scientific reports

OPEN

A hierarchical framework to evaluate the usability of smartphone health applications

Ansar Siddique[1]✉, Qudsia Sajjad[2], Shabbab Ali Algamdi[3]✉, Basit Shahzad[4] & Jamal M. Hussien[5]

In the current age of digitalization, smartphone health applications are rapidly increasing. It could be of great help for society to use health apps for regular monitoring of health instead of personally visiting medical consultants. Several health apps exist but their utilization is low mainly owing to usability problems. To enhance usage, it is essential to identify the key parameters that significantly affect the usability of health apps. This insight can help in developing an evaluation model for their usability. Although numerous usability models exist but these models often overlook critical usability aspects such as trust, security, response time and interruptibility and diverse stakeholder needs specific to health apps. This research emphasized identifying and prioritizing usability parameters of health apps using a hierarchical model. First, to determine the key parameters influencing the usability of health apps, a thorough analysis of relevant literature was carried out. Secondly, only the parameters with the most significant effects among the recognized ones were considered. A survey was conducted to identify the key usability parameters. The survey instrument was distributed among 195 subjects, including medical doctors, pharmacists, paramedical staff, medical students, designers of health apps, and the general public. Lastly, a second instrument was used with 49 participants to conduct pairwise comparisons among the parameters and sub-parameters, ranking them according to their relative importance. The identified and prioritized sub-parameters were grouped into four major categories. Among these, efficiency was recognized as the key criterion while effectiveness was considered the least significant criterion. The comparison revealed that the proposed model is more comprehensive regarding usability than existing models.

Smartphones are important inventions in technological development. It is a new inclusion to the bracket of the mobile phone family and one of the rapidly proliferating sections of the mobile market with a year-to-year increasing penetration rate[1]. According to Statista, there are 6.648 billion users of smartphones which points out that almost 84% of the world's population owns smartphones[2]. The statistical research department of Pakistan predicted nearly 51% usage during the year 2020, now which has increased further[1]. According to the statistics of the Pakistan Telecommunication Authority (PTA), there are 164 million cellular users, among which 77% of subscribers belong to youth[2]. Smartphones are technologically advanced and sophisticated devices (i.e. iPhone, Android, Windows phones, tablets, etc.,) that can perform all of the operations of a computer such as running operating system, downloading software applications, internet access using 4G & 5G networks, web browsing, navigation system, email services and perform multimedia functionality along with communication facility[2,3]. The landmark in the smartphone industry occurred with the launch of a business device, BlackBerry, which facilitated email, instant messaging, and HTML browsing[4]. Currently, smartphones are constantly introducing new functions and applications that attract users to explore new things and lead them towards the use of this emerging technology for various daily activities. Therefore, in the current digital era, smartphones have become an important part of our lives[3]. A mobile application is a software, specifically developed to run on mobile devices including smartphones and tablets. Smartphone usage has been expanded to several fields, especially healthcare,

[1]Department of Computer Sciences, Bahria University, Lahore Campus, Lahore, Punjab 54600, Pakistan. [2]Department of Information Technology, University of Gujrat, Gujrat 50700, Pakistan. [3]Department of Software Engineering, Prince Sattam bin Abdul Aziz University, Al-Kharj, Saudi Arabia. [4]Department of Software Engineering, University of Prince Mugrin, Medina, Saudi Arabia. [5]Information and Communication Technology (ICT) Department, University of Prince Mugrin, Medina, Saudi Arabia. ✉email: ansar.bulc@bahria.edu.pk; s.algamdi@psau.edu.sa

where they offer significant potential and practical advantages. Smartphone health applications (a.k.a. health apps) can be downloaded effortlessly onto mobile devices. Such applications are highly accepted among medical students and young clinicians. Research reports that nearly 50% medical students and professionals rely on health apps to access medical information. Young health professionals, and technologically capable patients depend on apps to perform their daily health activities[5]. Health apps are being developed considering various health care delivery requirements including diagnosis of disease, drug resources, medical calculations, a search of resources, clinical communications, and medical education to be used by several groups of medical specialists, medical students, pharmacists, paramedics, and patients[5].

The World Health Organization's (WHO) global observatory, states that mobile health (mHealth) is a public health practice supported through mobile devices, including smartphones, tablets, personal digital assistants, and other wireless devices. In general, mHealth offers various healthcare services to the public, including online appointments, remote monitoring of patients, video consultation, medical diagnosis, and disease prevention and management. These applications are categorized as drug and clinical apps, telemedicine apps, diagnostic apps, medical education apps, and health and fitness apps developed for medical professionals, patients, and the general public. Health apps aimed at improving the productivity and efficiency of healthcare professionals working in clinical settings. These applications serve to enhance clinical knowledge, decision-making, provide health guidelines, and facilitate patient care. Despite offering such an extensive functionality, there is a low adoption of health apps by relevant users, that is a great challenge[6]. The major cause of such low adoption is poor usability. Research indicates that these applications usually fail to provide value because the issue of usability has not been considered. It also shows that users do not usually spend more than 30 s learning how to use an app before switching or deleting[7]. In broader terms, usability refers to both product quality and user experience. The International Organization for Standardization (ISO) defines usability as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use[7]. Several studies highlighted that well-designed health apps can empower both clinicians and patients, while also improving medical adherence. Although a plethora of health apps has been developed so far, however, the effectiveness of such applications is questionable mainly owing to usability problems[8–11]. Users frequently experience critical usability problems such as difficulty navigating complex interfaces, lower digital literacy for patients, non-personalization of information, poorly distinguished alerts, design issues, poor feedback about system state or behavior, no provision for online help, and no support of local language low accessibility for elderly or disabled users, poorly distinguished alerts, poor feedback about system state, and concerns over data security and trust. Health apps are also used in proactive environments where responsiveness, information clarity, and managing interruptions are important. These challenges raise the need for a comprehensive usability evaluation framework with stakeholder input that is specific to healthcare context[11,12].

Poor usability may be caused by factors such as neglecting the users' needs, limited performance expectancy (i.e., the extent to which the users are able to have support in attaining the intended task goals while they are interacting with the product), and insufficient provision of domain knowledge. It is suggested by research that performance expectancy and a thorough analysis of user needs are two important usability factors to be taken into account to enhance the adoption rate of health apps[13]. To evaluate and advance the usability of health apps, various usability evaluation methods exist which identify usability problems through user-application interaction. Such methods include heuristic evaluation, cognitive walk-through, task analysis, field study, goals, operators, methods and selection rules (GOMS) analysis, keystroke level model (KLM), structured interview, cluster analysis, and severity ratings[13]. These methods require the availability of experts from various domains to evaluate application usability, which leads to significant cost and time consumption. Therefore, a sound mechanism is needed, to enable designers to evaluate the usability of applications independently and efficiently without relying on expert support. For effective evaluation outcomes, designers should be guided by a comprehensive framework that clearly define the key usability parameters for health apps. Developing such a framework requires identifying a diverse set of usability parameters preferably by primary users of health apps. Several usability evaluation models have already been developed to evaluate the usability of health apps. However, these models often lack comprehensive, stakeholder-driven parameters that consider context-specific usability concerns such as trust, security, and responsiveness. This study proposes an evaluation framework based on a hierarchical AHP structure with multi-stakeholder viewpoints and a considerably broader range of usability parameters specifically intended for health apps. Thus, the major scope of this work is to present a decision-making model for the usability evaluation of health apps. Considering that the usability evaluation of health apps is a multidimensional problem based on various parameters and sub-parameters, and given the lack of usability evaluation models with a strong theoretical foundation for health apps. This study employs a multi-criteria decision-making method, followed by an extensive literature review, to propose a model for evaluating the usability of health apps[14]. The study addresses three key questions which are as follows: (i) what parameters and sub-parameters affect the usability of health apps? (ii) how are these parameters and sub-parameters determined? (iii) how will these parameters and sub-parameters be prioritized? the primary objectives of this research are to (i) identify the key parameters and sub-parameters that significantly impact the usability of health apps; (ii) prioritize these parameters and sub-parameters by outlining their relative significance; (iii) propose a hierarchical framework to evaluate the usability of health apps.

The rest of this paper is organized as follows: "Related work" section discusses related work and "Methodology" section defines research methodology. "Findings and discussion" section discusses the findings of this research, presents the proposed usability evaluation framework, comparative analysis and empirical validation of the framework. "Conclusion and future work" section is about the conclusion and future work.

Related work

Previous research has introduced various usability evaluation models and frameworks. However, these models have limitations concerning the usability aspects of health apps. The details of these models and frameworks are discussed below:

A study examined the usability of four commonly available mobile devices including iPhone, iPad, and Android devices, both with touchscreen and built-in keyboards in terms of accessing healthcare applications and information (e.g. diet tracking) by adolescents. The Fit between Individual, Task and Technology (FITT) framework was used for usability evaluation. The results indicated that interface quality is an important factor in health apps that should be considered in developing future applications. The work was limited to a certain group of users of urban adolescent so results cannot be generalized for populations of diverse age groups residing in various zones of urban and rural settings[15]. An interactive TV application ALL@MEO provide home care services to the elderly population. The elderly population feel uncomfortable while using smartphones due to their complex interfaces. The usability of ALL@MEO was evaluated using an observational study based on different instruments such as the Post-Study System Usability Questionnaire (PSSUQ) along with general usability questions, ICF-based Usability Scale I (ICF-US_I) and ICF-based Usability Scale II (ICF-US_II), performance evaluation technique and critical incidents records. The usability evaluation results showed that elderly users have a considerable degree of satisfaction with the interaction with the TV set[16].

Stoyanov et al. (2015) developed a criterion to assess the quality of apps, which is known as the Mobile App Rating Scale (MARS). The authors asserted that MARS is a simple and reliable instrument to classify and measure the quality of health apps. It can also be used as a guideline or checklist for the design and development of new high-quality health applications[17]. Househ et al. (2015) presented a usability analysis of four diabetes health apps using the Health IT Usability Evaluation Model (Health-ITUEM). The health ITUEM model consists of various elements including, information needs, flexibility/customizability, learnability, performance speed, and competency, which direct the classification and analysis of the data. Results showed that diabetes applications have a positive impact on diabetic patients; however, the applications with multifunction have relatively low usability in comparison to uni-functional or bi-functional applications[18]. Schnall et al. (2016) investigated the utility of the Information Systems Research (ISR) framework as a guide in designing health apps. The iterative framework identifies barriers and facilitators for the use of health apps to control the spread of HIV. The findings highlight that using ISR is a useful approach for designing health apps, as it incorporates end users' design preferences[19]. However, the research did not address factors related to users' literacy and cognitive functioning, which are crucial for the effective use of health apps. A research team designed a mHealth application usability questionnaire (MAUQ) based on several existing validated instruments. Four versions of MAUQ were created to accommodate different types of mHealth application, including interactive and standalone apps, with specific versions tailored for patients and medical consultants. The questionnaire was employed to evaluate the usability of apps, including an interactive mHealth app and a standalone mHealth app[20]. A quality-in-use integrated measurement (QUIM) is presented to evaluate usability. The QUIM model consists of ten factors each of which corresponds to a particular aspect of usability that is defined in existing standards. Such 10 factors were decomposed into 26 subfactors which were further broken down into 127 specific metrics. This model explains the measurement theory of usability[21]. Another study recognizes weaknesses in existing usability definitions (e.g., ISO 9241 11, ISO 9126 1) which have limited agreement and specificity. They suggest a comprehensive taxonomy for usability attributes to reflect different stages of system development - specification, design and evaluation. They also intend to provide a structured reference for researchers and practitioners to define and measure parts of usability in a consistent and non-redundant way, which is especially useful for user-centered design and creating better evaluation artifacts[22]. An integrated usability model was introduced which comprises five parameters including effectiveness, efficiency, satisfaction, comprehensibility, and safety. The detailed taxonomy of these parameters and their corresponding sub-parameters is presented in a structured format[23]. The PACMAD is proposed by Harrison et al. (2013), combines numerous usability attributes from several models to introduce an exhaustive model. The usability attributes of this model include effectiveness, efficiency, satisfaction, learnability, memorability, error prevention, and cognitive load. The model modifies usability characteristics to tackle the particular challenges in mobile contexts, like small screen size, and input through touch and working in unpredictable environments. However, it does not have specific domain parameters relevant to health-care applications, including trust, security, accuracy, compliance, or data integrity, which are an important consideration when evaluating usability of health apps[24]. This study systematically reviews the current literature on mobile usability and proposes a model specific to mobile contexts. Mobile usability guidelines are disparate and predominantly desktop-focused. This article explores the gap in usable dimensions that are relevant to mobile applications, by making sense of usability dimensions pertaining to mobile apps existing in the literature. Authors identify 25 usable dimensions using content analysis of previous studies, however, only 10 usable dimensions were outlined based on their importance and frequency in the literature they reviewed[25]. To address the changing nature of mobile platforms, a dynamic usability evaluation model was proposed based on the Goal–Question–Metric (GQM) approach for mobile phone applications. Their framework translates qualitative usability goals (effectiveness, efficiency, satisfaction) into specific questions and measurable metrics, allowing developers and evaluators to systematically evaluate usability. While it is a good model for normal applications, the dynamic usability evaluation approach does not include various stakeholder participation or domain specificity; they also did not incorporate both criteria like security, trust or contextual usability related to health applications[26]. This study reported the frequently employed approaches to usability evaluations of mobile applications, with a goal to characterize approaches and to recognize commonly assessed dimensions. The evaluation approaches were mostly surveys and questionnaires, accounting for around 57% of cases, followed by participant observation (23%), interviews (11%), and methods such as think-aloud protocols, heuristic evaluations, and cognitive walkthroughs used much less frequently (3% or less). In light of these

findings, there continues to be a need for a multi-method, theoretically-informed evaluation framework, using stakeholder-informed criteria and subjective and objective usability measurements to improve usability evaluations of health apps in real world settings[27]. Research highlights the fact that usability is a baseline factor of success for mobile applications, particularly in relation to the constraints of the platform (e.g., small screens, limited input, and different orientations) found in smartphones. The article emphasizes the close relationship between usability, usability satisfaction, adoption rate, and performance of mobile applications. They have identified parameters, including ease of usage, response time, and informative and navigational ease[28]. A study describes key usability evaluation methods such as usability testing (including think-aloud protocols and task scenarios), heuristic evaluation, cognitive walkthroughs, pluralistic walkthroughs, and rapid iterative testing (RITE). The paper suggests that to account for usability dimensions in a broad way throughout the design and evaluation process, the multiple methods should be integrated, rather than exclusively relying on one method[29]. The study highlights that usability is an important aspect of health apps which influence the sustained engagement of users with apps. Improving usability through intuitive design, clear navigation and personalized features can support improvements to satisfaction, trust and continued usage. The results identified the importance of user-centered design principles when developing effective digital health interventions[30]. Research analyzed the usability of diabetes apps and found moderate user satisfaction but identified serious usability flaws including poorly structured interface, inconvenient data entry, and insufficient accessibility features. The results of the study are aligned with broader evidence indicating that multifunctional diabetes apps exhibit poor usability, and stressing on the need for user centered design and domain expert involvement[31]. Research criticized provides a comprehensive taxonomy presented with a distinct structure across multiple level for example parameter, sub-parameter and characteristics. However, the emphasize of this comprehensive framework is on general software usability rather than domain-specific (like health or mobile apps), limiting it to mobile health usability even with modification[32]. A study proposed a multi-stage fuzzy inference system for measuring the usability of software applications in order to address the subjectivity and uncertainty involved in usability evaluation. The model was viable; however, it only accounted for a basic set of usability attributes, and it did not consider more contemporary issues like cognitive load, affective response, and context-aware usability[33]. Another research presents a fuzzy-logic-based hierarchical usability model and discusses its implementation for predicting and ranking SDLC model usability. The content of the model has only been validated in terms of SDLC ranking and has not been practically applied to actual software systems (e.g. health apps) or real-life systems, interfaces[34]. A review describes the role of smart mobile and ambient technologies in improving quality of life through health and well-being feedback and personalized care. The review also identified developments in wearable and context-aware systems but highlights important gaps in usability testing, scalability, and privacy[35]. This study introduces the MAUEM framework to assess mobile application usability, including app domain-specific factors such as cognitive load, interruptibility and simplicity. While it shows promise through expert-based evaluation, it has not been used more broadly with users in user testing or cross-domain usage. There is also minimal comparison with other frameworks[36]. The Pentagram model for the evaluation of a teleconsultation app in terms of Quality of Experience (QoE), where the model consists of five dimensions namely availability, integrity, instantaneousness, retainability, and usability. This model provides a balanced assessment of technical reliability and user interaction. However, it lacks emphasize on cognitive and behavioral aspects of usability, which include learnability, memorability, and satisfaction[37]. An enhance usability model focused on mobile health applications was developed by incorporating mobile-specific metrics like accuracy of information, efficiency of task performance, and safety—as well as traditional usability facets. They validate the framework through expert evaluation and comparative evaluation but describe its importance for health-specific app evaluation in constrained contexts[38].

There are several usability models available, but most of these models have either been generalized for all mobile applications or depend only on expert evaluations, with no recognition of the different usability needs concerning health apps to be run on smartphones. Usability parameters such as trust, security, integrity, compliance, availability, and interruptibility are certainly very critical in clinical settings but are seldom found in usability models because they fail to integrate those specific parameters. Additionally, stakeholder diversity is commonly absent in these models as they basically exclude the voices of major user groups such as patients, paramedics, and medical students. Therefore, this proposed new model accommodates a hierarchical structure application of the AHP with inclusion of broader-and-more-relevant usability parameters obtained from both the reviewing kind of literature and multi-stakeholder contributions to ensure more context-aware structured and inclusive model for evaluations of usability specifically targeting health apps.

Methodology

The methodology used in this research consists of four phases: in the first phase, an extensive literature review was conducted to explore and identify the relevant factors that impact the usability of health apps. As a result of this extensive literature review, the relevant usability parameters were identified. In the second phase, a survey was conducted to determine the importance of various identified usability parameters, and only those with a significant impact were further considered in this study. In the third phase, the relative importance (weight) of each key usability parameter was estimated using the most commonly used pairwise comparison method—that is, the analytic hierarchy process (AHP). AHP is one of the MCDM decision-making techniques that is unquestionably effective at reaching judgments through pairwise comparisons of qualitative and quantitative elements. A good method for determining the weight of the parameters used in experts' reasoning processes is presented by AHP[14]. The output of third phase is a ranked list of usability parameters for health apps, which were the basis of the proposed usability evaluation model. Finally, in the fourth phase, the proposed model was compared with existing models. After a thorough review of the study objectives and methodology, ethical approval for this study was granted by institutional review board of Department of IT, University of Gujrat.

Figure 1 graphically illustrates the research methodology. Below we discuss the details of the four phases of research methodology.

Phase 1: systematic literature review to extract usability parameter

To make sure the transparency and reproducibility in determining the pertinent literature, PRISMA 2020 (Preferred Reporting Items for Systematic Literature Review and Meta Analyses) guidelines were used[39]. The purpose of the systematic literature review was to recognize the recurring usability parameters that have been emphasized in recent studies. The search was made in six major scholarly repositories including IEEE Xplore, ScienceDirect, PubMed, SpringerLink, Google Scholar, and Wiley Online Library to cover literature published in the years 2010 to 2024 as this period marks the rapid growth of smartphone-based health applications and related usability research. The search strings and Boolean combinations used to identify the relevant literature are as follows: ("usability" AND ("health apps" OR "mobile health" OR "mHealth applications") AND ("evaluation" OR "framework" OR "model")). The preliminary search returned 87 studies. After scanning titles and abstracts of the papers and removing duplicates, 70 records were reserved for full paper evaluation. The inclusion criteria consider studies (i) concentrate on research related to usability of smartphone or mobile health applications (ii) discuss usability parameters of smartphone apps, evaluation frameworks or models (iii) publications in English language. The exclusion criteria comprised (a) studies which are not relevant to smartphone or mobile phone health apps (b) publication lacking explicit usability criteria, and (c) editorials, grey literature and non-peer reviewed papers. Considering such criteria, 54 studies were finalized for detailed qualitative synthesis whereas
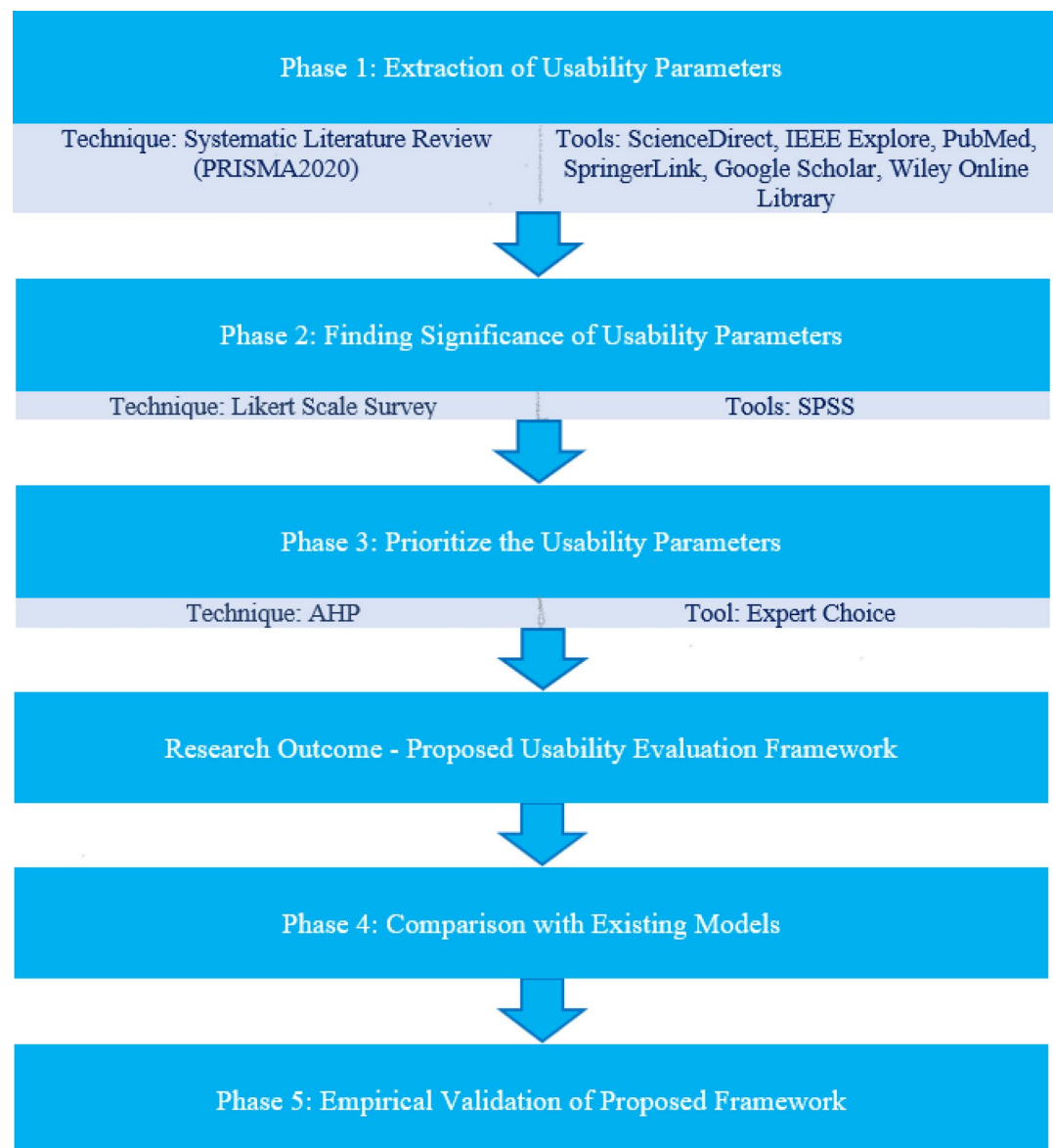

Fig. 1. Research methodology.

33 studies were discarded as they did not fulfill the inclusion criteria. The overall summary of the selection process is shown in the Fig. 2. This systematic approach ensures that the process is transparent, reproducible and consistent with the standards to develop comprehensive development of framework.

Phase 2: finding the significance of usability parameters

In phase 2, the significance of usability parameters was determined, and only those exceeding a predetermined threshold value were taken for further analysis. To identify the key usability parameters, this study utilized a survey method, as surveys are an effective means of collecting data from a broad population. The participants were requested to provide their opinion and mention the level of importance of a specific usability parameter based on 5-point Likert scales, with the choices ranging from 'very important' to 'not important'. Once participants had completed the survey, instruments were obtained for further analysis. To perform the qualitative analysis of the gathered data, the IBM SPSS statistics tool version 21.0 was used. The numeric values from 1 to 5 were assigned to each choice of the Likert scale, where 5 was assigned to 'very important' and 1 to 'not important'. The targeted population of this survey was healthcare professionals (patients, doctors, nurses, paramedical staff, etc.), who used health apps and filled out the questionnaire. The non-probability snowball sampling method was used to reach health-app users and professionals with relevant experience. The technique utilized the professional networks and academic circles to invite participants, ensuring coverage across multiple health related groups and general users[40].

The target population for the present research consisted of healthcare stakeholders who either use or have used smartphone health applications. Their practical experience and domain knowledge were considered vital for the identification and prioritization of usability parameters. Therefore, the following inclusion and exclusion criteria were applied: (i) Those people included who were employed in healthcare sector or health-related education (e.g., doctors, pharmacists, paramedics, and medical students). (ii) Those users from general population were involved who had at least some prior experience with smartphone health applications either for professional or personal health management purposes. and (iii) voluntary consent to the participation in the study. The exclusion criteria include (a) respondents who had no experience in using health apps. (b) incomplete or inconsistent answers. The justification for choosing these groups of participants was to ensure the representation of multiple stakeholders including the perspectives from both the clinical and non-clinical sides, which is crucial to develop a usability framework that reflects the wide-ranging expectations of end users in the context of smartphone health apps. The selection of health applications was open to all kinds of apps; instead, participants were instructed to reply on the basis of their experience with any smartphone health apps (e.g., fitness tracking, telemedicine, medication management) and so forth. This open-ended method enabled capturing usability perceptions across a broader spectrum of commonly used apps, thus making the findings more broadly applicable. The past experience of the participants with smartphone health apps was measured directly by a background section in the questionnaire that had items on the duration of app use (less than 1 year, 1–3 years, 3–5 years, more than 5 years). During the AHP phase, only those participants who had at least three years of app-usage experience were included for making informed and consistent pairwise comparisons.

A total of 195 questionnaires were distributed to participants who belong to different stakeholder groups. Among them, the six incomplete and vague survey questionnaires were excluded. The rest of the 189 were utilized to analyze the collected data. After getting the data from healthcare professionals and general users, the SPSS software was utilized for further analysis. The questionnaire was used mainly for two purposes: firstly, to identify and filter out the less significant usability sub-parameters derived from literature by applying a threshold mean value; and secondly, to empirically validate the conceptual framework by translating expert judgments into quantifiable inputs for the AHP weighting process.

The demographic profile of the respondents is as follows: The respondents in terms of usage experience were 120(61.54%) respondents with < 1 year of experience, 26(13.33%) respondents with < 3 years of experience, 35(17.95%) respondents with 3 years of experience, 12(6.15%) respondents with 3–5 years of experience, and 02(1.03%) respondents with > 5 years of experience of using health apps. In terms of qualification, 108(57.14%) respondents were MBBS, 21(11.11%) were pharmacists, 24(12.69%) were nursing diploma holders, 11(5.2%) were midwifery diploma holders, 01(0.52%) were undergraduate, 03(1.58%) were graduate, 17(8.99%) were postgraduate, 14(2.11%) were with M.Phil. qualifications. Amongst them, 108 (57.14%) were doctors, 21 (11.11%) were pharmacists, 35(18.51%) were paramedical staff and 25(13.22%) were related to other academics, as shown in below Table 1.

The priori power analysis was conducted using standard parameters for behavioral and usability research such as (significance level $\alpha = 0.05$, statistical power = 0.80, and medium effect size as recommended by Cohen). According to the analysis, at least 150 participants were needed in order to identify significant effects with the given power. With 189 valid responses, the final dataset surpasses this criterion and guarantees enough statistical power and representativeness for the usability parameters of smartphone health apps.

Phase 3.3: prioritization of usability parameters

The main objective of this phase is to find out the relative weights and rankings of the usability parameters established in Phase 2. The usability dimension weights have been estimated using the pairwise comparison approach. Pairwise comparison refers to any method of comparing two entities to determine which is preferred, which possesses a higher quantity of a particular quantitative attribute, or whether the two entities are similar. The origin of the approach can be traced back to the renowned multi-criteria decision-making framework known as AHP which is employed in various areas of research[14,41]. This study employed the pairwise comparison method to determine the relative importance of the key usability parameters which are as follows:
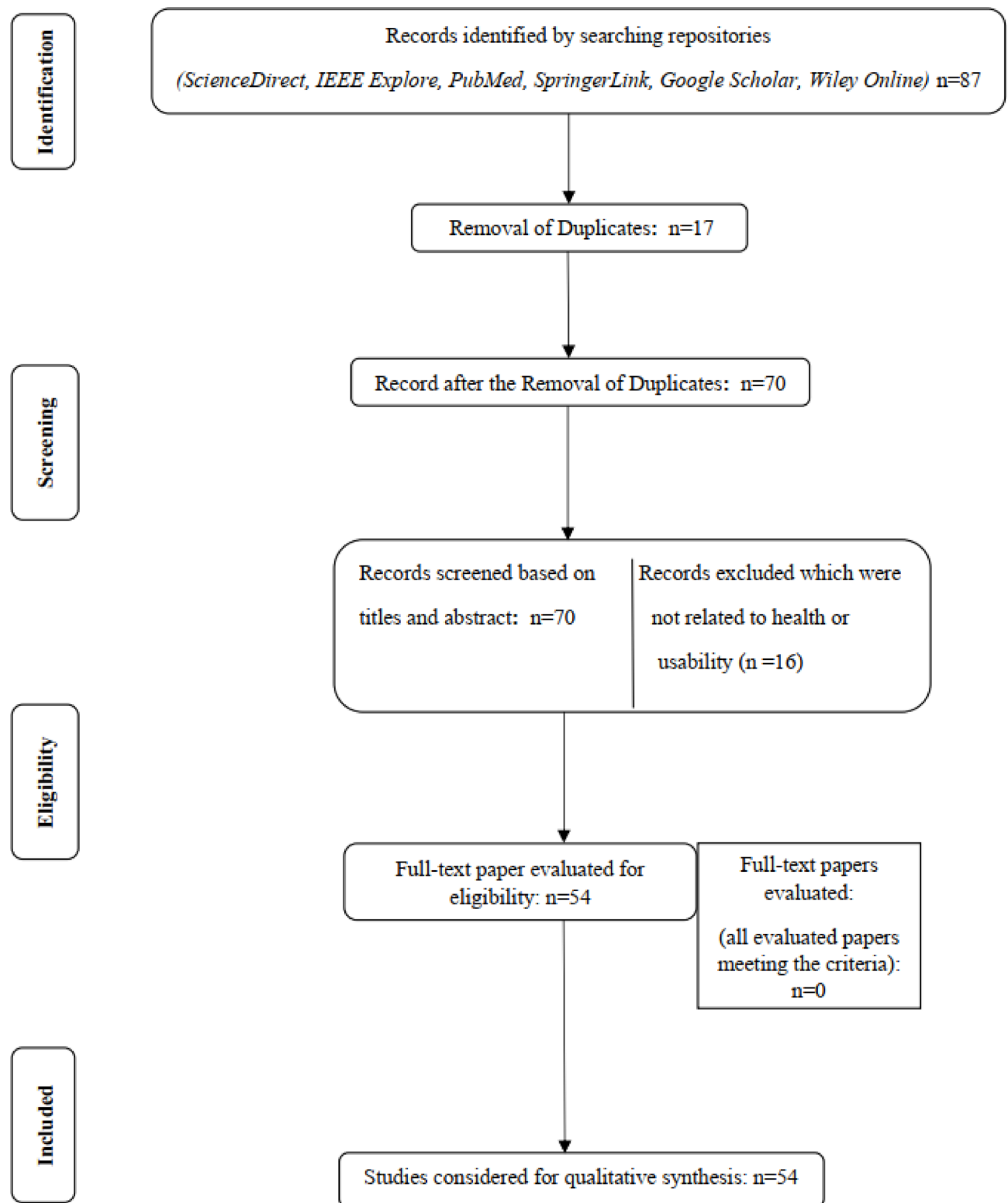
Fig. 2. PRISMA2020 flow diagram.

Fill out the matrix for pairs of comparisons
In this method, the relative importance of two parameters is examined through a scale containing values ranging from 1 to 9. The pair of parameters is assigned a value of 1 if parameter *Pi* is certainly as significant as parameter *Pj*. The value 9 is assigned if one parameter, *Pi*, is much more important than the second parameter, *Pj*. Intermediate values are used for varying degrees of importance as shown in Table 2. For example, where *Pi*

	Frequency	Percent
Experience of apps usage		
< 1 year	120	61.54
1- <3 year	26	13.33
3–5 year	47	24.10
> 5 year	02	1.03
Qualification		
MBBS	108	55.38
Pharmacy/Pharm D.	21	10.77
Nursing diploma	24	12.31
Midwifery diploma	11	5.64
Under graduate	05	2.56
Graduate	03	1.54
Postgraduate	17	8.72
M.Phil.	06	3.08
Profession		
Doctors	108	57.14
Pharmacists	21	11.11
Paramedical staff	35	18.51
Other users	25	13.22

Table 1. Demographic profile of the participants.

Definition	Index	Definition	Index
Equally important	1	Equally important	1/1
Moderately more important	3	Moderately less important	1/3
Much more important	5	Much less important	1/5
Far more important	7	Far less important	1/7
Extremely more important	9	Extremely less important	1/9

Table 2. Pairwise comparison matrix (AHP) scale index values.

is less significant than *Pj*, fractional values from 1/1 to 1/9 are utilized. The fractions 1/1 to 1/9 are provided for 'less important' relationships; 1/9 specifies that *Pi* is significantly less important than *Pj*.

A questionnaire was developed and distributed to participants with at least three years of experience using health apps, to assess the index values based on expert opinions. Experts were asked to evaluate the significance of each usability parameter in relation to other usability parameters, using scale shown in Table 2, and document their evaluations. If there is a variance in the estimates of experts, a consensus technique can be applied to minimize the divergence. A cross-matrix C (n x n) is populated row by row with the estimates approved finally. Equation (1) first populates the diagonal of C with values of 1. Second, until every parameter has been compared to every other parameter, the right upper half of C is filled. If P_i to P_j was evaluated with the relative significance of m (*i.e.*, $C_{ij} = m$), *Pj* to *Pi* must be rated with 1/m (*i.e.*, $C_{ji} = 1/m$). Finally, through Eq. (2), the corresponding fractions are filled in the lower left side of *C*. (Note that the parameters of *C* in row *i* and column *j* are denoted by C_{ij}, and that *i* and *j* are positive integers ≤ n).

$$C_{ij} = 1, \ i = j \tag{1}$$

$$C_{ij} = \frac{1}{C_{ji}}, \ i \neq j \tag{2}$$

Determine the comparison matrix that is normalized
By dividing each parameter in matrix C by the total of the parameters in its column, a normalized comparison matrix C' is produced. Equation (3) shows this.

$$C'_{ij} = \ C_{ij} / \sum {}^{n}_{i=1} C_{ij} \tag{3}$$

Determine the factors relative weights
Equation (4) shows how to calculate the mean of each row in C' to obtain the weight w_i of each parameter *Fi*.

$$w_i = \frac{1}{n} \sum_{j=1}^{n} C'_{ij} \tag{4}$$

Equation (5) demonstrates that these weights are already normalized, with a sum of 1.

$$0 \leq w_i \leq 1$$

$$\sum_{i=1}^{n} w_i = 1 \tag{5}$$

Verify the consistency of the pairwise comparison results
Saaty states that a consistency ratio of less than 10% is acceptable; if not, pairwise comparisons need to be adjusted (Lane & Verdini, 1989). Equation (6) provides the ratio of consistency (CR).

$$CR = \frac{CI}{RI} \tag{6}$$

Where CI is consistency index that is shown by Eq. (7)

$$CI = \frac{\lambda_{max} - n}{n - 1} \tag{7}$$

The rank of pairwise comparison matrix is denoted by *n*, and λ_{max} is maximum eigenvalue.

The random index (RI) of consistency varies in value based on the number of parameters, as shown in Table 3.

Several software tools exist to implement the AHP approach. Some important tools are AUTOMAN, Criterium, HIPRE3, and Expert Choice. The Expert Choice is known as standard AHP software. It is therefore Expert Choice is used in this research to implement AHP[41]. AHP is implemented through the following steps.

(1) Specify research goal

The research goal of this study is to evaluate and rank the important usability parameters of health apps.

(2) Arrange goal and evaluation parameters in hierarchical format

The initial level of hierarchical structure, level 1 of the hierarchy defines the research goal. The second level defines the major usability parameters and the third level outlines the sub-parameters corresponding to each usability parameter.

(3) Calculate the relative weights

Calculate the relative weights for parameters and sub-parameters through above discussed steps 1 to 4. A pairwise comparison involves n $(n-1)/2$ comparisons where 'n' denotes number of parameters or sub-parameters[41].

Research outcome: proposed usability framework
In this stage, usability parameters were evaluated by end users and domain experts such as doctors, pharmacists, paramedics, medical students and regular smartphone health application users. This combination confirmed that the data reflected both practical exposure of users and professional judgements. Every construct of usability including efficiency, effectiveness, satisfaction and comprehensibility in turn was operationalized with certain directly measurable sub- parameters derived from existing usability frameworks. Participants used a five-point Likert scale to rate how important each sub-parameter was, (from 1 = not important at all, to 5 = extremely important). After collecting the scores, the average for each one was calculated to determine their importance, and only sub-parameters with above average importance (i.e. determined using threshold value) were retained for further analysis. These filtered and validated sub-parameters were further employed in AHP pairwise comparison process to ascertain relative weights and rank the parameter and sub-parameters. This multi-step process guaranteed that the constructs operationalization was both empirically grounded and methodologically aligned, boosted framework's reliability and replicability.

Phase 4: comparison with existing models
The comparison of proposed model with previously introduced models has been made in terms of usability parameters recognized by each model for the evaluation of smartphone health apps. The comparison is presented in subsequent Table 9. The table includes two columns; 1st column was consisting of different models with the

N	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49

Table 3. Random index values.

proposed model at the top and 2nd column was consist of usability parameters identified in each model. The comparison was made by highlighting the presence and absence of usability parameters in a particular model.

Phase 5: empirical validation of the proposed framework

A pilot validation was performed on two commonly used smartphone health apps to develop the practical relevance and preliminary empirical validity of the proposed usability evaluation framework. To ensure diversity in user interface design and functionality the chosen applications represent different types of smartphone health apps such as Welltory (i.e. diagnostic and monitoring app) and Oladoc (i.e. a facilitation app). Three experience evaluators participated in the validation process. Every evaluator had prior usability evaluation experience in mobile apps. Additionally, a brief orientation session was also given regarding the structure and usage of the proposed framework. The finalized usability parameters of the framework weighted using AHP technique were converted into a structure evaluation sheet using a 5-point Likert scale (1 = extremely poor usability, 5 = excellent usability). Every evaluator independently interacted with these two apps for at least 20 min, exploring the user interface design, navigational structure, feedback mechanism, information presentation and overall user experience according to the proposed framework. The weighted usability score (WUS) for each app was calculated by multiplying individual evaluator ratings with corresponding AHP weight. The WUS Eq. 8 is as follows:

$$WUS = \sum_{i=1}^{27}(Rating_i \times Weight_i) \tag{8}$$

To evaluate whether the framework allows agreeing scoring among evaluators, inter-rate reliability (IRR) was calculated using the interclass correlation coefficient (ICC) that is a commonly known reliability indicator for multi-rater usability research.

Findings and discussion

The findings related to previous discussed four phases are as follows:

Findings of phase 1

The extensive literature review results in the identification of a large number of relevant usability parameters that could play a vital role in evaluating the usability of health apps. The identified parameters are classified into four major categories including efficiency, effectiveness, satisfaction, and comprehensibility. The resulting usability parameters and sub-parameters are shown in Table 4.

Findings of phase 2

A total of 195 questionnaires were administered among participants. The sample size was selected carefully to gather data, only those participants were selected who have usage experience with health apps. Of these, 189 valid responses were subjected to analysis. Six responses fell outside the scope of analysis due to incompleteness, vague, or ambiguous answers, thus giving an outstanding response rate of approximately 96.9%, reported as excellent for survey-based usability research. The fact that there is a high response rate minimizes the risk of significant non-response bias.

The results of the analysis are shown in Table 5. The detailed literature review produces a large number of usability sub-parameters which needs to be reduced because they further are presented to respondents. So, there is a need to eliminate some less important sub-parameters before applying the pairwise comparison of AHP in the next phase. The less important sub-parameters are eliminated using a threshold (mean value) that is 3.9. The usability parameters that have a mean value less than 3.9, were excluded from further analysis. Resultantly, of the 42 sub-parameters, 15 fell below 3.9 and were eliminated, leaving 27 sub-parameters for AHP analysis. The threshold of 3.9 was set as a slightly conservative cut-off point below 4.0 and it was based on mean ratings from a 5-point Likert scale for sub-parameters relevance that were carried out using a survey. This threshold is not only practical and discriminative for keeping the most meaningful items, but also it can filter out the less significant ones. A strict threshold of 4.0 would have been too stringent and, as such, it could have ended up removing sub-parameters still recognized as important by the majority of respondents, but that would have had some minor differences because of stakeholder perspectives. The value 3.9 permits the inclusion of those items that are rated as "important to very important" by different user groups, thus ensuring that the decision-making process is inclusive and at the same time does not lose its focus in the AHP model.

The results shown in Table 5 indicate that among the sub-parameters of usability parameter "Efficiency", 'performance' is marked as very important with the highest mean value of 4.22, and 'multimedia' is less important with a mean value of 3.57. This finding indicates that users look for smoother, more stable performances of applications rather than adding rich media to scores as the measure of efficiency. Among the sub-parameters of usability parameter "Effectiveness", 'accuracy' is very important with the highest mean value of 4.15, and 'layout' is less important with the lowest mean value of 3.55. This finding suggests that users consider in their evaluation, accuracy and correctness of functioning more critical than the visual arrangement of elements.

Among the sub-parameters of usability parameter "Satisfaction", 'security' is the most important parameter with the highest mean value of 4.63, and, 'design' is less important with the lowest mean value of 3.22. These results imply that users value protection of their data and security in interactions more than aesthetic design when gauging satisfaction.

Parameters	Sub- parameters	References
Efficiency	Productivity	27,29,32,33
	Performance	42,43,45
	Features	26
	Integrity	33,37
	Retainability	26
	Instantaneousness	37
	Availability	33,37,53
	Multimedia	43
Effectiveness	Accessibility	29,49,52–54
	Operability	22,31,38
	Universality	27,33,35,38
	Flexibility	43
	Robustness	22
	Error prevention	29,30,36,50
	Completeness	43
	Simplicity	25,26,36
	Accuracy	26
	Layout	54
	Content	44,54
Satisfaction	Comfort/ease of use	28,31,34,42,50
	Aesthetic/attractiveness	26,29,31,38,53
	Design	45
	Competency	34,43
	Security/Safety	26,27,33,35,46,47,51
	Trustfulness	46,47,51,54
	Reputation	44
	Information needs	43,47
	Awareness	34,51
	Mobility	34
	Preparedness	34
	Immediacy	34,37
	Compliance	5,31,47
	Navigation	28,45,50
Comprehensibility	Usefulness	25,54
	Response time/time taken	26,28,34
	Learnability	30,31,36,38,48,51
	Memorability	33,35–37,48
	Understandability	25,29,31,49,51
	Intuitiveness	25
	Cognitive load	24,30,36,48
	Interruptibility	36
	Appropriateness	29,30

Table 4. Extracted usability parameters and sub-parameters.

Among sub-parameters of usability parameter "Comprehensibility", 'response time' is marked as very important as it has the highest mean value of 4.25 whereas 'cognitive load' is marked as less important as it relatively has the lowest mean value of 3.84. Hence, fast interaction and quick system response are crucial for health app users to operate them effectively.

Findings of phase 3
AHP was applied through Expert Choice software to find weights and ranking of the usability parameters. AHP sets problems in a tree such as objectives, parameters, and alternatives. The study utilized the AHP technique for the ranking and prioritization of usability parameters and then proposed a usability evaluation model. The implementation steps and their results are shown below.

(1) Defining the Goal: In the first step of AHP, the goal of the study is defined as forming a hierarchical model to evaluate the usability of health apps.

(2) Identification of Parameters and Sub-Parameters: The major usability parameters of health apps identified in this study include efficiency, effectiveness, satisfaction, and comprehensibility. Furthermore, 27 sub-parameters were extracted previously, and incorporated through Expert Choice software.

(3) Hierarchical Model Construction: A three-level hierarchical model is constructed using Expert Choice software. The first level of hierarchy defines the main objective or problem; the second level describes the major parameters and level three describes sub-parameters.

(4) Processing gathered data: A survey questionnaire with pair-wise comparisons is processed through a software tool

(5) Pair-wise Comparison: In this step, pairwise comparison is performed to determine the importance of one parameter over other parameters. The results based on participants' responses are shown in Tables 5, 6, 7, 8 and 9, and 10.

(6) Consistency Test: The consistency check was performed at this step. Based on the consistency ratio (CR) > 10%, of the 49 expert responses collected, 24 were excluded as they were not meeting the consistency check. Rest of the 26 valid responses were considered for AHP calculations.

(7) Calculating Local and Global Weights: The local and global weights of the usability parameters and sub-parameters were calculated based on the data loaded into the software. The weights were categorized into global weights (weights of four usability parameters) and local weights (weights of sub-parameters). The global weights and ranks of usability parameters including efficiency, effectiveness, satisfaction, and

Parameters	Sub- parameters		Contribution level					Mean	Std dev
			Very important	Important	Neutral	Less important	Not important		
Efficiency	Productivity	189	61	105	23	0	0	4.20	0.64
		%	32.3	55.5	12.2	0	0		
	Performance	189	62	106	21	0	0	4.22	0.63
		%	32.8	56.1	11.1	0	0		
	Features	189	25	95	67	1	1	3.75	0.71
		%	13.2	50.4	35.4	0.5	0.5		
	Integrity	189	33	107	47	2	0	3.90	0.68
		%	17.5	56.6	24.8	1.1	0		
	Retainability	189	32	105	52	0	0	3.89	0.66
		%	16.9	55.6	27.5	0	0		
	Instantaneousness	189	26	91	70	2	0	3.75	0.70
		%	13.8	48.1	37	1.1	0		
	Availability	189	40	108	40	1	0	3.99	0.67
		%	21.2	57.1	21.2	0.5	0		
	Multimedia	189	28	58	97	5	1	3.57	0.79
		%	14.8	30.7	51.4	2.6	0.5		
Effectiveness	Accessibility	189	33	126	30	0	0	4.02	0.58
		%	17.5	66.7	15.8	0	0		
	Operability	189	29	124	36	0	0	3.96	0.59
		%	15.4	65.6	19	0	0		
	Universality	189	32	116	41	0	0	3.95	0.62
		%	16.9	61.4	21.7	0	0		
	Flexibility	189	28	112	48	1	0	3.88	0.64
		%	14.8	59.3	25.4	0.5	0		
	Robustness	189	40	117	32	0	0	4.04	0.62
		%	21.2	61.9	16.9	0	0		
	Error prevention	189	47	109	33	0	0	4.07	0.65
		%	24.8	57.7	17.5	0	0		
	Completeness	189	38	127	23	0	1	4.06	0.61
		%	20.1	67.2	12.2	0	0.5		
	Simplicity	189	42	122	25	0	0	4.09	0.59
		%	22.2	64.6	13.2	0	0		
	Accuracy	189	54	110	25	0	0	4.15	0.63
		%	28.6	58.2	13.2	0	0		
	Layout	189	24	67	87	11	0	3.55	0.79
		%	12.8	35.4	46	5.8	0		
	Content	189	37	72	80	0	0	3.77	0.76
		%	19.6	38.1	42.3	0	0		
Continued									

Parameters	Sub- parameters		Contribution level					Mean	Std dev
			Very important	Important	Neutral	Less important	Not important		
Satisfaction	Comfort/ease of use	189	35	121	33	0	0	4.01	0.60
		%	18.5	64	17.5	0	0		
	Aesthetic/attractiveness	189	33	84	66	5	1	3.76	0.79
		%	17.5	44.4	34.9	2.7	0.5		
	Design	189	12	54	88	33	2	3.22	0.85
		%	6.3	28.5	46.6	17.5	1.1		
	Competency	189	46	103	40	0	0	4.03	0.68
		%	24.3	54.5	21.2	0	0		
	Security/safety	189	121	66	2	0	0	4.63	0.51
		%	64	34.9	1.1	0	0		
	Trustfulness	189	67	99	23	0	0	4.23	0.65
		%	34.6	53.2	12.2	0	0		
	Reputation	189	38	63	74	14	0	3.66	0.88
		%	20.1	33.3	39.2	7.4	0		
	Information needs	189	25	134	29	1	0	3.97	0.56
		%	13.3	70.9	15.3	0.5	0		
	Awareness	189	28	120	39	1	1	3.92	0.65
		%	14.8	63.6	20.6	0.5	0.5		
	Mobility	189	27	60	93	8	1	3.55	0.81
		%	14.4	31.7	49.2	4.2	0.5		
	Preparedness	189	31	105	52	0	1	3.87	0.69
		%	16.4	55.6	27.5	0	0.5		
	Immediacy	189	34	112	42	1	0	3.95	0.65
		%	18	59.3	22.2	0.5	0		
	Compliance	189	40	103	44	2	0	3.96	0.70
		%	21.2	54.4	23.3	1.1	0		
	Navigation	189	27	104	54	4	0	3.81	0.69
		%	14.3	55	28.6	2.1	0		
Comprehensibility	Usefulness	189	43	110	35	1	0	4.03	0.66
		%	22.8	58.2	18.5	0.5	0		
	Response time/time taken	189	62	113	14	0	0	4.25	0.58
		%	32.8	59.8	7.4	0	0		
	Learnability	189	51	115	23	0	0	4.15	0.61
		%	27	60.8	12.2	0	0		
	Memorability	189	32	113	42	2	0	3.93	0.66
		%	16.9	59.8	22.2	1.1	0		
	Know/understandability	189	45	122	22	0	0	4.12	0.58
		%	23.8	64.6	11.6	0	0		
	Intuitiveness	189	42	84	59	4	0	3.87	0.78
		%	22.3	44.4	31.2	2.1	0		
	Cognitive load	189	26	108	54	0	1	3.84	0.67
		%	13.8	57.1	28.6	0	0.5		
	Interruptibility	189	39	106	42	2	0	3.96	0.69
		%	20.6	56.1	22.2	1.1	0		
	Appropriateness	189	39	105	41	3	1	3.94	0.73
		%	20.6	55.6	21.7	1.6	0.5		

Table 5. Analysis of data collected through survey.

comprehensibility are shown in Table 6. The local weights of corresponding sub-parameters are shown in Tables 7, 8, 9 and 10.

The operationalization of all usability parameters and sub-parameters represented in the proposed framework are presented by the thorough explanation in Sect. 5. Each sub-parameter characterizes a measurable attribute of usability and consequently serves as the direct operational definition of major parameter. The assessment of each sub-parameter is performed using experts scoring on a 5-point Likert scale (1 = strongly disagree, 5 = strongly

Main parameters	Weights	Ranks
Efficiency	0.369	1
Comprehensibility	0.326	2
Satisfaction	0.165	3
Effectiveness	0.140	4

Table 6. Global weights and ranks of main parameters.

Sub-parameters	Weights	Ranks
Productivity	0.389	1
Availability	0.258	2
Integrity	0.217	3
Performance	0.136	4

Table 7. Local weights and ranks of sub-parameters of efficiency.

Sub-parameters	Weights	Ranks
Learnability	0.178	1
Usefulness	0.156	2
Memorability	0.147	3
Understandability	0.136	4
Interruptibility	0.129	5
Appropriateness	0.127	6
Response-time	0.127	6

Table 8. Local weights and rank of sub-parameters of comprehensibility.

Sub-parameters	Weights	Ranks
Security	0.178	1
Information needs	0.131	2
Immediacy	0.128	3
Competency	0.120	4
Comfort	0.119	5
Awareness	0.117	6
Trustfulness	0.109	7
Compliance	0.098	8

Table 9. Local weights and ranks of sub-parameters of satisfaction.

agree). The relative significance (local weights) of usability sub-parameters was computed using AHP process and equations discussed in preceding "Phase 3.3: prioritization of usability parameters" subsection. The expert judgements were aggregated through geometric mean method and local priority vectors were gained through the eigenvector method with consistency confirmed using CI and CR calculations. Such local weights were further multiplied by the major parameter weights to obtain global weights.

Table 6 demonstrates the weights and ranks of usability parameters. The usability parameter "Efficiency" measured value is (0.369, 36.9%), which indicates its significance and importance in terms of evaluation of usability of health apps. "Comprehensibility" is found as the second most important usability parameter and its measured value is (0.326, 32.6%). Similarly, 'Satisfaction' is ranked in the third position and its measured value is (0.165, 16.5%) whereas 'Effectiveness' is placed in the fourth position and measured at (0.140, 14%) and ranked number 4. Table 7 shows that 'productivity' is the most important usability sub-parameter of the "Efficiency" parameter with the highest weight which is (0.389, 38.9%) while 'performance' is found relatively least important sub-parameter, it's a weight value is (0.136, 13.6%). Table 8 shows that 'learnability' is the most important usability sub-parameter of "Comprehensibility" with the highest weight value which is (0.178, 17.8%) while the 'appropriateness' and 'response time' are found to be relatively least important sub-parameters and their weight value is (0.127, 12.7%). Similarly, Table 9 shows that 'security' is the most important usability sub-parameter of the "Satisfaction" parameter with the highest weight that is (0.178, 17.8%) whereas 'trustfulness' is

Sub-parameters	Weights	Ranks
Simplicity	0.165	1
Universality	0.142	2
Accessibility	0.132	3
Robustness	0.130	4
Error-prevention	0.121	5
Completeness	0.115	6
Accuracy	0.098	7
Operability	0.097	8

Table 10. Local weights and ranks of sub-parameters of effectiveness.

found relatively least important sub-parameter and its weight is (0.098, 9.8%). Table 10 shows that 'simplicity' is the most important usability sub-parameter of the "Effectiveness" parameter with the highest weight that is (0.165, 16.5%) while 'operability' is found relatively least important sub-parameter and its weight is (0.097, 9.7%). Table 11 delineates the overall weights and rankings of four usability parameters and their corresponding sub-parameters. It presents the global weights and rankings of the parameters and sub-parameters, along with their local weights and rankings. Global ranking indicates the overall significance of all parameters whereas local ranking refers to importance solely within the parent parameter.

Proposed usability evaluation framework

The proposed framework (Fig. 3) consists of four major usability parameters (efficiency, comprehensibility, satisfaction, effectiveness) weighted using the AHP. In the AHP, experts performing pairwise comparisons on each parameter, and translated this judgment into numerical priorities (weights). The obtained weights were normalized to sum up to 1.0, meaning that the weight of about 0.369 for efficiency implies that it accounts for 36.9% of the total importance in the usability score. In other words, it means that efficiency is considered to be the most critical factor (0.369, 36.9%), next comes comprehensibility (0.326, 32.6%), then satisfaction (0.165, 16.5%), and lastly effectiveness (0.140, 14%). Each weight indicates the degree of influence that parameter has on the general usability evaluation; for instance, almost 37% of the score comes from efficiency. This was derived directly from judgments by the experts participating in the analysis via AHP.

These findings present the practical realities of usage among users, and especially among healthcare professionals, for whom efficiency has been and is at the top priority in supporting the rapid and accurate completion of tasks in clinical settings that are very time-sensitive. The next parameter has been comprehensibility, which typically concerns the clear and understandable interface for a wide user spectrum-from healthcare professionals to patients. Satisfaction and effectiveness, though important, were rated lower because in high-stakes environments like healthcare, functional performance and clarity take precedence over subjective enjoyment or general task accomplishment. These rankings have very great implications for app developers and designers in that usability improvements should first focus on optimizing speed, responsiveness, and ease of navigation, followed by improving clarity, learnability, and error prevention especially in medical or clinical use cases. The four major usability parameters which are associated with twenty-seven sub parameters are as follows.

Efficiency
In the context of health apps, efficiency is how much a user is able to complete intended tasks with proper speed and accuracy, in relation to time and effort expended. It refers to system's capacity to support interaction by minimizing barriers of time or focus distraction, so that tasks can be carried out successfully and productively. Efficiency was, therefore, composed of four sub-parameters in the proposed model: productivity, availability, integrity, and performance-each of which speaks to system responsiveness, reliability, and functional adequacy while actually being used. The prominent global weight of (0.369, 36.9%) ranking it as the most important usability parameter, efficiency is clearly something that stakeholders value most when it comes to accessing health information and features quickly and seamlessly, especially in time-critical or high-risk medical situations. Thus, it adds to the importance of having the interface and workflow designed in such a way as to put the least load on the person cognitively and operationally, thus allowing the maximum output on any task.

i. With regard to the efficiency parameter, productivity is ranked highest among the four sub-parameters. Productivity relates to the number of useful outputs over the length of time a user interacts with an application. A high level of productivity implies that users can finish more relevant tasks in that timeframe, which supports overall task efficiency. So, for health apps, the more an application can do with actionable outcomes, whether that means a correct diagnosis, timely alerts, or access to health records, the greater its contribution to perceived usability[32].

ii. Availability is the second sub-parameter under efficiency. This refers to the degree to which users can conveniently access the application and its resources successfully real-time in the meaningful format. Availability means the basic functionalities should be easily accessible and continuous because that is the core of perceived efficiency because it reduces unpleasant user experience and therefore supports the smooth completion of the task. Thus, being high in availability assumes direct contribution to overall

Main parameters	Global weight with contribution %	Global ranking of main parameters	Sub- parameters	Local weight with contribution %	Local ranking of sub factors	Global ranking of sub-factors
Efficiency	0.369 (36.9%)	1	Productivity	0.389(38.9%)	1	1
			Availability	0.258(25.8%)	2	2
			Integrity	0.217(21.7%)	3	3
			Performance	0.136(13.6%)	4	9
Comprehensibility	0.326 (32.6%)	2	Learnability	0.178(17.8%)	1	4
			Usefulness	0.156(15.6%)	2	6
			Memorability	0.147(14.7%)	3	7
			Understandability	0.136(13.6%)	4	9
			Interruptibility	0.129(12.9%)	5	13
			Appropriateness	0.127(12.7%)	6	15
			Response-time	0.127(12.7%)	6	15
Satisfaction	0.165 (16.5%)	3	Security	0.178(17.8%)	1	4
			Information needs	0.131(13.1%)	2	11
			Immediacy	0.128(12.8%)	3	14
			Competency	0.120(12.0%)	4	17
			Comfort	0.119(11.9%)	5	18
			Awareness	0.117(11.7%)	6	19
			Compliance	0.109(10.9%)	7	21
			Trustfulness	0.098(9.8%)	8	22
Effectiveness	0.140 4 (14.0%)	4	Simplicity	0.165(16.5%)	1	5
			Universality	0.142(14.2%)	2	8
			Accessibility	0.132(13.2%)	3	10
			Robustness	0.130(13.0%)	4	12
			Error-prevention	0.121(12.1%)	5	16
			Completeness	0.115(11.5%)	6	20
			Accuracy	0.098(9.8%)	7	22
			Operability	0.097(9.7%)	8	23

Table 11. Overall weights and ranking of usability parameters and sub-parameters.

usability when it comes to health apps, especially in time-critical instances for accessing medical information[33].

iii. Integrity is locally ranked in third position, which shows the stability and correctness of an application. A more stable application will complete a greater number of tasks per unit of time, so integrity is a usability sub-parmeter that directly affects the efficiency of any application[33].

iv. Performance is ranked in the fourth position and it specifies the extent to which a design is expected to improve user performance[33]. Improved performance directly contributes to the efficiency of an application. Experimental research supports this usability sub-parameter, showing that during interaction with health apps, users paid special attention to how effectively the interface design facilitates efficient operations[55].

Comprehensibility

Comprehensibility refers to the degree to which a health app offers information in a clear, easy to learn, and memorable way. It specifically points to the user's ability to easily interpret and retain necessary medical information presented by the interface. Due to the wide range of users (healthcare professionals, patients, and medical students), it is important that the content is presented in an understandable and accessible format[56]. The global weight assigned to comprehensibility in the proposed model is (0.326, 32.6%) making it the second most important usability parameter. This top position clearly shows its significance in several aspects, for example, the application that users can efficiently move around, understand the features of the application, recall how to use the application if they want to use it later, which are especially important in the healthcare setting where understanding can be directly linked with user trust, compliance, and safety. The sub-parameters associated with comprehensibility are as follows:
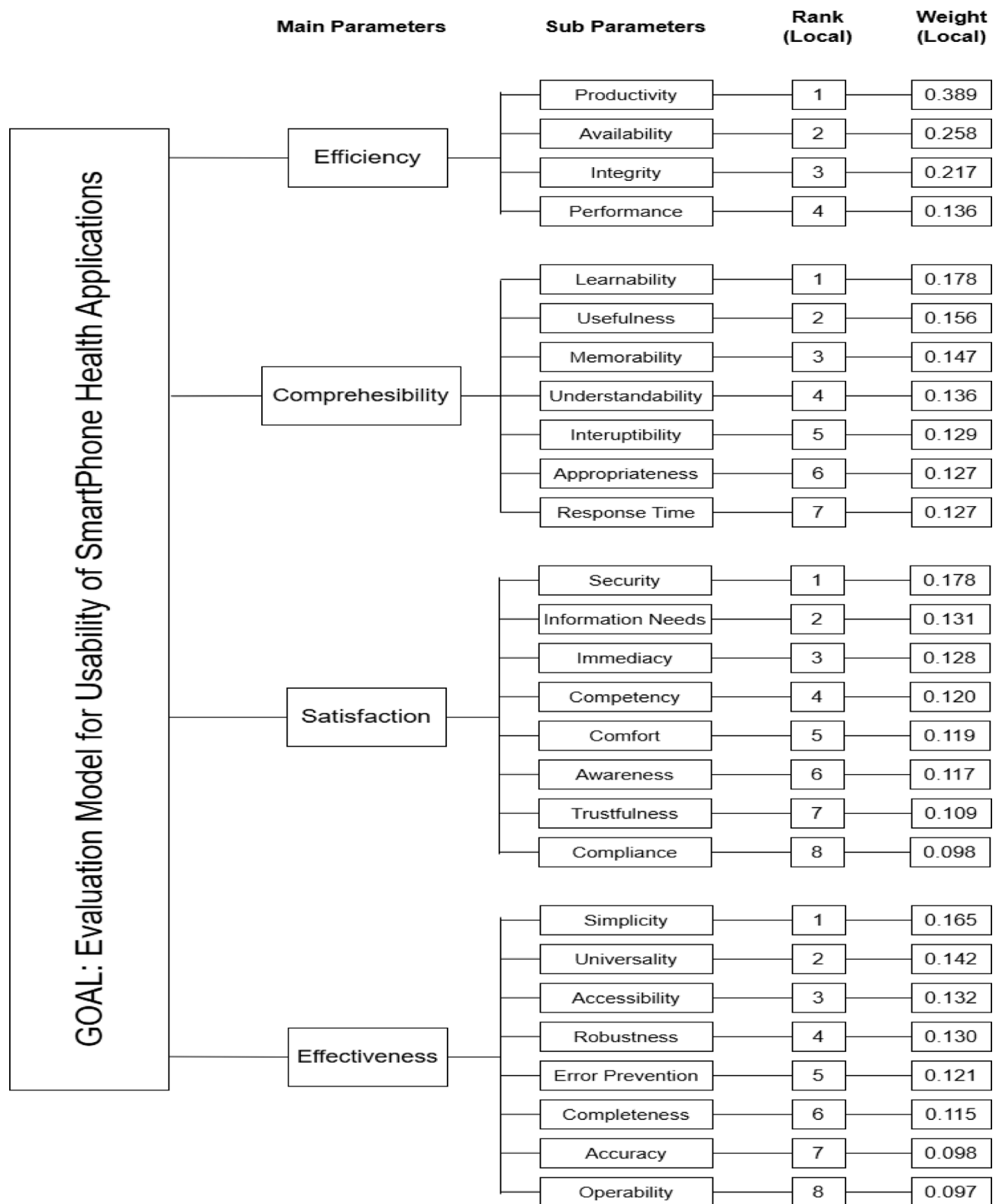
Fig. 3. Proposed usability evaluation framework.

i.   Learnability is the first locally ranked sub-parameter of comprehensibility is learnability which indicates the ability of the user to easily learn and operate the application so if the learnability of an app is high then it would be highly comprehended.

ii.   Usefulness is the second locally ranked sub-parameter which highlights users' ability to achieve the desired results to fulfill their needs and expectations. There should be an existence of important and meaningful information related to medication and treatment results[56].

iii.   Memorability is locally ranked in the third position which indicates that users should be able to remember how to perform tasks even after a long period. So, the design should minimize the memory load by showing visible options, objects, and actions[57].

iv.   Understandability, ranked fourth among the sub-parameters which represents the ability of a user to comprehend navigating or using the application. For this to happen, app content must be clearly stated, accurate, and relevant. App content must present information and visuals in such a way as to effectively guide users while preventing them from drawing on their cognitive resources[56].

v.   Interruptibility, which is the fifth sub-parameter, refers to a new action starting before the initial action has been finished. High interruptibility has the potential to cause ambiguity and interruptions in the task flow of users which can impact applications usability and understanding of the application[36].

vi.   Appropriateness and response time are locally ranked in the sixth position; appropriateness is the meaningfulness of visual metaphors used in the application. If the metaphors are appropriate, the system will be easy to learn and remember. The response time indicates the time taken to complete a task, respond to error messages, read, understand, and make decisions on various feedback messages[30].

Satisfaction

Satisfaction is a usability parameter that is globally ranked in third position. It indicates the extent to which the application fulfills the user's expectations or the level of comfort experienced by the user while interacting with an application[55,56,58]. If there are fewer issues or difficulties experienced while interacting and performing the task through an application there is a higher probability of user satisfaction with an application. The usability parameter satisfaction represents the eight usability sub-parameters which are as follows.

i.   With security being rated as a highest sub-parameter of the satisfaction parameter, it is a vital factor that contributes to users' perception of trust and comfort. When users are confident that their personal data and interactions are shielded, they will likely regard these levels of trust higher than users, who feel possibility of a risk. Therefore, greater attention should be given to the risks and security associated with health apps[55].

ii.   Information is ranked as the second highest sub-parameter of satisfaction parameter, which indicates the importance of considering varying level literacy of users while designing the health apps. This ensures that the content is usable, relevant, and appropriately tailored to support informed decision-making[56].

iii.   Immediacy is ranked third, indicating that users should be able to perform certain operations with fewer clicks while minimizing the risk of errors. If the design supports the user in such a way, then s/he would feel more satisfied[37].

iv.   The competency sub-parameter is ranked fourth, it is about the confidence level of a user in their ability to perform certain tasks. It is the innovative design of an application that makes the user confident that s/he can interact without any difficulty and impediments[42].

v.   Comfort is the degree to which an app produces positive feelings towards its user through an interaction. The user-centered design can play a vital role in enhancing comfort. It is ranked in fifth position[31,34].

vi.   Awareness is the ability of the user to perceive objects, thoughts, and events. It is ranked in sixth position.

vii.   Compliance is ranked in seventh position indicating that the application should have compliance with the domain (i.e. health sector) and usability guidelines.

viii.   Trustfulness is ranked at eighth position; it is about the trust level that an app provides to its users. Trust in health apps is very crucial because if a user does not trust an application s/he will not use it[54].

Effectiveness

The usability parameter effectiveness is globally ranked in fourth position; it indicates the extent to which an interface supports a user in completing the task for which it was intended[55,56]. The usability parameter effectiveness represents eight usability sub-parameters. These sub-parameters are as follows.

1.   Among these eight sub-parameters, simplicity is locally ranked at first position. Simplicity usually leads to clarity and effectiveness. It enables the interface to convey functions more effectively. Moreover, it also offers aesthetic appeal to the design.

2.   Universality is a sub-parameter that is placed at the second number. It accommodates the diversity in the population in terms of background, culture, experience, etc.

3. Accessibility is placed in the third position; it indicates the application's ability to be usable by users with disabilities.
4. The robustness of interface design assists users in the successful completion and evaluation of tasks. If the user is facing any difficulty in performing the task the interface assists him/her in accomplishing the task. This sub-parameter is locally ranked in the fourth position.
5. Error prevention refers to the ability of an interface to minimize the possibility of user errors and to effectively recover from errors that may occur during interaction with the application. This sub-parameter is locally ranked in the fifth position.
6. Completeness ensures that the interface presents all those objects and actions required to effectively perform the task. It is ranked in sixth position.
7. Accuracy is another sub-criterion of effectiveness parameter that ensures the accuracy of different design elements used on the interface of health apps. The accuracy is ranked in the seventh position.
8. Operability measures how successfully users' complete tasks in terms of achieving their goals. It is ranked in the eighth position.

The practical implications of the proposed framework are as follows:

9. The proposed framework offers a structured and hierarchical approach which developers and designers can use to evaluate the usability of health apps based on explicit and contextually relevant usability parameters.
10. As the framework developed by involving all pertinent stakeholders so it ensures that usability evaluations will be relevant across the healthcare landscape best reflecting real world need and facilitate the development of user-centered health apps.
11. The identified usability parameters can act as a design benchmark, as well as awareness for teams thinking about important usability areas that may have been overlooked, specifically trust, appropriateness, cognitive load, and response-time which are even more important with high-risk health-related applications.
12. With the growth of digital health applications, the framework could provide a useful way to inform regulatory criteria.

Findings of phase 4

An analysis was performed to compare the proposed usability evaluation framework against the existing usability models introduced by prior research studies. The comparison aimed to prove the completeness and conceptual coverage of the proposed framework by correlating its four main dimensions including efficiency, effectiveness, satisfaction, and comprehensibility and their corresponding sub-parameters to those in existing models (see Table 12). The analysis showed that the proposed framework not only integrates all essential usability factors (a.k.a. sub-parameters) recognized in previous studies but also includes new sub-parameters like trust, security, interruptibility, and appropriateness, which are especially important in the case of smartphone health apps. This proves that the proposed model has increased comprehensiveness and contextual validity and thus has confirmed its suitability as a robust and domain-specific usability evaluation framework for smartphone health apps. The enhanced usability of the proposed model stems from the systematic approach taken in its development. By thoroughly reviewing relevant research literature and incorporating insights from all key stakeholders, this method resulted in a more authentic and reliable evaluation model.

The findings of this study in terms of usability parameters and sub-parameters are supported by some recent research studies that highlighted the pertinent role and importance of efficiency, effectiveness, satisfaction, and comprehensibility in evaluating and enhancing the usability of health apps[55–59]. Furthermore, the credibility of this research's findings is reinforced by adherence to the International Standard ISO 9241-11, which emphasizes efficiency, effectiveness, and satisfaction as core components of usability in interactive applications[57].

Findings of phase 5

The two smartphone health apps were evaluated using the proposed usability framework, evaluation results showed that both apps are usable. Oladac app achieved a higher usability score (4.18) signifying superior user experience particularly in navigation, information quality and efficacy. Welltory app scores moderately high (3.84) but displays few usability challenges related to interface complexity, navigation clarity and cognitive load. Oladoc is a purpose driven app aimed to book doctor appointment, and consultation. Its design is simple, clean, easy to understand and quality of information is excellent. There are minor issues which do not significantly affect overall usability. On the other hand, welltory provides advanced metrics which may require cognitive effort to interpret them. The app offers rich functionality but cluttered interface, onboarding complexity and hefty content curtail learnability and efficiency. These results highlight that the proposed framework can eloquently differentiate usability and serve as a practical evaluation tool for smartphone health apps.

The calculated value of ICC was (0.82) which signify the strong agreement among the three evaluators. Such high degree of consistency validates that the framework can be reliably employed by different evaluators without significant variation in interpretation or scoring supporting its IRR. The preliminary validation provides empirical evidence about the proposed framework. It showed that the framework is practical and straightforward to use in real world situations. Secondly, it produces usability ratings that is consistent across different smartphone health applications. Strong rater agreement signifies the reliability of application.

Conclusion and future work

A structured usability evaluation framework for smartphone health applications has been proposed, having been developed through a rigorous process involving review of extensive literature, expert opinion, and application of the AHP method. The framework operates by ranking major usability parameters and sub-parameters and

Models		Usability parameters														
	Proposed model	Efficiency	Effectiveness	Satisfaction	Comprehensibility	Productivity	Performance	Integrity	Availability	Accessibility	Operability	Universality	Robustness	Error-prevention	Completeness	Simplicity
22		✓	×	✓	×	×	×	×	×	×	✓	×	✓	×	×	×
23		✓	✓	✓	×	×	×	×	×	×	×	×	×	×	×	×
24		✓	✓	✓	✓	×	×	×	×	×	×	×	×	✓	×	×
25		✓	✓	✓	×	×	×	×	×	×	×	×	×	×	×	✓
26		×	×	×	×	×	×	×	×	×	×	×	×	×	×	✓
27		✓	✓	✓	×	✓	×	×	×	×	×	✓	×	×	×	×
28		×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
29		✓	✓	✓	×	✓	×	×	×	✓	×	✓	×	✓	×	×
30		✓	✓	✓	×	×	×	×	×	×	×	×	×	✓	×	×
31		×	×	×	×	×	×	×	×	×	✓	×	×	×	×	×
33		✓	✓	✓	×	✓	×	×	×	×	×	✓	×	×	×	×
34		✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×
35		✓	✓	✓	×	✓	×	×	×	×	×	✓	×	×	×	×
36		✓	✓	✓	×	×	×	×	✓	×	×	×	×	✓	×	✓
37		×	×	×	×	×	×	✓	×	×	✓	×	×	×	×	×
38		×	✓	✓	×	×	×	×	×	×	✓	✓	×	✓	×	×
43		×	×	×	×	×	✓	×	×	×	×	×	×	×	✓	×
44		×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
45		×	✓	×	×	×	✓	×	×	×	×	×	×	×	×	×
46		✓	×	✓	×	×	×	×	×	×	×	×	×	×	×	×
47		×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
48		×	×	×	×	✓	×	×	×	×	×	×	×	×	×	×
49		×	✓	×	✓	×	×	×	×	✓	×	✓	×	×	×	×
51		×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
53		✓	✓	✓	×	×	×	×	✓	×	×	×	×	×	×	×
54		×	✓	✓	×	✓	×	×	×	✓	×	✓	×	✓	×	×

Models		Usability parameters														
	Proposed model	Accuracy	Comfort	Competency	Security	Trustfulness	Information needs	Awareness	Immediacy	Compliance	Usefulness	Response time	Learnability	Memorability	Understandability	Interruptibility	Appropriateness
22		×	×	×	✓	×	×	×	×	×	×	×	×	×	✓	×	×
23		×	×	×	✓	×	×	×	×	×	×	×	×	×	×	×	×
24		×	×	×	×	×	×	×	×	×	×	×	✓	✓	✓	×	×
25		×	×	×	×	×	×	×	×	×	✓	✓	✓	×	✓	×	×
26		✓	×	×	✓	×	×	×	×	×	×	×	×	×	×	×	×
27		×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
28		×	✓	×	×	×	✓	×	×	×	×	×	×	×	×	×	✓
29		×	×	×	×	×	×	×	×	×	×	×	✓	×	×	×	×
30		×	×	×	×	×	×	×	×	×	×	×	✓	✓	×	×	×
31		×	×	×	✓	×	×	×	×	✓	×	×	✓	×	✓	×	×
33		×	×	×	✓	×	×	×	×	×	×	×	×	✓	×	×	×

Continued

20

Models	Usability parameters															
Proposed model	Accuracy	Comfort	Competency	Security	Trustfulness	Information needs	Awareness	Immediacy	Compliance	Usefulness	Response time	Learnability	Memorability	Understandability	Interruptibility	Appropriateness
34	×	√	×	×	×	×	√	√	×	×	√	×	×	×	×	×
35	×	×	×	√	×	×	×	×	×	×	×	×	√	×	×	×
36	×	×	×	×	×	×	×	×	×	×	×	√	√	×	√	×
37	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
38	×	×	×	×	×	×	×	×	×	×	×	√	×	×	×	×
43	×	×	√	×	×	√	×	×	×	×	×	√	√	×	×	×
44	×	√	×	×	×	×	×	×	×	×	×	×	×	×	×	×
45	×	×	×	√	√	×	×	×	×	×	×	×	×	×	×	×
46	×	×	×	√	√	×	×	×	√	×	×	×	×	×	×	×
47	×	×	×	√	√	×	×	×	×	×	×	×	√	√	×	×
48	×	×	×	×	×	×	×	×	×	×	×	×	×	√	×	×
49	×	×	×	√	√	×	√	×	×	×	×	×	√	×	×	×
51	×	×	×	×	√	×	√	×	×	×	×	√	√	×	×	×
53	×	×	×	×	×	×	×	×	×	√	×	×	×	×	×	×
54	×	×	×	×	√	×	×	×	×	√	×	√	×	×	×	×

Table 12. Comparative validation of proposed framework with existing models. √—Criterion met, ×—Criterion not met.

thus provides a systematic and cost-effective alternative to traditional usability evaluation methods that are often resource-intensive. The framework will present real value to the developers, designers, and researchers by serving as a decision-support tool for evaluating health applications usability. However, its contribution should be viewed within certain boundaries. The first limitation of this framework is that it is based only on the AHP method, which, though good for hierarchical decision-making, pays little attention to alternative establishing a priori degree of importance. Hence, AHP can be combined with any other MCDM methodologies, such as TOPSIS, DEMATEL, or fuzzy AHP, in future research so as to contribute towards validation and generalization of the study. Although participants were selected from a local population and may not fully represent cultural and demographic views, a wider-ranging sample of various cultures and demographics would strengthen the external validity of the outcomes. Though the preliminary empirical validation provides promising results; we acknowledge that a more comprehensive multi user and multi app validation would improve the framework's generalizability. Such research is planned to be done in future. In addition, while the consistency ratios were employed, respondent bias in AHP is also a major concern to be addressed in future research. Lastly, while this research is restricted to smartphone health apps, modifications would have to be made for extending the framework into areas such as wellness, and fitness apps. In a nutshell, the framework provides and initial step toward a structured approach evaluating the usability of health apps, a pathway that will allow for improvement and more comprehensive usability evaluation frameworks in subsequent research.

Data availability

The datasets generated and/or analyzed during the current study are not publicly available due to participant confidentiality and institutional policy but are available from the corresponding author on reasonable request.

References

1. Shahzadi, R. Relationship between smartphone usage and psychological well-being of working women with different socioeconomic background in Punjab-Pakistan. *Pakistan Social Sci. Rev.* **4**(III), 302–314. https://doi.org/10.35484/pssr.2020(4-iii) (2020).
2. Bajwa, R. S., Abdullah, H., Zaremohzzabieh, Z., Jaafar, W. M. W. & Samah, A. A. Smartphone addiction and phubbing behavior among university students: A moderated mediation model by fear of missing out, social comparison, and loneliness. *Front. Psychol.* **13**, 1072551 (2023). https://doi.org/10.3389/fpsyg.2022.1072551 (2023).
3. Jabeen, U., Sarvat, H. & Hashmi, Z. Smartphone addiction and family communication in adults. *Humanit. Social Sci. Reviews*. **9**(3), 1288–1294. https://doi.org/10.18510/hssr.2021.93127 (2021).
4. Cecere, G., Corrocher, N. & Battaglia, R. D. Innovation and competition in the smartphone industry: is there a dominant design? *Telecomm. Policy*. **39**(3–4), 162–175. https://doi.org/10.1016/j.telpol.2014.07.002 (2015).
5. Hisam, A. et al. Usage and types of mobile medical applications amongst medical students of Pakistan and its association with their academic performance. *Pakistan J. Med. Sci.* **35**(2), 403–408. https://doi.org/10.12669/pjms.35.2.672 (2019).
6. Jembai, J. V. J. et al. Mobile health applications: Awareness, attitudes, and practices among medical students in Malaysia. *BMC Med. Educ.* **22**(1), 544 (2022). https://doi.org/10.1186/s12909-022-03603-4 (2022).
7. Liew, M. S., Zhang, J., See, J. & Ong, Y. L. Usability challenges for health and wellness mobile apps: Mixed-methods study among mHealth experts and consumers. *JMIR Mhealth Uhealth*. **7**(1), e12160. https://doi.org/10.2196/12160 (2019).
8. Alessa, T., Hawley, M., Everson-Hock, E. & De Witte, L. Smartphone apps to support self-management of hypertension: review and content analysis. *JMIR Mhealth Uhealth*. **7**(5), e13645. https://doi.org/10.2196/13645 (2019).
9. Aranda-Jan, C. B., Mohutsiwa-Dibe, N. & Loukanova, S. Systematic review on what works, what does not work and why of implementation of mobile health (mHealth) projects in Africa. *BMC Public. Health*. **14**(1), 188. https://doi.org/10.1186/1471-2458-14-188 (2014).
10. Free, C. et al. The effectiveness of mobile-health technologies to improve healthcare service delivery processes: A systematic review and meta-analysis. *PLoS Med.* **10**(1), e1001363. https://doi.org/10.1371/journal.pmed.1001363 (2013).
11. McNair, J. B. Theoretical basis of health IT evaluation. *PubMed* **222**, 39–52 (2016). https://pubmed.ncbi.nlm.nih.gov/27198091
12. Bhutkar, G., Konkani, A., Katre, D. & Ray, G. A review: healthcare usability evaluation methods. *Biomedical Instrum. Technol.* **47**(s2), 45–53. https://doi.org/10.2345/0899-8205-47.s2.45 (2013).
13. Monfort, G. M., Paluzié, G., Díaz-Gegúndez, J. M. & Chabrera, C. Usability of a mobile application for health professionals in home care services: A user-centered approach. *Sci. Rep.* **13**(1), 2607 (2023). https://doi.org/10.1038/s41598-023-29640-7 (2023).
14. Muhammad, A. et al. Evaluating usability of academic websites through a fuzzy analytical hierarchical process. *Sustainability* **13**(4), 2040 (2021). https://doi.org/10.3390/su13042040 (2021).
15. Sheehan, B., Lee, Y. J., Rodriguez, M., Tiase, V. L. & Schnall, R. A comparison of usability factors of four mobile devices for accessing healthcare information by adolescents. *Appl. Clin. Inf.* **3**(4), 356–366. https://doi.org/10.4338/aci-2012-06-ra-0021 (2012).
16. Ribeiro, V. S., Martins, A. I., Queirós, A., Silva, A. & Rocha, N. P. Usability evaluation of a healthcare application based on IPTV. *Procedia Comput. Sci.* **64**, 635–642. https://doi.org/10.1016/j.procs.2015.08.577 (2015).
17. Stoyanov, S. et al. Mobile app rating scale: A new tool for assessing the quality of health mobile apps. *JMIR Mhealth Uhealth*. **3**(1), e27. https://doi.org/10.2196/mhealth.3422 (2015).
18. Househ, M., Shubair, M. M., Yunus, F., Jamal, A. & Aldossari, B. The use of an adapted health IT usability evaluation model (Health-ITUEM) for evaluating consumer-reported ratings of diabetes mHealth applications: implications for diabetes care and management. *Acta Informatica Med.* **23**(5), 290–295. https://doi.org/10.5455/aim.2015.23.290-295 (2015).
19. Schnall, R. et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). *J. Biomed. Inform.* **60**, 243–251. https://doi.org/10.1016/j.jbi.2016.02.002 (2016).
20. Zhou, L., Bao, J., Setiawan, I. M. A., Saptono, A. & Parmanto, B. The mHealth app usability questionnaire (MAUQ): development and validation study. *JMIR Mhealth Uhealth*. **7**(4), e11500. https://doi.org/10.2196/11500 (2019).
21. Ahmed, S., Donyaee, M., Kline, R. B. & Padda, H. K. Usability measurement and metrics: A consolidated model. *Software Qual. J.* **14**(2), 159–178. https://doi.org/10.1007/s11219-006-7600-8 (2006).
22. Alonso-Ríos, D., Vázquez-García, A., Mosqueira-Rey, E. & Moret-Bonillo, V. Usability: A critical analysis and a taxonomy. *Int. J. Hum Comput Interact.* **26**(1), 53–74. https://doi.org/10.1080/10447310903025552 (2009).
23. Dubey, S., Gulati, A. & Rana, A. Integrated model for software usability (2012). Available at: https://www.semanticscholar.org/paper/Integrated-Model-for-Software-Usability-Dubey-Gulati/e4a7168cd5d80aa7d28966d72cc46457226f14d3.
24. Harrison, R., Flood, D. & Duce, D. A. Usability of mobile applications: literature review and rationale for a new usability model. *J. Interact. Sci.* **1**(1), 1. https://doi.org/10.1186/2194-0827-1-1 (2013).

25. Baharuddin, R., Singh, D. & Razali, R. Usability dimensions for mobile applications: A review. *Res. J. Appl. Sci. Eng. Technol.* **11**(9), 2225–2231. https://doi.org/10.19026/rjaset.5.4776 (2013).
26. Hussain, A., Hashim, N. L., Nordin, N. & Tahir, H. M. A metric-based evaluation model for applications on mobile phones. *J. ICT*. **12**. https://doi.org/10.32890/jict.12.2013.8137 (2013).
27. Nugroho,A., Santosa, P. I. & Hartanto, R. Usability evaluation methods of mobile applications: A systematic literature review. in *2022 International Symposium on Information Technology and Digital Innovation (ISITDI)* 92–95. (IEEE, 2022). https://doi.org/10.1109/ISITDI55734.
28. Razak, A. A. & Ahmad, Z. The importance of usability for mobile applications. *Int. J. Comput. Appl.* **97**(9), 17–20. https://doi.org/10.5120/16924-3098 (2014).
29. Sinha, A. & Kaur, A. An overview of usability engineering and its methods. *Int. J. Comput. Appl.* **138**(3), 36–40. https://doi.org/10.5120/12027-8862 (2016).
30. McKeown, C. The impact of app usability on user engagement in healthcare applications. *Front. Public. Health*. **8**, 370. https://doi.org/10.3389/fpubh.2020.00370 (2020).
31. Abd El-Hafeez, S. Usability evaluation of mobile health applications for diabetes management. *Egypt. J. Otolaryngol.* **37**(1), 1–7. https://doi.org/10.1186/s43163-021-00157-6 (2021).
32. Gupta, D., Ahlawat, A. & Sagar, K. A critical analysis of a hierarchy-based usability model. in *2014 International Conference on Contemporary Computing and Informatics (IC3I)* 255–260 (2014). https://doi.org/10.1109/IC3I.2014.7019810.
33. Gupta, D. & Ahlawat, A. Usability determination using multistage fuzzy system. *Procedia Comput. Sci.* **78**, 263–270. https://doi.org/10.1016/j.procs.2016.02.042 (2016).
34. Gupta, D., Ahlawat, A. & Sagar, K. Usability prediction & ranking of SDLC models using fuzzy hierarchical usability model. *Open. Eng. (Warsaw)*. **7**(1), 161–168. https://doi.org/10.1515/eng-2017-0021 (2017).
35. Chen, T. C. & Chiu, M. C. Smart technologies for assisting the life quality of persons in a mobile environment: A review. *J. Ambient Intell. Humaniz. Comput.* **9**(2), 319–327. https://doi.org/10.1007/s12652-016-0396-x (2016).
36. Saleh, A. M., Ismail, R. & Fabil, N. Evaluating usability for mobile applications. *Proc. Int. Conf. Intell. Syst. Comput. Appl.* 71–77. https://doi.org/10.1145/3178212.3178232 (2017).
37. De La Díez, T., Alonso, I., Cruz, S. G., Franco, M. & E. M., & Measuring QOE of a teleconsultation app in mental health using a Pentagram model. *J. Med. Syst.* **43**(7). https://doi.org/10.1007/s10916-019-1342-1 (2019).
38. Kasali, F. et al. An enhanced usability model for mobile health applications. *Int. J. Comput. Sci. Inform. Secur. (IJCSIS)*. **17**(2), 20–29. https://doi.org/10.1063/1.4960948 (2019).
39. Page, M. J. et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. bmj **372**, n71 (2021).
40. Ibrahim, M. S. Validating service quality (SERVQUAL) in healthcare: measuring patient satisfaction using their perceptions in Jordan. *J. Inform. Knowl. Manage.* **19**(01), 2040021. https://doi.org/10.1142/s0219649220040213 (2020).
41. Muhammad, A. H. et al. A hierarchical model to evaluate the quality of web-based e-learning systems. *Sustainability* **12**(10), 4071. https://doi.org/10.3390/su12104071 (2020).
42. Bass, L. & John, B. E. Linking usability to software architecture patterns through general scenarios. *J. Syst. Softw.* **66**(3), 187–197. https://doi.org/10.1016/S0164-1212(02)00076-6 (2013).
43. Brown, W., Yen, P. Y., Rojas, M. & Schnall, R. Assessment of the health IT usability evaluation model (Health-ITUEM) for evaluating mobile health (mHealth) technology. *J. Biomed. Inform.* **46**(6), 1080–1087. https://doi.org/10.1016/j.jbi.2013.08.001 (2013).
44. Singh, K. & Kumar, P. A model for website quality evaluation—a practical approach. *Int. J. Res. Eng. Technol.* **2**(3), 61–68 (2014).
45. Shitkova, M., Holler, J., Heide, T., Clever, N. & Becker, J. Towards usability guidelines for mobile websites and applications. in *Proceedings of the 12th International Conference on Wirtschaftsinformatik (WI 2015)* 1603–1617 (2015).
46. Deniz-Garcia et al. WARIFA Consortium. Quality, usability, and effectiveness of mHealth apps and the role of artificial intelligence: Current scenario and challenges. *J. Med. Internet Res.* **25**, e44030. https://doi.org/10.2196/44030 (2023).
47. Rezaee, R., Khashayar, M., Saeedinezhad, S., Nasiri, M. & Zare, S. Critical criteria and countermeasures for mobile health developers to ensure mobile health privacy and security: mixed methods study. *JMIR mHealth uHealth*. **11**, e39055. https://doi.org/10.2196/39055 (2023).
48. Zayim, N., Yıldız, H. & Yüce, Y. K. Estimating cognitive load in a mobile personal health record application: A cognitive task analysis approach. *Healthc. Inf. Res.* **29**(4), 367–376. https://doi.org/10.4258/hir.2023.29.4.367 (2023).
49. Shin, J. H. et al. Quality and accessibility of home assessment mHealth apps for community living: systematic review. *JMIR mHealth uHealth*. **12**(1), e52996. https://doi.org/10.2196/52996 (2024).
50. Galavi, Z., Montazeri, M. & Khajouei, R. Which criteria are important in usability evaluation of mHealth applications: an umbrella review. *BMC Med. Inf. Decis. Mak.* **24**(1), 365. https://doi.org/10.1186/s12911-024-02738-2 (2024).
51. Giebel,G. D. et al. Problems and barriers related to the use of mHealth apps from the perspective of patients: focus group and interview study. *J. Med. Internet Res.* **26**, e49982. https://doi.org/10.2196/49982 (2024).
52. Khamaj, A. & Ali, A. M. Examining the usability and accessibility challenges in mobile health applications for older adults. *Alexandria Eng. J.* **102**, 179–191. https://doi.org/10.1016/j.aej.2024.06.002 (2024).
53. Stapelfeldt, P. M., Müller, S. A. & Kerkemeyer, L. Assessing the accessibility and quality of mobile health applications for the treatment of obesity in the German healthcare market. *Front. Health Serv.* **4**, 1393714. https://doi.org/10.3389/frhs.2024.1393714 (2024).
54. Seffah, A., Donyaee, M., Kline, R. B. & Padda, H. K. Usability measurement and metrics: A consolidated model. *Software Qual. J.* **14**(2), 159–178. https://doi.org/10.1007/s11219-006-7600-8 (2016).
55. Shen, Y., et al. Evaluating the usability of mHealth apps: an evaluation model based on task analysis methods and eye movement data. *Healthcare* **12**(13), 1310. https://doi.org/10.3390/healthcare12131310 (2024).
56. Kim, G., Hwang, D., Park, J., Kim, H. K. & Hwang, E. S. How to design and evaluate mHealth apps? A case study of a mobile personal health record app. *Electronics* **13**(1), 213. https://doi.org/10.3390/electronics13010213 (2024).
57. Shareef, S. & Khan, M. N. A. Evaluation of usability dimensions of smartphone applications. *Int. J. Adv. Comput. Sci. Appl.* **10**(9). https://doi.org/10.14569/IJACSA.2019.0100956 (2019).
58. Dahri, A. S., Al-Athwari, A. & Hussain, A. Usability evaluation of mobile health application from AI perspective in rural areas of Pakistan. *International Association Online Engineering* (2019). https://www.learntechlib.org/p/216620/.
59. Chao, S. M., Pan, C. K., Wang, M. L., Fang, Y. W. & Chen, S. F. Functionality and usability of mHealth apps in patients with peritoneal dialysis: A systematic review. *Healthcare* **12**(5), 593. https://doi.org/10.3390/healthcare12050593 (2024).

Acknowledgements

Author contributions

Conceptualization, A.S and Q.S.; Formal analysis, A.S., Q.S., and B.S.; Funding acquisition, B.S., S.A.A.; Investigation, A.S. and Q.S.; Methodology, A.S. and B.S.; Project administration, A.S and A.A.S.; Resources, A.S and

B.S.; Software, Q.S. J.M.H.; Supervision, A.S.; Validation, A. S. and A.A.S.; Writing—original draft, A.S. and Q.S.; Writing—review & editing, A.S. All authors have read and agreed to the published version of the manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Ethical approval
The research was conducted in compliance with the ethical standards set forth by the institutional research committee. The Department of Information Technology Review Board, University of Gujrat (UOG), proceeded with the approval of the study after briefing the objectives and methodology of the research. Though no reference number is issued by committee, official approval was taken before data collection.

Informed consent
Informed consent was obtained through explanation and voluntary participation from all the subjects.

Additional information
Correspondence and requests for materials should be addressed to A.S. or S.A.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.