	20	20
A37	App (Saheb et al., 2021)	38	37	38	37	24	45	45	38
A38	App (Balagurunathan et al., 2021)	24	26	34	18	20	23	23	23
A39	App (Bania & Halder, 2021)	46	45	44	42	34	49	46	46
A40	App (Saba et al., 2020)	42	41	40	46	30	37	40	42
A41	App (Washington et al., 2020)	43	42	41	47	31	38	41	43
A42	App (Leal et al., 2021)	9	9	8	15	11	7	9	8
A43	App (Rieke et al., 2020)	44	43	42	48	32	39	42	44
A44	App (Wenzel & Wiegand, 2020)	45	44	43	49	33	40	43	45
A45	App (Oprescu et al., 2022)	15	15	16	20	5	16	15	15
A46	App (Holzinger et al., 2021)	36	35	31	35	50	36	36	36
A47	App (Séroussi et al., 2020)	21	21	17	26	35	22	21	21
A48	App (González-Gonzalo et al., 2022)	4	4	5	6	2	4	4	4
A49	App (Yang et al., 2022)	23	23	18	19	43	24	24	24
A50	App (Abdar et al., 2023)	49	48	49	45	38	43	49	49

- **6th and 7th Scenarios**: the ranks getting in scenarios of 6th and 7th have high similarity compared to the ranks getting from proposed framework, where 6th scenario was 69 of alternatives got same orders in rank compared to the proposed framework, and in 7th scenario were 72 of alternatives got same orders in rank compared to the proposed framework.

Overall, the sensitivity analysis demonstrates that the rankings of the trustworthy AI applications are influenced by the variation in weights assigned to the evaluation attributes. Different scenarios result in changes to the relative importance and positions of the frameworks. These sensitivity results emphasize the need for careful consideration of weights and the potential impact of different scenarios on the evaluation outcomes. They highlight the importance of examining the robustness of the rankings and the significance of individual evaluation attributes in the decision-making process. Researchers can use these sensitivity

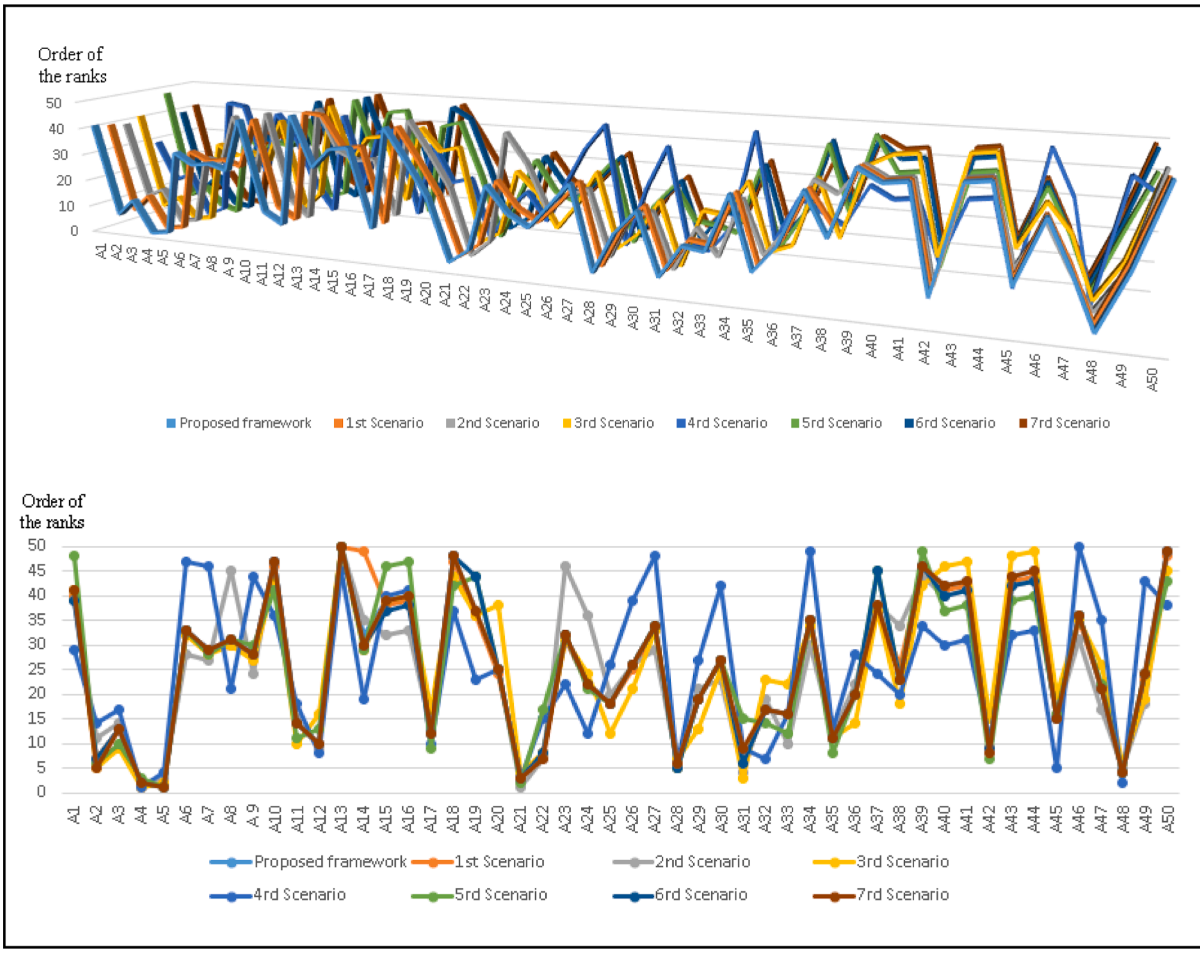

Fig. 2. Ranks Variation by Different Sensitivity Analysis Weights Scenarios with the q-ROF2TL-CODAS method.

results to gain insights into the stability and reliability of the rankings and make informed decisions based on the variations observed. It provides a comprehensive understanding of the trustworthy AI applications' performance under different weight distributions, enhancing the credibility and validity of the benchmarking results.

5. Research implications

This research presented several beneficial new knowledge and insights for society, healthcare providers, and practitioners, the study confirmed the necessity of handling ethical and societal issues related to AI, this research also proposed a benchmarking framework that can enable users' determination if the AI Applications compatible with the ethical and values standards of society. The proposed hybrid framework offers a novel evaluation and benchmarking approach of trustworthy AI systems in healthcare. The amalgamation of MCDM methods with the fuzzy environment facilitates an all-encompassing and dependable evaluation of AI systems according to their trustworthy attributes. The findings of the study offer insightful insights into the relative weight of trustworthiness attributes and the ranking of various AI healthcare systems. Adopting the proposed framework in evaluating healthcare-oriented AI applications based on fairness and accountability standards contributes to helping to obtain fair applications without bias or discrimination against any party. The proposed framework enables healthcare institutions to obtain AI applications that meet the requirements of trustworthiness by comparing available AI-based Healthcare applications. Thus, it allows healthcare institutions to make the right decisions regarding their healthcare applications. The proposed approach considered security, privacy, and compliance issues in evaluating AI applications, which makes them comply with applicable legal and regulatory requirements. To sum up, this study's implications generally focus on fostering several considerations such as transparency, confidence, ethical considerations, accountability, and responsible AI application deployment in healthcare. All that can make the users feel trust about integrating AI with healthcare applications and increase their acceptance of them, as well as enhance societal acceptance.

6. Conclusion

It is crucial to evaluate and benchmark AI-based healthcare applications in terms of trustworthiness. Confidence and satisfaction with AI healthcare applications are essential for delivering reliable services. The evaluation and benchmarking of the trustworthiness of these systems pose a significant challenge due to their complexity, involving qualitative concepts and multiple evaluation attributes that must be considered simultaneously. This study aimed to develop a new MCDM framework based on the q-ROF2TL environment to evaluate and benchmark trustworthy AI applications in healthcare.

To achieve this goal, the FWZIC method has been extended to a new version designed to operate within the q-ROF2TL environment, named q-ROF2TL-FWZIC. This method has been used for weighting the evaluation attributes and integrated with q-ROF2TL-CODAS for benchmarking the AI applications. The FWZIC method has been extended in this paper with q-ROF2TL, incorporating two types of distance measurements, allowing for a comprehensive consideration of both ED and TD. Q-ROF2TL combines 2TL terms and q-ROF sets, enhancing the adaptability of q-ROFS.

The combination of q-ROF2TL-FWZIC and q-ROF2TL-CODAS techniques, along with sensitivity analysis, contributes to the robustness of the proposed MCDM mathematical model. This framework is intended to support the development of trustworthy healthcare AI applications and facilitate the transition toward trustworthy AI. By adopting evaluation attributes that encompass all components of AI trustworthiness, the evaluation activities become more comprehensive, supporting precise decision-making regarding the trustworthiness of AI applications.

The implications of this study center around promoting transparency, confidence, ethical considerations, accountability, and responsible AI application deployment in healthcare. These considerations aim to enhance user confidence in the use of artificial intelligence in healthcare and foster societal acceptance of such applications.

Despite the contributions of this study, there are limitations. One limitation is the failure to consider the relative contribution or weight importance of each expert's expertise or viewpoints. This oversight may potentially impact the final weighting of criteria and the selection of alternatives in the case study under investigation, as well as in prospective future cases. A forthcoming endeavor involves addressing this issue by proposing a novel technique for assigning distinct weights to each expert, which can be used in determining criteria weights and subsequent alternative evaluations.

In conclusion, the integration of additional fuzzy sets and precise fuzzy operators with the suggested MCDM techniques holds promise as a potential avenue for future research contributions.

CRediT authorship contribution statement

M.A. Alsalem: Writing – original draft. **A.H. Alamoodi:** Writing – original draft. **O.S. Albahri:** Conceptualization, Methodology. **A.S. Albahri:** Conceptualization, Methodology. **Luis Martínez:** Methodology, Writing – review & editing. **R. Yera:** Writing – review & editing. **Ali M. Duhaim:** Formal analysis. **Iman Mohamad Sharaf:** Formal analysis.

Declaration of competing interest

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

Data availability

Data will be made available on request.

References

Abbas, F., Ali, J., Mashwani, W. K., & Syam, M. I. (2023). q-rung orthopair fuzzy 2-tuple linguistic clustering algorithm and its applications to clustering analysis. *Scientific Reports, 13*(1), 2789.

Abdar, M., Salari, S., Qahremani, S., Lam, H.-K., Karray, F., Hussain, S., … Nahavandi, S. (2023). UncertaintyFuseNet: Robust uncertainty-aware hierarchical feature fusion model with ensemble Monte Carlo dropout for COVID-19 detection. *Information Fusion, 90*, 364–381.

Abou-Nassar, E. M., Iliyasu, A. M., El-Kafrawy, P. M., Song, O.-Y., Bashir, A. K., & Abd El-Latif, A. A. (2020). DITrust chain: Towards blockchain-based trust models for sustainable healthcare IoT systems. *IEEE Access, 8*, 111223–111238.

Adedjouma, M., Adam, J.-L., Aknin, P., Alix, C., Baril, X., Bernard, G., Bonhomme, Y., Braunschweig, B., Cantat, L., & Chale-Gongora, G. (2022). *Towards the engineering of trustworthy AI applications for critical systems-The Confiance. ai program*.

Administration, U. S. F. and D. (2022). *Artificial intelligence and machine learning (AI/ML)-enabled medical devices*.

Al-Samarraay, M., Al-Zuhairi, O., Alamoodi, A. H., Albahri, O. S., Deveci, M., Alobaidi, O. R., … Kou, G. (2024). An integrated fuzzy multi-measurement decision-making model for selecting optimization techniques of semiconductor materials. *Expert Systems with Applications, 237*, Article 121439.

Alamoodi, A. H., Garfan, S., Zaidan, B. B., Zaidan, A. A., Shuwandy, M. L., Alaa, M., … Albahri, O. S. (2020). A systematic review into the assessment of medical apps: Motivations, challenges, recommendations and methodological aspect. *Health and Technology, 10*, 1045–1061.

Alamoodi, A. H., Mohammed, R. T., Albahri, O. S., Qahtan, S., Zaidan, A. A., Alsattar, H. A., … Baqer, M. J. (2022). Based on neutrosophic fuzzy environment: A new development of FWZIC and FDOSM for benchmarking smart e-tourism applications. *Complex & Intelligent Systems, 8*(4), 3479–3503.

Albahri, A. S., Duhaim, A. M., Fadhel, M. A., Alnoor, A., Baqer, N. S., Alzubaidi, L., … Salhi, A. (2023). A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. *Information Fusion*.

Albahri, O. S., Alamoodi, A. H., Deveci, M., Albahri, A. S., Mahmoud, M. A., Al-Quraishi, T., … Sharaf, I. M. (2023). Evaluation of organizational culture in companies for fostering a digital innovation using q-rung picture fuzzy based decision-making model. *Advanced Engineering Informatics, 58*, Article 102191.

Albahri, O. S., Alamoodi, A. H., Deveci, M., Albahri, A. S., Mahmoud, M. A., Sharaf, I. M., & Coffman, D. (2023). Multi-perspective evaluation of integrated active cooling systems using fuzzy decision making model. *Energy Policy, 182*, Article 113775.

Albahri, O. S., Zaidan, A. A., Albahri, A. S., Alsattar, H. A., Mohammed, R., Aickelin, U., … Alamoodi, A. H. (2022). Novel dynamic fuzzy decision-making framework for COVID-19 vaccine dose recipients. *Journal of Advanced Research, 37*, 147–168.

Albahri, O. S., Zaidan, A. A., Salih, M. M., Zaidan, B. B., Khatari, M. A., Ahmed, M. A., … Alazab, M. (2021). Multidimensional benchmarking of the active queue management methods of network congestion control based on extension of fuzzy decision by opinion score method. *International Journal of Intelligent Systems, 36*(2), 796–831. https://doi.org/10.1002/int.22322

Albahri, O. S., Zaidan, A. A., Zaidan, B. B., Albahri, A. S., Mohsin, A. H., Mohammed, K. I., & Alsalem, M. A. (2022). New mHealth hospital selection framework supporting decentralised telemedicine architecture for outpatient cardiovascular disease-based integrated techniques: Haversine-GPS and AHP-VIKOR. *Journal of Ambient Intelligence and Humanized Computing*, 1–21.

Albrecht, U.-V., Hillebrand, U., & von Jan, U. (2018). Relevance of trust marks and CE labels in German-language store descriptions of health apps: Analysis. *JMIR MHealth and UHealth, 6*(4), e10394.

Ali, A., Ullah, K., & Hussain, A. (2023). An approach to multi-attribute decision-making based on intuitionistic fuzzy soft information and Aczel-Alsina operational laws. *Journal of Decision Analytics and Intelligent Computing, 3*(1), 80–89.

Almahdi, E. M., Zaidan, A. A., Zaidan, B. B., Alsalem, M. A., Albahri, O. S., & Albahri, A. S. (2019). Mobile-based patient monitoring systems: A prioritisation framework using multi-criteria decision-making techniques. *Journal of Medical Systems, 43*(7). https://doi.org/10.1007/s10916-019-1339-9

Alsalem, M. A., Alamoodi, A. H., Albahri, O. S., Dawood, K. A., Mohammed, R. T., Alnoor, A., … Jumaah, F. M. (2022). Multi-criteria decision-making for coronavirus disease 2019 applications: A theoretical analysis review. *Artificial Intelligence Review, 55*(6), 4979–5062.

Alsalem, M. A., Alsattar, H. A., Albahri, A. S., Mohammed, R. T., Albahri, O. S., Zaidan, A. A., … Zaidan, B. B. (2021). Based on T-spherical fuzzy environment: A combination of FWZIC and FDOSM for prioritising COVID-19 vaccine dose recipients. *Journal of Infection and Public Health, 14*(10), 1513–1559.

Alsalem, M. A., Mohammed, R., Albahri, O. S., Zaidan, A. A., Alamoodi, A. H., Dawood, K., … Aickelin, U. (2022). Rise of multiattribute decision-making in combating COVID-19: A systematic review of the state-of-the-art literature. *International Journal of Intelligent Systems, 37*(6), 3514–3624.

Alsalem, M. A., Zaidan, A. A., Zaidan, B. B., Albahri, O. S., Alamoodi, A. H., Albahri, A. S., … Mohammed, K. I. (2019). Multiclass benchmarking framework for automated acute leukaemia detection and classification based on BWM and group-VIKOR. *Journal of Medical Systems, 43*(7). https://doi.org/10.1007/s10916-019-1338-x

Angerschmid, A., Zhou, J., Theuermann, K., Chen, F., & Holzinger, A. (2022). Fairness and explanation in AI-informed decision making. *Machine Learning and Knowledge Extraction, 4*(2), 556–579.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., … Benjamins, R. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115.

Arshad, M., Saeed, M., Rahman, A. U., Mohammed, M. A., Abdulkareem, K. H., Nedoma, J., … Deveci, M. (2024). A robust framework for the selection of optimal COVID-19 mask based on aggregations of interval-valued multi-fuzzy hypersoft sets. *Expert Systems with Applications, 238*, Article 121944.

Balagurunathan, Y., Mitchell, R., & El Naqa, I. (2021). Requirements and reliability of AI in the medical context. *Physica Medica, 83*, 72–78.

Bania, R. K., & Halder, A. (2021). R-HEFS: Rough set based heterogeneous ensemble feature selection method for medical data classification. *Artificial Intelligence in Medicine, 114*, Article 102049.

Barclay, I., & Abramson, W. (2021). Identifying roles, requirements and responsibilities in trustworthy AI systems. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers* (pp. 264–271).

Barometer, E. T. (2019). Edelman trust barometer global report. *Retrieved November, 2* (2021), 2002–2019.

Cheney, S., & Thompson, E. (2018). The 2017–2022 app economy forecast: 6 billion devices, $157 billion in spend & more. *App Annie, 2*.

Chou, Y.-L., Moreira, C., Bruza, P., Ouyang, C., & Jorge, J. (2022). Counterfactuals and causability in explainable artificial intelligence: Theory, algorithms, and applications. *Information Fusion, 81*, 59–83.

Collins, J. W., Marcus, H. J., Ghazi, A., Sridhar, A., Hashimoto, D., Hager, G., … Marz, K. (2022). Ethical implications of AI in robotic surgical training: A Delphi consensus statement. *European Urology Focus, 8*(2), 613–622.

Darras, K. E., van Merriënboer, J. J. G., Toom, M., Roberson, N. D., de Bruin, A. B. H., Nicolaou, S., & Forster, B. B. (2019). Developing the evidence base for M-learning in

undergraduate radiology education: Identifying learner preferences for mobile apps. *Canadian Association of Radiologists' Journal, 70*(3), 320–326.
Delacroix, S., & Wagner, B. (2021). Constructing a mutually supportive interface between ethics and regulation. *Computer Law & Security Review, 40*, Article 105520.
Deperlioglu, O., Kose, U., Gupta, D., Khanna, A., Giampaolo, F., & Fortino, G. (2022). Explainable framework for Glaucoma diagnosis by image processing and convolutional neural network synergy: Analysis with doctor evaluation. *Future Generation Computer Systems, 129*, 152–169.
Du, Y., Rafferty, A. R., McAuliffe, F. M., Wei, L., & Mooney, C. (2022). An explainable machine learning-based clinical decision support system for prediction of gestational diabetes mellitus. *Scientific Reports, 12*(1), 1170.
El-Sappagh, S., Alonso, J. M., Islam, S. M. R., Sultan, A. M., & Kwak, K. S. (2021). A multilayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer's disease. *Scientific Reports, 11*(1), 2660.
Enaizan, O., Zaidan, A. A., Alwi, N. H. M., Zaidan, B. B., Alsalem, M. A., Albahri, O. S., & Albahri, A. S. (2020). Electronic medical record systems: Decision support examination framework for individual, security and privacy concerns using multi-perspective analysis. *Health and Technology, 10*, 795–822.
Fan, W., Liu, J., Zhu, S., & Pardalos, P. M. (2020). Investigating the impacting factors for the healthcare professionals to adopt artificial intelligence-based medical diagnosis support system (AIMDSS). *Annals of Operations Research, 294*, 567–592.
Freier, A. (2018). App revenue reaches $92.1 billion in 2018 driven by mobile gaming apps. *Business Apps, 13*(09).
Ghoushchi, S. J., & Sarvi, S. (2023). Prioritizing and evaluating risks of ordering and prescribing in the chemotherapy process using an extended SWARA and MOORA under fuzzy Z-numbers. *Journal of Operations Intelligence, 1*(1), 44–66.
Gille, F., Jobin, A., & Ienca, M. (2020). What we talk about when we talk about trust: Theory of trust for AI in healthcare. *Intelligence-Based Medicine, 1*, Article 100001.
González-Gonzalo, C., Thee, E. F., Klaver, C. C. W., Lee, A. Y., Schlingemann, R. O., Tufail, A., … Sánchez, C. I. (2022). Trustworthy AI: Closing the gap between development and integration of AI systems in ophthalmic practice. *Progress in Retinal and Eye Research, 90*, Article 101034.
Guiñazú, M. F., Cortés, V., Ibáñez, C. F., & Velásquez, J. D. (2020). Employing online social networks in precision-medicine approach using information fusion predictive model to improve substance use surveillance: A lesson from Twitter and marijuana consumption. *Information Fusion, 55*, 150–163.
Hasani, N., Morris, M. A., Rahmim, A., Summers, R. M., Jones, E., Siegel, E., & Saboury, B. (2022). Trustworthy artificial intelligence in medical imaging. *PET Clinics, 17*(1), 1–12.
Herrera, F., & Martínez, L. (2000). A 2-tuple fuzzy linguistic representation model for computing with words. *IEEE Transactions on Fuzzy Systems, 8*(6), 746–752.
Ho, C.-W.-L., & Caals, K. (2021). A call for an ethics and governance action plan to harness the power of artificial intelligence and digitalization in nephrology. *Seminars in Nephrology, 41*(3), 282–293.
Holl, K., & Elberzhager, F. (2019). Mobile application quality assurance. *Advances in computers* (Vol. 112,, 1–77.
Holzinger, A., Dehmer, M., Emmert-Streib, F., Cucchiara, R., Augenstein, I., Del Ser, J., … Díaz-Rodríguez, N. (2022). Information fusion as an integrative cross-cutting enabler to achieve robust, explainable, and trustworthy medical artificial intelligence. *Information Fusion, 79*, 263–278.
Holzinger, A., Malle, B., Saranti, A., & Pfeifer, B. (2021). Towards multi-modal causability with graph neural networks enabling information fusion for explainable AI. *Information Fusion, 71*, 28–37.
Hsu, C.-Y., Huang, B.-R., Wang, M.-Y., & Yun-Jie, F. (2023). Exploring competitive sports technology development: Using a MCDM model. *Journal of Physical Education & Sport, 23*(8).
Hussein, A. S., Omar, W. M., Li, X., & Ati, M. (2012). Efficient chronic disease diagnosis prediction and recommendation system. *IEEE-EMBS Conference on Biomedical Engineering and Sciences, 2012*, 209–214.
Kale, A., Nguyen, T., Harris, F. C., Jr, Li, C., Zhang, J., & Ma, X. (2022). Provenance documentation to enable explainable and trustworthy AI: A literature review. *Data Intelligence*, 1–41.
Karim, M. R., Islam, T., Lange, C., Rebholz-Schuhmann, D., & Decker, S. (2022). Adversary-aware multimodal neural networks for cancer susceptibility prediction from multiomics data. *IEEE Access, 10*, 54386–54409.
Kaur, D., Uslu, S., & Durresi, A. (2021). Requirements for trustworthy artificial intelligence–a review. *Advances in Networked-Based Information Systems: The 23rd International Conference on Network-Based Information Systems (NBiS-2020) 23*, 105–115.
Kaur, D., Uslu, S., Rittichier, K. J., & Durresi, A. (2022). Trustworthy artificial intelligence: A review. *ACM Computing Surveys (CSUR), 55*(2), 1–38.
Kerasidou, A. (2021). Ethics of artificial intelligence in global health: Explainability, algorithmic bias and trust. *Journal of Oral Biology and Craniofacial Research, 11*(4), 612–614.
Keshavarz Ghorabaee, M., Zavadskas, E. K., Turskis, Z., & Antucheviciene, J. (2016). A new combinative distance-based assessment (CODAS) method for multi-criteria decision-making. *Economic Computation & Economic Cybernetics Studies & Research, 50*(3).
Khatari, M., Zaidan, A. A., Zaidan, B. B., Albahri, O. S., & Alsalem, M. A. (2019). Multi-criteria evaluation and benchmarking for active queue management methods: Open issues, challenges and recommended pathway solutions. *International Journal of Information Technology and Decision Making, 18*(4). https://doi.org/10.1142/S0219622019300039
Lakhan, A., Mohammed, M. A., Abdulkareem, K. H., Hamouda, H., & Alyahya, S. (2023). Autism Spectrum Disorder detection framework for children based on federated learning integrated CNN-LSTM. *Computers in Biology and Medicine, 166*, Article 107539.
Larasati, R., De Liddo, A., & Motta, E. (2021). Ai healthcare system interface: Explanation design for non-expert user trust. *ACMIUI-WS 2021: Joint Proceedings of the ACM IUI*.
Leal, F., Chis, A. E., Caton, S., González-Vélez, H., García-Gómez, J. M., Durá, M., … Gerogiannis, V. C. (2021). Smart pharmaceutical manufacturing: Ensuring end-to-end traceability and data integrity in medicine production. *Big Data Research, 24*, Article 100172.
Leung, C. K., Madill, E. W. R., Souza, J., & Zhang, C. Y. (2022). Towards trustworthy artificial intelligence in healthcare. In *2022 IEEE 10th International Conference on Healthcare Informatics (ICHI)* (pp. 626–632).
Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., … Zhou, B. (2023). Trustworthy ai: From principles to practices. *ACM Computing Surveys, 55*(9), 1–46.
Li, Z., Pan, Q., Wang, D., & Liu, P. (2022). An extended PROMETHEE II method for multi-attribute group decision-making under q-rung orthopair 2-tuple linguistic environment. *International Journal of Fuzzy Systems, 24*(7), 3039–3056.
Liang, W., Tadesse, G. A., Ho, D., Fei-Fei, L., Zaharia, M., Zhang, C., & Zou, J. (2022). Advances, challenges and opportunities in creating data for trustworthy AI. *Nature Machine Intelligence, 4*(8), 669–677.
Limboo, B., & Dutta, P. (2022). A q-rung orthopair basic probability assignment and its application in medical diagnosis. *Decision Making: Applications in Management and Engineering, 5*(1), 290–308.
Lucieri, A., Bajwa, M. N., Braun, S. A., Malik, M. I., Dengel, A., & Ahmed, S. (2022). ExAID: A multimodal explanation framework for computer-aided diagnosis of skin lesions. *Computer Methods and Programs in Biomedicine, 215*, Article 106620.
Lucieri, A., Bajwa, M. N., Dengel, A., & Ahmed, S. (2020). Achievements and challenges in explaining deep learning based computer-aided diagnosis systems. *ArXiv Preprint ArXiv:2011.13169*.
Markus, A. F., Kors, J. A., & Rijnbeek, P. R. (2021). The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. *Journal of Biomedical Informatics, 113*, Article 103655.
Martínez-Agüero, S., Soguero-Ruiz, C., Alonso-Moral, J. M., Mora-Jiménez, I., Álvarez-Rodríguez, J., & Marques, A. G. (2022). Interpretable clinical time-series modeling with intelligent feature selection for early prediction of antimicrobial multidrug resistance. *Future Generation Computer Systems, 133*, 68–83.
Matell, M. S., & Jacoby, J. (1971). Is there an optimal number of alternatives for Likert scale items? Study I: Reliability and validity. *Educational and Psychological Measurement, 31*(3), 657–674.
Mattioli, J., Sohier, H., Delaborde, A., Pedroza, G., Amokrane-Ferka, K., Awadid, A., … Khalfaoui, S. (2023). Towards a holistic approach for AI trustworthiness assessment based upon aids for multi-criteria aggregation. In *SafeAI 2023-The AAAI's Workshop on Artificial Intelligence Safety* (p. 3381).
Moslem, S., Solieman, H., Oubahman, L., Duleba, S., Senapati, T., & Pilla, F. (2023). Assessing public transport supply quality: A comparative analysis of analytical network process and analytical hierarchy process. *Journal of Soft Computing and Decision Analytics, 1*(1), 124–138.
Müller, H., Holzinger, A., Plass, M., Brcic, L., Stumptner, C., & Zatloukal, K. (2022). Explainability and causability for artificial intelligence-supported medical image analysis in the context of the European In Vitro Diagnostic Regulation. *New Biotechnology, 70*, 67–72.
Müller, J., Stoehr, M., Oeser, A., Gaebel, J., Streit, M., Dietz, A., & Oeltze-Jafra, S. (2020). A visual approach to explainable computerized clinical decision support. *Computers & Graphics, 91*, 1–11.
Naz, S., Akram, M., Sattar, A., & Al-Shamiri, M. M. A. (2022). 2-tuple linguistic q-rung orthopair fuzzy CODAS approach and its application in arc welding robot selection. *AIMS Math, 7*(9), 17529–17569.
Nezhad, M. Z., Nazarian-Jashnabadi, J., Rezazadeh, J., Mehraeen, M., & Bagheri, R. (2023). Assessing dimensions influencing IoT implementation readiness in industries: A fuzzy DEMATEL and fuzzy AHP analysis. *Journal of Soft Computing and Decision Analytics, 1*(1), 102–123.
Nicora, G., Rios, M., Abu-Hanna, A., & Bellazzi, R. (2022). Evaluating pointwise reliability of machine learning prediction. *Journal of Biomedical Informatics, 127*, Article 103996.
Oprescu, A. M., Miró-Amarante, G., García-Díaz, L., Rey, V. E., Chimenea-Toscano, A., Martínez-Martínez, R., & Romero-Ternero, M. C. (2022). Towards a data collection methodology for Responsible Artificial Intelligence in health: A prospective and qualitative study in pregnancy. *Information Fusion, 83*, 53–78.
Ouhbi, S., Fernández-Alemán, J. L., Pozo, J. R., Bajta, M. E., Toval, A., & Idri, A. (2015). Compliance of blood donation apps with mobile OS usability guidelines. *Journal of Medical Systems, 39*, 1–21.
Pal, N. R. (2020). In search of trustworthy and transparent intelligent systems with human-like cognitive and reasoning capabilities. *Frontiers in Robotics and AI, 7*, 76.
Procter, R., Tolmie, P., & Rouncefield, M. (2023). Holding AI to account: Challenges for the delivery of trustworthy AI in healthcare. *ACM Transactions on Computer-Human Interaction, 30*(2), 1–34.
Qader, M. A., Zaidan, B. B., Zaidan, A. A., Ali, S. K., Kamaluddin, M. A., & Radzi, W. B. (2017). A methodology for football players selection problem based on multi-measurements criteria analysis. *Measurement: Journal of the International Measurement Confederation, 111*, 38–50. https://doi.org/10.1016/j.measurement.2017.07.024
Rahman, A. U., Saeed, M., Mohammed, M. A., Abdulkareem, K. H., Nedoma, J., & Martinek, R. (2023). An innovative mathematical approach to the evaluation of susceptibility in liver disorder based on fuzzy parameterized complex fuzzy hypersoft set. *Biomedical Signal Processing and Control, 86*, Article 105204.

Rahman, A. U., Saeed, M., Saeed, M. H., Zebari, D. A., Albahar, M., Abdulkareem, K. H., … Mohammed, M. A. (2023). A framework for susceptibility analysis of brain tumours based on uncertain analytical cum algorithmic modeling. *Bioengineering, 10*(2), 147.
Rahman, M. A., Hossain, M. S., Showail, A. J., Alrajeh, N. A., & Alhamid, M. F. (2021). A secure, private, and explainable IoHT framework to support sustainable health monitoring in a smart city. *Sustainable Cities and Society, 72*, Article 103083.
Ranjan, M. J., Kumar, B. P., Bhavani, T. D., Padmavathi, A. V., & Bakka, V. (2023). Probabilistic linguistic q-rung orthopair fuzzy Archimedean aggregation operators for group decision-making. *Decision Making: Applications in Management and Engineering, 6*(2), 639–667.
Rathi, V. K., Rajput, N. K., Mishra, S., Grover, B. A., Tiwari, P., Jaiswal, A. K., & Hossain, M. S. (2021). An edge AI-enabled IoT healthcare monitoring system for smart cities. *Computers & Electrical Engineering, 96*, Article 107524.
Rehman, A., Saba, T., Haseeb, K., Larabi Marie-Sainte, S., & Lloret, J. (2021). Energy-efficient IoT e-health using artificial intelligence model with homomorphic secret sharing. *Energies, 14*(19), 6414.
Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., … Maier-Hein, K. (2020). The future of digital health with federated learning. *NPJ Digital Medicine, 3*(1), 119.
Rong, G., Mendez, A., Assi, E. B., Zhao, B., & Sawan, M. (2020). Artificial intelligence in healthcare: Review and prediction case studies. *Engineering, 6*(3), 291–301.
Rostami, M., & Oussalah, M. (2022). A novel explainable COVID-19 diagnosis method by integration of feature selection with random forest. *Informatics in Medicine Unlocked, 30*, Article 100941.
Saba, T., Haseeb, K., Ahmed, I., & Rehman, A. (2020). Secure and energy-efficient framework using Internet of Medical Things for e-healthcare. *Journal of Infection and Public Health, 13*(10), 1567–1575.
Sachan, S., Almaghrabi, F., Yang, J.-B., & Xu, D.-L. (2021). Evidential reasoning for preprocessing uncertain categorical data for trustworthy decisions: An application on healthcare and finance. *Expert Systems with Applications, 185*, Article 115597.
Saheb, T., Saheb, T., & Carpenter, D. O. (2021). Mapping research strands of ethics of artificial intelligence in healthcare: A bibliometric and content analysis. *Computers in Biology and Medicine, 135*, Article 104660.
Sahoo, S. K., & Goswami, S. S. (2023). A comprehensive review of multiple criteria decision-making (MCDM) Methods: Advancements, applications, and future directions. *Decision Making Advances, 1*(1), 25–48.
Saif, I., & Ammanath, B. (2020). Trustworthy AI'is a framework to help manage unique risk. *MIT Technology Review, Vol. Artificial Intelligence*.
Saleh, N., & Salaheldin, A. M. (2022). A benchmarking platform for selecting optimal retinal diseases diagnosis model based on a multi-criteria decision-making approach. *Journal of the Chinese Institute of Engineers, 45*(1), 27–34.
Saqlain, M., & Saeed, M. (2024). From ambiguity to clarity: Unraveling the Power of similarity measures in multi-polar interval-valued intuitionistic fuzzy soft sets. *Decision Making Advances*.
Séroussi, B., Hollis, K. F., & Soualmia, L. F. (2020). Transparency of health informatics processes as the condition of healthcare professionals' and patients' trust and adoption: The rise of ethical requirements. *Yearbook of Medical Informatics, 29*(01), 7–10.
Setchi, R., Dehkordi, M. B., & Khan, J. S. (2020). Explainable robotics in human-robot interactions. *Procedia Computer Science, 176*, 3057–3066.
Sharaf, I. M., Alamoodi, A. H., Albahri, O. S., Deveci, M., Talal, M., Albahri, A. S., … Pedrycz, W. (2024). Architecture selection for 5G-radio access network using type-2 neutrosophic numbers based decision making model. *Expert Systems with Applications, 237*, Article 121420.
Sheikh, A., Anderson, M., Albala, S., Casadei, B., Franklin, B. D., Richards, M., … Mossialos, E. (2021). Health information technology and digital innovation for national learning health and care systems. *The Lancet Digital Health, 3*(6), e383–e396.
Shi, Z., Chen, W., Liang, S., Zuo, W., Yue, L., & Wang, S. (2019). Deep interpretable mortality model for intensive care unit risk prediction. *Advanced Data Mining and Applications: 15th International Conference, ADMA 2019, Dalian, China, November 21–23, 2019, Proceedings 15*, 617–631.
Smuha, N. A. (2019). The EU approach to ethics guidelines for trustworthy artificial intelligence. *Computer Law Review International, 20*(4), 97–106.
Stanton, B., & Jensen, T. (2021). Trust and artificial intelligence. *Preprint*.
Tang, A. K. Y. (2019). A systematic literature review and analysis on mobile apps in m-commerce: Implications for future research. *Electronic Commerce Research and Applications, 37*, Article 100885.
Tešić, D., & Marinković, D. (2023). Application of fermatean fuzzy weight operators and MCDM model DIBR-DIBR II-NWBM-BM for efficiency-based selection of a complex combat system. *Journal of Decision Analytics and Intelligent Computing, 3*(1), 243–256.
Thiebes, S., Lins, S., & Sunyaev, A. (2021). Trustworthy artificial intelligence. *Electronic Markets, 31*, 447–464.
Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine, 25*(1), 44–56.
Tzeng, G.-H., & Huang, J.-J. (2011). *Multiple attribute decision making: Methods and applications*. CRC Press.
Vahdani, B., Tavakkoli-Moghaddam, R., Mousavi, S. M., & Ghodratnama, A. (2013). Soft computing based on new interval-valued fuzzy modified multi-criteria decision-making method. *Applied Soft Computing, 13*(1), 165–172.
Wang, J., Jin, H., Chen, J., Tan, J., & Zhong, K. (2022). Anomaly detection in Internet of medical Things with Blockchain from the perspective of deep neural network. *Information Sciences, 617*, 133–149.
Washington, P., Leblanc, E., Dunlap, K., Penev, Y., Varma, M., Jung, J.-Y., Chrisman, B., Sun, M. W., Stockham, N., & Paskov, K. M. (2020). Selection of trustworthy crowd workers for telemedical diagnosis of pediatric autism spectrum disorder. *Biocomputing 2021: Proceedings of the Pacific Symposium*, 14–25.
Wei, G., Gao, H., & Wei, Y. (2018). Some q-rung orthopair fuzzy Heronian mean operators in multiple attribute decision making. *International Journal of Intelligent Systems, 33*(7), 1426–1458.
Wenzel, M., & Wiegand, T. (2020). Toward global validation standards for health AI. *IEEE Communications Standards Magazine, 4*(3), 64–69.
Yager, R. R. (2016). Generalized orthopair fuzzy sets. *IEEE Transactions on Fuzzy Systems, 25*(5), 1222–1230.
Yang, G., Ye, Q., & Xia, J. (2022). Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond. *Information Fusion, 77*, 29–52.
Yu, K.-H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature Biomedical Engineering, 2*(10), 719–731.
Zaidan, A. A., Zaidan, B. B., Alsalem, M. A., Albahri, O. S., Albahri, A. S., & Qahtan, M. Y. (2020). Multi-agent learning neural network and Bayesian model for real-time IoT skin detectors: A new evaluation and benchmarking methodology. *Neural Computing and Applications, 32*, 8315–8366.
Zarour, M., Ansari, M. T. J., Alenezi, M., Sarkar, A. K., Faizan, M., Agrawal, A., … Khan, R. A. (2020). Evaluating the impact of blockchain models for secure and trustworthy electronic healthcare records. *IEEE Access, 8*, 157959–157973.
Zerka, F., Urovi, V., Vaidyanathan, A., Barakat, S., Leijenaar, R. T. H., Walsh, S., … Dumontier, M. (2020). Blockchain for privacy preserving and trustworthy distributed machine learning in multicentric medical imaging (C-DistriM). *IEEE Access, 8*, 183939–183951.
Amri, M. M., & Abed, S. A. (2023). The data-driven future of healthcare: A review. *Mesopotamian Journal of Big Data, 2023*, 70–76.
Obaid, O. I. (2023). From machine learning to artificial general intelligence: A roadmap and implications. *Mesopotamian Journal of Big Data, 2023*(SE-Articles), 81–91. https://doi.org/10.58496/mjbd/2023/012