Evaluating Usability of Artificial Intelligence (AI) Based M-Health Applications Through Cognitive Walkthrough

Majed A. Alshamari
Department of Information Systems, College of Computer Sciences and Information Technology
King Faisal University
Hofuf- Kingdom of Saudi Arabia
Email: smajed@kfu.edu.sa

***Abstract*—Artificial Intelligence (AI) technology has been adopted and employed in healthcare section to develop applications for providing various healthcare services. However, the effectiveness of these apps depends on their usability, which is a critical factor in their success. One approach to evaluating the usability of these apps is through a cognitive walkthrough. In our study, we aimed to evaluate the usability of AI-based features in 3 mHealth applications, including Ada, Babylon, and Ornament. We conducted a cognitive walkthrough by providing a list of tasks in order to carry out the process. After each task completion, evaluators were presented with questionnaire to assess the application's usability attributes. A total of 27 distinct problems were identified. The highest number of problems were related to health information and symptom checking features. The reported severity of identified issues in Ada, Babylon and Ornament are 7.4, 8.0 and 4.2 respectively. Some of the identified usability problems are irrelevant health information, limited disease enlistment, no search option, tiresome navigation, unsatisfactory results, and delayed responses. These issues impact effectiveness, and efficiency of AI models, and ultimately user satisfaction, thus, highlighting the need to improve AI based mHealth applications' functionality and design. Further, the evaluators provide recommendations on these identified problems.**

Keywords- AI based mobile applications Introduction; Artificial Intelligence; Cognitive Walkthrough; mHealth application; Usability Evaluation.

I. INTRODUCTION

Nowadays, everyone irrespective of their age can access smart devices including smart televisions, tablets, phones, and other internet-connected devices because of digital media. Every day, thousands of apps, with a wide range of functions, are added to dedicated (iOS and Android) app stores (Apple App Store and Google Play Store), and this number is constantly growing [1]. In the past decade, the health industry has seen phenomenal growth and pushed healthcare delivery to new levels. Therefore, m-health is becoming an essential sector for delivering and spreading health in our society as a whole [2]. The mobile health (mHealth) app market is anticipated to develop at a compound annual growth rate of 17.7% throughout the forecast period, according to the most recent report by Grand View Research, Inc. [3], reaching US $149.3 billion by 2028. Users' interest in mHealth applications has grown significantly over the past decades, making healthcare a significant category in these mobile app catalogs. According to research, up to 34% of smartphone owners have at least one health app loaded on their mobile devices [4]. Also, the usage of artificial intelligence (AI) in mobile apps for healthcare systems, finance, and entertainment has increased primarily due to smartphones and tablets [5]. In this era of rapid technological development, people from all walks of life utilize artificially intelligent mobile applications (apps) on a global scale. Conclusively, AI is progressively playing a larger role in people's daily lives [6] [7].

Using AI-based applications in healthcare is of particular importance to patients; therefore, it is important that their use does not harm them, but rather benefits them. Thus, AI systems should provide patient satisfaction across multiple healthcare environments and be effective and efficient [8]. Hence, by examining the usability of mobile health apps, we can uncover issues, help redesign systems, spend less time and money, and improve user acceptance [9]. Effectiveness, efficiency, learnability, ease of use, and user satisfaction considered some of the most common usability attributes when defining the usability of system. An often-used analytical method of usability evaluation is Cognitive Walkthrough (CW) [10].

Lewis and Wharton developed cognitive walkthrough for evaluating the usability of interfaces using theory-based evaluation [11]. It is employed to identify problems and generate proposals about their causes. Learning through CWs is aimed at simplifying learning, especially through exploratory learning. In medical equipment evaluation, CW is used to evaluate depression screening models [12], Nurse information systems [13], Diabetes management systems [14] and other healthcare systems. It is advantageous to use CW in healthcare since it can be used to identify important usability problems quite easily, quickly, and cheaply when real usability testing is not feasible [15]. This paper aims to contribute in the identification of critical issues in mobile health applications that affect the usability attributes and that impact the adoption and effective use of AI applications in healthcare.

The rest of the paper is organized as follows: In Section 2 we discuss the previous publications which evaluates the AI based system. Section 3 is the detailed methodology of our evaluation process. In Section 4, we present the qualitative

 ISBN: 978-1-68558-096-4

and quantitative findings from evaluation, followed by Section 5, which is the conclusion of our research work.

II. LITERATURE REVIEW

There are nearly 165,000 mobile health (mHealth) apps available in the Apple iTunes and Android app stores in the United States [16], which are used by two-thirds (66%) of Americans [17] Survey Finds 66% of Americans eager to leverage digital tools to manage personal health. Many mHealth apps have designed with little input from end users, and they continue to expand despite limited evidence of user engagement [18][16] Applications are routinely created with low-quality designs and with insufficient consideration of end-user demands. Such applications may be challenging to use, misunderstood, or underutilized, and may ultimately fall short of their objectives [19] [20] Apps must therefore guarantee quality and offer the necessary functionality. This emphasizes how crucial it is to assess usability of mHealth applications. Also, medical technology focuses more on usability than user experience [21]. This section provides a summary several research studies that have been performed to calculate usability of mHealth and mobile systems through the usability evaluation methods.

A mobile accounting software has been developed by [22] using Rapid Application Participatory Development (RAPD) method. Further, they evaluated the usability of accounting software using a cognitive walkthrough performed by 16 participants. Their objective was to identify the effect of COVID-19 on the usability of new software. Their research identifies new factors that influence application usability, including user experience, remote work, security, privacy, internet speed and Artificial Intelligence. [23] proposed a conceptual model names " GenDAI", which is an AI assisted laboratory Diagnostic Solution for the genomic applications. In GenDAI, the AI-driven AI2VIS4BigData abstract architecture for metagenomics is combined with the CRISP4BigData-based model for the gene expression diagnostics. An evaluation of this conceptual model was conducted by partnering with small medical laboratory of ImmBioMed GmbH & Co. KG in Heidelberg, Germany. Platflow was developed to perform analysis on the raw data and was evaluated through cognitive walkthroughs. Preliminary study results indicate that there are several areas in laboratory workflow that could be automated.

A depression-screening model has been evaluated by N. Fasihah Jamaludin [12] to examine how effective it was at addressing adolescent motivation during gamification-based depression screening, through a cognitive walkthrough. The evaluation was conducted by five respondents with expertise in adolescent counseling and human-computer interaction. According to the analysis, all respondents gave positive feedback on the sets of tasks provided. These results confirmed the model's usability in detecting depression through the cognitive walkthrough. They concluded that the model might be used as a blueprint for creating a real depression screening system. A. S. Dahri [24] investigates how well the mobile health application "mHealth" is used by patients by accessing their satisfaction with their tasks. 15 patients completed tasks on task success rate, mistakes, efficiency (time spent), and satisfaction using System Usability Scale (SUS) and the International Organization for Standardization (ISO) 9241-11 standard criteria. Effectiveness was measured in terms of how many users have successfully completed task, while efficiency was measured in terms of time taken by each task to get completed. The findings of this study showed that finding a medical professional was the most challenging step for users and registering was the easiest task. The usability scores in this study are also influenced by educational level and mobile expertise.

In addition, M. N. Islam et al [25] developed a mobile-based solution "Muktomon" which means open one's mind, for providing mental health support to the people of Bangladesh. This application provides virtual therapy through videos and audio, a chatbot service for mental health assistance. They evaluated the usability by conducting a system usability survey and pots questionnaire from 37 participants. Their application got SUS score of 79.875%, which means acceptable system in terms of usability. In the context of the COVID-19 pandemic, the application proved useful and usable for improving mental health. N. A. Zaini et al [26] designed a low-fidelity APi prototype of a game to provide fire safety education to children. They used interactive learning as a key to promoting preschool children's knowledge of fire safety basics. APi prototypes were designed based on the user requirements of preschool children focused on cognitive, psychomotor, and behavioral aspects. A small group of 6 people including professional designers and developers. They conducted the cognitive walkthrough evaluation to evaluate the usability and learnability of the APi interface. Participants evaluated prototype on color, background theme, font, and consistency in design etc. From the findings of the cognitive walkthrough, they designed the high-fidelity prototype of APi interface for fire safety education.

The cognitive evaluation method has been used by [13] to evaluate the usability of a user interface of a Nursing Information System (NIS). The system was evaluated by five evaluators according to given scenarios and the problems identified were assigned to usability attributes. Evaluators also determined the severity of each identified problem. M. Georgsson [14] proposed a technique called user-centered cognitive walkthrough, to address the flaws of the original cognitive walkthrough. They also perform a preliminary validation using the think-aloud protocol to gauge the method's efficacy, and user acceptability in a study with diabetes patient which are users of a mHealth self-management application. They divided the Diabetes patients into 2 groups, one as UC-CW and the other as think-aloud (TA) groups at the University of Utah Health in the United States. They identified 26 different usability problems (heuristics violation) with UC-CW and 20 usability problems using the think-aloud method, in Recall and Recognition, Consistency and Standards, and Match between System and Real world. The study reported that UC-CW is an effective method for finding usability problems than TA becuase patients' diseases required customized qualities that could not

be determined by TA. [27] proposed a study which compares two expert-based evaluation methods (Heuristic Evaluation & Cognitive Walkthrough) in a nursing module of a Hospital Information System (HIS). Five evaluators use the system and identifies 104 problems with the heuristic method and 24 usability problems with cognitive walkthrough method. They reported a significant change between severity of recognized usability problems and the number by these methods. As a result of the cognitive walkthrough, issues of learnability, efficiency, and memorability have been identified, whereas as a result of heuristic evaluation, issues of effectiveness, satisfaction, and errors have been identified. methods.

A usability test involving 18 healthcare professionals has been conducted by [28] to evaluate the effectiveness of an electronic health record (EHR) display prototype for emergency medicine. Participants were asked to complete 2 questionnaires for rating usability, usefulness, and effectiveness. Study findings emphasize the need for user-centered design when developing EHR systems for emergency medicine. [2] developed and evaluated an e-health prototype with five health professionals including information system experts and six health consumers. The Post-Study System Usability Scale (PSSUS) was modified and adapted by the authors, who developed the post-Study e-Health Usability Questionnaire (PSHUQ), which consists of 19 items describing five characteristics of system usability: easy learning, functional adequacy, rapid acquisition of usability experts and several different user groups, rapid completion of work, and high-quality online documentation. A number of users have provided feedback on the system, suggesting improvements and recommendations for future enhancements. The most common suggestion was that consumers' personal information should be kept confidential and secure. Moreover, optimization of resource utilization and quality are desired, along with meeting consumer demands.

TABLE I. LITERATURE REVIEW

Reference #	Objective	No. of Participants	Evaluation method	Evaluated App
[22]	Developed accounting software using RAPD Identified new usability factors	16	Cognitive walkthrough, interviews	Accounting mobile app
[23]	Genome diagnostic tool for laboratory use Developed an application for analyzing results	-	Cognitive walkthrough, on-site visits, interviews	Platflow Tool
[12]	Usability evaluation of the depression screening model	5	Cognitive walkthrough	Gamification Model
[2]	Developed and evaluated an e-health prototype	11	Post-Study System Usability Scale	Heal-me.co
[25]	Development and Usability evaluation of Mental Health care app	37	SUS, Interviews	Muktomon
[13]	Usability evaluation of an information system	5	Cognitive Walkthrough	Nursing Information System
[27]	A comparative study to evaluate usability and learnability of a system with different methods	5	Cognitive Walkthrough +Heuristic Evaluation	Health Information System
[24]	Investigates UE of the Mobile Health application by patients' task performance evaluation and satisfaction	15	SUS, ISO 9241-11	mHealth
[26]	Develop and evaluate the	6	Cognitive Walkthrough	Fire Safety Education

	usability of the prototype for preschool children			
[14]	A case study to evaluate the usability of a mobile-based healthcare system	12	Cognitive Walkthrough	Diabetes self-management application
[28]	Evaluate the usability of Emergency Medicine HER Prototype	18	Questionnaires	Electronic Health Record Display
Proposed Work	Evaluate usability and learnability of AI applications in healthcare	15	Cognitive Walkthrough	Ornament, Ada Health, Babylon Health

The proposed work evaluates the effectiveness, efficiency, ease of use and satisfaction of 3 AI applications (Ada, Babylon and Ornament) in healthcare using cognitive walkthroughs. A significant contribution of this study will be the identification of critical issues in these applications that affect the usability attributes and that impact the adoption and effective use of AI applications in healthcare, and they will contribute to knowledge of usability and learnability. In order to develop more user-friendly AI applications that are easy to use, learn, and adopt in healthcare, the study aims to provide useful recommendations from experts.

III. MATERIALS AND METHODS

In this section, we present the methodology that was used to evaluate the usability of AI-based features in three mobile health applications: Ada Health, Ornament, and Babylon Health. To assess the effectiveness, efficiency, satisfaction, and ease of use of these applications, we used the cognitive walkthrough evaluation method. We chose these three applications because they are well-established and widely used in the healthcare industry and each provides unique AI-based features. Further, we recruited a team to conduct the evaluation. We then conducted a survey that included both qualitative data, that are problems and suggestions, and yes/no responses. The complete workflow diagram is shown in Figure 1.

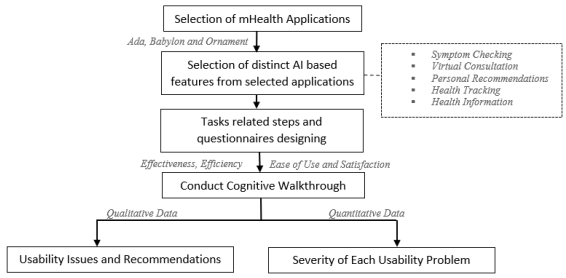

Figure 1. Complete Workflow Diagram of conducting CW to Evaluate usability 3 Applications.

A. *Cognitive Walkthrough Evaluation Method*

A group-based expert approach called CW was created by Polson and Lewis and is based on theories of the cognitive exploratory learning or users' capacities to understand through their activities [29], [30]. Experts identify system flaws using CW by simulating users' problem-solving skills. For systems that need cognitive support or feedback when users lack basic knowledge, this point is crucial [14]. It involves evaluators simulating users' cognitive processes when thinking about the actions they took to accomplish tasks that based on their background knowledge. It is important for evaluators to put themselves in the user's shoes in order to produce good results [15]. The assessor evaluates the user interface and assesses how simple each step is for new [10].

Firstly, we identified the applications to evaluate and assess usability. Then we identified the tasks and users, who are actually evaluators, and determine the sequence of actions that user will take to carry out the task. After it, evaluators conducted the walkthrough and answer the following four questions after each step of task. These questions are aids to stimulating the user's cognitive process.

1. Will the user be trying to achieve the right effect?
2. Will the user discover that the correct action is available?
3. Will the user associate the correct action with the desired effect?
4. If the correct action is performed, will the user see that progress is being made?

In response to these questions, the user will answer a YES or NO along with reasons why their action was successful or unsuccessful. To further evaluate the main usability attributes which are effectiveness (accuracy of predictions), efficiency (time taken by AI model to give results), satisfaction of user and ease of use, users will be posed to multiple questions after the completion of each task. In response of these questions, users will describe the usability problems found and their recommendations.

IV. AI BASED M-HEALTH APPLICATIONS

We have selected three mHealth applications that uses Artificial Intelligence. These applications include, Ada, Babylon, and Ornament Health application. The criteria of selecting applications is based on use of AI model, availability on android and iOS, and diversity of applications. All the 3 applications use AI to predict disease from symptoms, personal recommendations etc., and available on Android and iOS.

Ada Health [31] is a mHealth application that uses AI algorithms to provide users with personalized symptom-checking and health information (Figure .2). In this application, users can identify potential health concerns and make informed decisions about seeking medical treatment. Users can input information about their symptoms, medical history, and other relevant health information into the application. A personalized report is generated based on this information, which suggests possible causes of the symptoms and recommends seeking medical care when necessary.

Babylon Health [32] is a another mHealth application that offers telehealth services to users. Home Screen of this application is shown in Figure 3. A virtual consultation can be scheduled with a healthcare professional, such as a doctor, nurse, or therapist, through the application. Medical records can also be accessed and prescriptions can be requested using the application. Specifically designed for non-emergency medical issues, Babylon Health provides users with convenient access to healthcare services. Using AI algorithms, the application guides patients to the most appropriate healthcare provider based on their symptoms and medical history. Ornament Health [33] is also a mHealth application that focuses on wellness and helping users achieve their health goals (Figure. 4). Users can track their physical activity, nutrition and other health metrics. Application's Ai models generated personalized recommendations for users using this collected information to help them improve their well-being and health. Users can also access wellness coaches through Ornament Health, who can answer questions and provide guidance on living a healthy lifestyle.

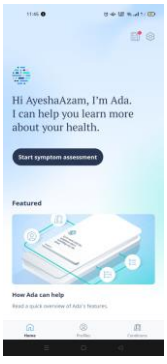

Figure 2. Ada Health Mobile App Home Screen

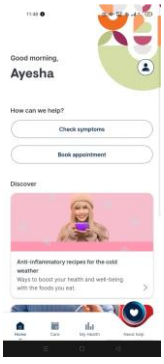

Figure 3. Babylon Health Mobile App Home Screen

Fugire 4. Ornament Health Mobile App Home Screen

A. Evaluators

For a cognitive walkthrough evaluation, a minimum of three evaluators is recommended to ensure that a variety of perspectives are represented [34]. It was found in a study by [35] that only three subjects are needed to uncover 65% of the problems, five are needed to uncover 80%, and nine are needed to uncover 95%. Our study was conducted by 15 evaluators. The evaluators include Ph.D students of which, 5 were developers, 5 were UX designers, 3 were the HCI experts, and 2 were the Ph.D. Scholars in Computer science. All of them have prior experience with mHealth applications. We aimed to enrich the results by bringing in different perspectives from individuals with different expertise, despite the higher cost associated with using more evaluators. This enabled us to evaluate the usability of mHealth applications in a comprehensive manner, which can lead to the design of more effective and user-friendly products.

V. DATA COLLECTION AND ANALYSIS METHOD

Each task was performed independently by evaluators on three applications in order to carry out the evaluation. If a problem arises afterward to achieve a task from a users' perspective, the evaluators could report back [13]. In addition to acting as an observer, the researcher, along with the evaluators, took notes on the data collection forms regarding questions, comments, and ambiguities relating to the evaluation process. To assess usability attributes, we present evaluators with a questionnaire following completion of each task. This questionnaire includes attributes are effectiveness, efficiency, ease of use, and satisfaction. The definition of these usability attributes are as follows:

1. Effectiveness: It refers to how well the AI model performs in accurately predicting a specific health outcome or disease diagnosis.
2. Efficiency: It refers to how quickly the application and user can perform a specific task or process.
3. Ease of Use: It refers to how easily and intuitively users can interact with the application to perform specific tasks or access relevant information.
4. Satisfaction: It refers to overall satisfaction that user get after using and experiencing the application.

The evaluators reviewed the details of the usability issues and made corrections or additions as needed after the evaluation process was complete. All the identified problems were added to the list of problems, while the repeated problems were removed from the list. To calculate the severity of each problem we used the (1) [36],

$$\text{Severity} = \text{Frequency} * \text{Impact} \quad (1)$$

Where, frequency is the number of times a problem is occurring and impact is severity of consequences of the problem. Symptom Checking given impact value 5, virtual consultation, personal recommendations, health tracking and health information were given 4, 3, 2, and 1 respectively.

 ISBN: 978-1-68558-096-4

VI. RESULTS AND DISCUSSION

An evaluation of three AI-based mHealth applications, including Ada health, Babylon health, and Ornament health, was conducted using Cognitive walkthrough method. The five tasks were performed by 15 evaluators to identify the problems with the application's usability. To further assess the usability of AI based tasks, we designed and presented a questionnaire to evaluators following the cognitive walkthrough of each task. In total, 56 problems were detected, of which 40 unique ones remained (14 from Ada health, 14 from Babylon and 12 from Ornament health application) after eliminating duplicates and combining the problems. Table 2 summarizes the unique problems identified and recommendations based on usability attributes.

TABLE II. A USABILITY ATTRIBUTE-BASED IDENTIFIED PROBLEMS AND RECOMMENDATIONS

Usability Attributes	Application	Identified Problem	Recommendations	Task
Effectiveness	Ada	Not relevant articles	Add articles relevant to history of patient.	Health Information
		Sometimes system provides a list of irrelevant diseases.	Re train the AI models	Symptom Checking
		AI model only enlist few possible diseases and nothing else.	Provide with medical treatment options as well.	Symptom Checking
	Babylon	No personal recommendation according to expectation	Retrain the model with updated data and new algorithms Or use collaborative filtering techniques	Personal Recommendations
		No option to search for articles of user's choice	Add search option so user can search for relevant information	Health Information
		Asks for driver license and other unnecessary details before booking consultation session.	Do not ask user for passport or driving license information.	Virtual Consultation
		AI model only enlist few possible diseases and nothing else.	Provide with medical treatment options as well.	Symptom Checking
	Ornament	limited options (disease) were given on "Ask Doctor page"	Add more variety of diseases to choose from	Health Tracking
		Presented data and statistics is difficult to understand.	Use user friendly language and visualization methods.	Health Information
Efficiency	Ada	No search option make user took a lot of time in order to search and get desired information.	Add search functionality	Health Information
	Babylon	No search option make user took a lot of time in order to search and get desired information.	Add search functionality	Health Information
		Duplicate buttons	Remove duplicated buttons of book appointment and symptom check	Virtual Consultation, Symptom Checking
	Ornament	Application give response to almost every touch after some time	Optimize the code to improve speed and responsiveness of application	Health information, Health Tracking, Personal recommendations
Ease of Use	Ada	Menus and options were not organized in logical manner	Restructure and group the menus and options in a more intuitive and user-friendly manner	Symptom checking
		It was not easy to find the information that we are looking for.	Categorize the data, Add search functionality on home page	Health Information, Symptom checking

		No "Go back" to home button	Add "Go back" button on every screen	Health Information
	Babylon	Presentation and design of application was not pleasant.	Use more pleasing color scheme and design elements	Symptom Checking, Health Information, Health Tracking, Virtual Consultation
		Navigation is tiresome.	Provide clear and concise labels for navigation. Optimize layout design for easy navigation.	Symptom Checking, Health Information, Health Tracking
	Ornament	Some evaluator reported that UI of application is not pleasing.	Use visually appealing color scheme, typography, and layout	Health Tracking
		Application gets stuck on Insight's page sometimes	Optimize the page loading and processing times by reducing unnecessary data from page	Health Tracking
		Restriction on Must select 3 topics to get insights on.	Remove this restriction	Health Information
		Navigating Back does not work on some pages	Make it work on every page.	Health Information, Health Tracking
Satisfaction	Ada	Sometimes results are not what user expected.	Improve the accuracy of the symptom checker algorithm by incorporating more comprehensive and up-to-date knowledge	Symptom checking
		Did not feel informed about health after using it	Provide comprehensive and personalized health information.	Health Information, Personal recommendations
		Some icons are different from their functions.	Use icons that have clear meanings for users.	Health Information, Personal recommendations
	Babylon	Very few articles to read	Add more user health history related articles.	Health Information
	Ornament	Due to time lagging, the most of the users are not very satisfied with application.	Work on improving speed and responsiveness.	Health Information, Personal recommendations, Health tracking

TABLE III. AVERAGE SEVERITY OF IDENTIFIED PROBLEMS AND AVERAGE TIME TAKEN BY EACH TASK

Tasks	**Severity in ADA**	**Severity in Babylon**	**Severity in Ornament**	**Average Time Taken in Ada (min)**	**Average Time Taken in Babylon (min)**	**Average Time Taken in Ornament (min)**
Symptom Checking	25	20	0	03:35	02: 58	00:00
Virtual Consultation	0	8	0	00:00	00:00	00:00
Personal Recommendations	6	3	6	03:50	04: 38	05:34
Health Tracking	0	4	10	00:00	02:10	04:23
Health Information	6	5	5	02: 35	01:25	03:59
Average	**7.4**	**8.0**	**4.2**	**03:20**	**02:48**	**04:39**

One of the problems that evaluators reported frequently is inaccurate and unexpected results of symptom checker model of applications and it is related to effectiveness of AI models and applications. Other research studies have found that health applications often suffer from poor accuracy, which can undermine their effectiveness [37], [38]. Patients are the most significant recipients and users of AI based mobile applications, thus ensuring its use in healthcare does not harm but rather benefits them should be a priority [39]. Hence, AI systems should be efficient and effective to provide user satisfaction [8]. AI in mobile health applications is a helpful tool [40]. AI has the potential to engage their users and develop significant and healthy

connections with them over time. Using advance machine and deep learning methods and large amount of data, the effectiveness, accuracy of algorithms can improve [41].

Another reoccurring problem that evaluators reported is no search bar (Figure. 5) when performing health information task, thus, affecting the efficiency usability attribute. Usability issues related to efficiency have also been identified in the literature [42], [43] reported frustration in users due to long waits. Hence, it is the need for applications to be faster and more responsive. Similarly, in another it was found to be frustrating and time-consuming for users to manually browse through a lot of information on health information pages without a search bar on interface [44]. A search bar added on page of an application makes it more efficient and enhances the user experience. A study by the Nielsen Norman Group [45] has shown that the presence of a search bar improves the efficiency of an application by allowing users to quickly find what they are looking for. Participants were asked to perform a series of tasks on a website, some with a search bar and some without. The results showed that participants were able to complete tasks faster when a search bar was present hence more efficient. To get health information, ornament restrict their user to select minimum of 3 topics to get insights on as shown in Figure. 6, ultimately affecting the ease-of-use usability attribute of application.

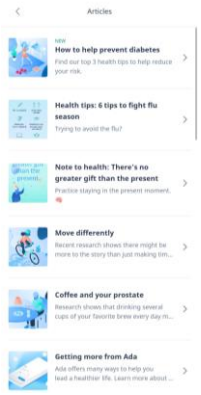

Figure 5. No search Bar on the Health Information Page

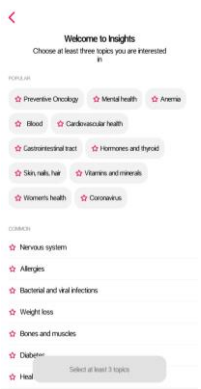

Figure 6. Restriction to select Minimum of 3 items to get insights on

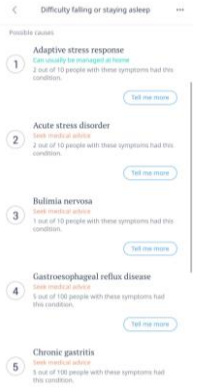

Figure 7. List of Possible diseases

Participants also reported that navigation was tiresome in Babylon. And in Ada Health app, Elements were not properly organized, making it difficult for them to look for the information they require. Go back button was also missing on some screens. However, AI applications need system behavior to be presented in a clear manner. Furthermore, despite dynamic system behavior, high consistency levels are achieved. The navigation design of an application impacts how easily the user can operate it [46].

The health applications Babylon and Ada do not offer personalized recommendations to users. Upon entering symptoms, these applications only generate a list of possible diseases as shown in Figure. 7. These symptoms are added to the user's history. However, evaluators have reported that they did not observe any articles personalized to their history that met their expectations. There are studies that identified this problem the lack of personalization in AI based mobile health applications. However, personalized recommendations are particularly important in AI based Healthcare applications, as they enable to provide personalized information to user to meet their expectations which ultimately affect the satisfaction of user [47]. Personalized recommendations significantly increased people's likelihood of adopting healthy lifestyle behaviors, according to a study [48]. Another study, shows that the personalized diabetes management recommendations generated by an AI algorithm improve glycemic control better [49]. Additionally, a study found that the lack of user-specific data is a major challenge in developing personalized mobile health applications [50]. Furthermore, a study concluded that personalization is crucial in AI systems used for clinical decision-making, as it allows for more accurate and effective diagnoses and treatment plans [51]. Health care AI apps have predominantly focused on AI's analytical capabilities, and data handling, but have neglected human factors perspectives, resulting in poorly designed apps [52]. Issues such as the need for simpler navigation and better design have been noted which affect the ease-of-use usability attribute [53]. Research has also high-lighted the importance of satisfaction in health applications, as a factor to determine the success of a health care facility [54].

Finally, the evaluators from our study recommended to retrain the models with updated data to provide personal recommendations that meet individual needs of the users, and results in overall good user experience. To address these issues, evaluators proposed following solutions. They suggested that the addition of a search bar in such applications can significantly enhance a user's experience since it makes it easier for them to find the desired information without having to navigate through multiple screens. Additionally, prior research has shown that search functionality in health applications is extremely important [55],[56]. To address the problem of inaccurate predictions, it is recommended that these applications provide users with more personalized recommendations. It is suggested that a more relevant and accurate recommendation can be provided by collecting more user data and incorporating it into the AI algorithms. This suggestion is also reinforced by previous research that emphasizes the importance of personalization and personalized recommendations in mHealth applications [57]–[60]. Users who are unfamiliar with the application may be confused and frustrated by the absence of a "navigate to back" button on certain screens. By adding this feature to all screens, the application's overall usability can be improved and user frustration can be reduced. Additionally, it is recommended that these applications should include more engaging and attractive functionalities, such personalized feedback, goal-setting, and performance reporting, to improve the overall user experience. These features can increase application commitment and user motivation, which will lead to improved health results.

 ISBN: 978-1-68558-096-4

VII. CONCLUSION

In conclusion, the widespread use of mobile healthcare application with making use of artificial intelligence technology provide numerous healthcare services to their users. However, as the number of mobile applications in increasing day by day, evaluating their usability in terms of effectiveness of AI systems they provide, efficiency in terms of time user take to perform a task, ease of using application and over experience of user is crucial factor in their success. Cognitive walkthrough is one of many methods of usability evaluation. In this study, we selected 3 applications that make use of AI in their features, on the basis of their popularity, and availability to evaluate their usability through cognitive walkthrough. The identified tasks are symptom checking, virtual consultation, personal recommendations, health tracking and health information. 27 unique problems were identified after eliminating the repeating ones. The most of the problems were reported in symptom checking and health information tasks, 9 and 16 respectively. Since health information and health tracking impact value are lesser than symptom checking and virtual consultation. Therefore, the average severity of problems in Ada, Babylon and Ornament are 7.4, 8.0 and 4.2 respectively. Babylon has the most severity due to the high impact of symptom checking and virtual consultation tasks. The average time taken by users in these applications is varied, ornament being taking the longest time to complete tasks, because evaluators reported unresponsiveness issues in Ornament application. The evaluators provided recommendations for the identified problems to improve the effectiveness, efficiency, ease of use, and satisfaction of these applications. From our study, it is clear that evaluating the usability during the development and design of mHealth applications, especially those that use AI-based features is crucial to ensure the success and effectiveness of application. Adding more usability evaluation methods can enrich such studies taking into account also additional mobile applications. Involving larger number of users can be seen as an extension of this research project.

AUTHORS CONTRIBUTION

The author has formalized the idea, conducted a comprehensive literature review and then conducted the evaluation. The author then completed data analysis, discussion and writing up the paper.

ACKNOWLEDGMENT

This research was funded by the Deputyship for Research and Innovation, Ministry of Education in Saudi Arabia, grant number 523.

REFERENCES

[1] A. Muro-Culebras et al., “Tools for Evaluating the Content, Efficacy, and Usability of Mobile Health Apps According to the Consensus-Based Standards for the Selection of Health Measurement Instruments: Systematic Review,” JMIR Mhealth Uhealth 2021;9(12):e15433 https://mhealth.jmir.org/2021/12/e15433, vol. 9, no. 12, p. e15433, Dec. 2021, doi: 10.2196/15433.

[2] S. S. E. Alwi, M. A. A. Murad, S. Abdullah, and A. Kamaruddin, “A Prototype Development and Usability Evaluation of an E-health System,” Proceedings - AiIC 2022: 2022 Applied Informatics International Conference: Digital Innovation in Applied Informatics during the Pandemic, pp. 34–39, 2022, doi: 10.1109/AIIC54368.2022.9914606.

[3] “mHealth Apps Market Size, Share & Trends Analysis Report by Type (Fitness, Medical), by Region (North America, Europe, Asia Pacific, Latin America, Middle East & Africa), and Segment Forecasts, 2022-2030.” https://www.researchandmarkets.com/reports/4396364/mhealth-apps-market-size-share-and-trends (accessed Feb. 22, 2023).

[4] D. E. Jake-Schoffman et al., “Methods for Evaluating the Content, Usability, and Efficacy of Commercial Mobile Health Apps,” JMIR Mhealth Uhealth 2017;5(12):e190 https://mhealth.jmir.org/2017/12/e190, vol. 5, no. 12, p. e8758, Dec. 2017, doi: 10.2196/MHEALTH.8758.

[5] R. Alturki, V. G.-J. formative research, and undefined 2019, “The development of an Arabic weight-loss app Akser Waznk: qualitative results,” formative.jmir.org, Accessed: Feb. 22, 2023. [Online]. Available: https://formative.jmir.org/2019/1/e11785/

[6] A. I. Alharbi, V. Gay, M. J. Alghamdi, R. Alturki, and H. J. Alyamani, “Towards an Application Helping to Minimize Medication Error Rate,” Mobile Information Systems, vol. 2021, 2021, doi: 10.1155/2021/9221005.

[7] K. Komalavalli and R. Hemalatha, S. D.-S. I. Journal, and undefined 2020, “A Survey of Artificial Intelligence in Smart Phones and Its Applications among the Students of Higher Education in and around Chennai City.,” ERIC, doi: 10.34293/education.v8i3.2379.

[8] C. Cutillo, K. Sharma, L. Foschini, … S. K.-N. digital, and undefined 2020, “Machine intelligence in healthcare—perspectives on trustworthiness, explainability, usability, and transparency,” nature.com, Accessed: Mar. 10, 2023. [Online]. Available: https://www.nature.com/articles/s41746-020-0254-2

[9] R. Khajouei and A. Ameri, Y. J.-I. journal of medical informatics, and undefined 2018, “Evaluating the agreement of users with usability problems identified by heuristic evaluation,” Elsevier, Accessed: Feb. 22, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1386505618301254

[10] C. Lewis, P. Polson, C. Wharton, and J. Rieman, “Testing a Walkthrough Methodology for Theory-Based Design of Walk-Up-and-Use Interfaces”.

[11] C. Lewis, C. Wharton. of human-computer interaction, and undefined 1997, “Cognitive walkthroughs,” Elsevier, Accessed: Feb. 22, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780444818621500960

[12] N. Fasihah Jamaludin, “The usability evaluation of adolescent depression screening model using cognitive walkthrough

[13] ,” Asia-Pacific Journal of Information Technology and Multimedia Jurnal Teknologi Maklumat dan Multimedia Asia-Pasifik, vol. 11, no. 2, pp. 71–88, 2022, doi: 10.17576/apjitm-2022-1102-06.

[14] M. Farzandipour, E. Nabovati, H. Tadayon, and M. Sadeqi Jabali, “Usability evaluation of a nursing information system by applying cognitive walkthrough method,” Int J Med

Inform, vol. 152, p. 104459, Aug. 2021, doi: 10.1016/J.IJMEDINF.2021.104459.

[15] M. Georgsson, N. Staggers, E. Årsand, and A. Kushniruk, "Employing a user-centered cognitive walkthrough to evaluate a mHealth diabetes self-management application: A case study and beginning method validation," J Biomed Inform, vol. 91, p. 103110, Mar. 2019, doi: 10.1016/J.JBI.2019.103110.

[16] L. O. Bligård and A. L. Osvalder, "Enhanced cognitive walkthrough: Development of the cognitive walkthrough method to better predict, identify, and present usability problems," Advances in Human-Computer Interaction, vol. 2013, 2013, doi: 10.1155/2013/931698.

[17] M. Maguire.-I. journal of human-computer studies and undefined 2001, "Methods to support human-centred design," Elsevier, vol. 55, pp. 587–634, 2001, doi: 10.1006/ijhc.2001.0503.

[18] "Fifth Annual 'Pulse of Online Health' Survey Finds 66% of Americans Eager To Leverage Digital Tools To Manage Personal Health." https://www.prnewswire.com/news-releases/fifth-annual-pulse-of-online-health-survey-finds-66-of-americans-eager-to-leverage-digital-tools-to-manage-personal-health-300039986.html (accessed Jun. 13, 2023).

[19] A. Roess . J. Med and undefined 2017, "The promise, growth, and reality of mobile health-another data-free zone," researchgate.net, doi: 10.1056/NEJMp1713180.

[20] H. Cho, P. Y. Yen, D. Dowding, J. A. Merrill, and R. Schnall, "A multi-level usability evaluation of mobile health applications: A case study," J Biomed Inform, vol. 86, pp. 79–89, Oct. 2018, doi: 10.1016/J.JBI.2018.08.012.

[21] R. Schnall, J. Mosley, … S. I.-J. mHealth and, and undefined 2015, "Comparison of a user-centered design, self-management app to existing mHealth apps for persons living with HIV," mhealth.jmir.org, Accessed: Feb. 21, 2023. [Online]. Available: https://mhealth.jmir.org/2015/3/e91/

[22] O. V. Bitkina, H. K. Kim, and J. Park, "Usability and user experience of medical devices: An overview of the current state, analysis methodologies, and future challenges," Int J Ind Ergon, vol. 76, p. 102932, Mar. 2020, doi: 10.1016/J.ERGON.2020.102932.

[23] Y. A. Daraghmi, B. Yahya, and Y. Daraghmi, " Has Covid-19 affected software usability: mobile accounting system as a case" J Theor Appl Inf Technol, vol. 31, no. 2, 2023, Accessed: Feb. 17, 2023. [Online]. Available: www.jatit.org

[24] T. Krause, E. Jolkver, S. Bruchhaus, P. M. Kevitt, M. Kramer, and M. Hemmje, "A Preliminary Evaluation of “GenDAI”, an AI-Assisted Laboratory Diagnostics Solution for Genomic Applications," BioMedInformatics 2022, Vol. 2, Pages 332-344, vol. 2, no. 2, pp. 332–344, Jun. 2022, doi: 10.3390/BIOMEDINFORMATICS2020021.

[25] A. S. Dahri, A. S. Dahri, A. Al-Athwari, and A. Hussain, Usability Evaluation of Mobile Health Application from AI Perspective in Rural Areas of Pakistan, vol. 14, no. 15. International Association of Online Engineering, 2019. doi: 10.3991/ijim.v13i11.11513.

[26] M. N. Islam, S. R. Khan, N. N. Islam, M. Rezwan-A-Rownok, S. R. Zaman, and S. R. Zaman, "A Mobile Application for Mental Health Care During COVID-19 Pandemic: Development and Usability Evaluation with System Usability Scale," Advances in Intelligent Systems and Computing, vol. 1321, pp. 33–42, 2021, doi: 10.1007/978-3-030-68133-3_4.

[27] N. A. Zaini, S. F. M. Noor, and T. S. M. T. Wook, "Evaluation of APi interface design by applying cognitive walkthrough," International Journal of Advanced Computer Science and Applications, vol. 10, no. 2, pp. 306–315, 2019, doi: 10.14569/IJACSA.2019.0100241.

[28] M. Farzandipour, E. Nabovati, and M. S. Jabali, "Comparison Usability Evaluation Methods in a Health Information System: Heuristic Evaluation versus Cognitive Walkthrough Method," Oct. 2021, doi: 10.21203/RS.3.RS-871961/V1.

[29] T. Kim et al., "Assessing the usability of a prototype emergency medicine patient-centered electronic health record display," Proceedings - 2018 IEEE International Conference on Healthcare Informatics, ICHI 2018, pp. 424–425, Jul. 2018, doi: 10.1109/ICHI.2018.00083.

[30] P. G. Poison and C. H. Lewis, "Theory-Based Design for Easily Learned Interfaces," Hum Comput Interact, vol. 5, no. 2–3, pp. 191–220, 1990, doi: 10.1080/07370024.1990.9667154.

[31] P. Polson, C. Lewis, J. Rieman, C. W.-I. J. of man, and undefined 1992, "Cognitive walkthroughs: a method for theory-based evaluation of user interfaces," Elsevier, Accessed: Mar. 01, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/002073739290039N

[32] "Health. Powered by Ada." https://ada.com/ (accessed Mar. 01, 2023).

[33] "Babylon Health UK - The Online Doctor and… | Babylon Health." https://www.babylonhealth.com/en-gb (accessed Mar. 01, 2023).

[34] "Homepage | Ornament." https://ornament.health/ (accessed Mar. 01, 2023).

[35] R. Khajouei, M. Zahiri Esfahani, and Y. Jahani, "Comparison of heuristic and cognitive walkthrough usability evaluation methods for evaluating health information systems," J Am Med Inform Assoc, vol. 24, no. e1, pp. e55–e60, Apr. 2017, doi: 10.1093/JAMIA/OCW100.

[36] R. A. Virzi, "Refining the Test Phase of Usability Evaluation: How Many Subjects Is Enough?," https://doi.org/10.1177/001872089203400407, vol. 34, no. 4, pp. 457–468, Nov. 2016, doi: 10.1177/001872089203400407.

[37] R. S. Pressman, "Software engineering : a practitioner's approach," p. 941.

[38] S. Jayakumar et al., "Quality assessment standards in artificial intelligence diagnostic accuracy systematic reviews: a meta-research study," NPJ Digit Med, vol. 5, no. 1, Dec. 2022, doi: 10.1038/S41746-021-00544-Y.

[39] M. Modiba, "Artificial intelligence for the improvement of records management activities at the Council for Scientific and Industrial Research," Journal of the South African Society of Archivists, vol. 55, pp. 16–26, Nov. 2022, doi: 10.4314/jsasa.v55i.2.

[40] P. Esmaeilzadeh, "Use of AI-based tools for healthcare purposes: A survey study from consumers' perspectives," BMC Med Inform Decis Mak, vol. 20, no. 1, Jul. 2020, doi: 10.1186/S12911-020-01191-1.

[41] S. Berrouiguet, M. L. Barrigón, J. L. Castroman, P. Courtet, A. Artés-Rodríguez, and E. Baca-García, "Combining mobile-health (mHealth) and artificial intelligence (AI) methods to avoid suicide attempts: The Smartcrises study protocol," BMC Psychiatry, vol. 19, no. 1, Sep. 2019, doi: 10.1186/S12888-019-2260-Y.

[42] A. S. Alzahrani, V. Gay, R. Alturki, and M. J. Alghamdi, "Towards Understanding the Usability Attributes of AI-Enabled eHealth Mobile Applications," J Healthc Eng, vol. 2021, 2021, doi: 10.1155/2021/5313027.

[43] D. Li, "5G and intelligence medicine—how the next generation of wireless technology will reconstruct healthcare?," Precis Clin Med, vol. 2, no. 4, pp. 205–208, Dec. 2019, doi: 10.1093/PCMEDI/PBZ020.

[44] O. Strachna and O. Asan, "Systems Thinking Approach to an Artificial Intelligence Reality within Healthcare: From Hype to Value," ISSE 2021 - 7th IEEE International Symposium on

Systems Engineering, Proceedings, Sep. 2021, doi: 10.1109/ISSE51541.2021.9582546.

[45] "Importance of Adding Search Bar to your Website - The Next Scoop." https://thenextscoop.com/add-search-bar-to-website/ (accessed Mar. 10, 2023).

[46] "Search: Visible and Simple." https://www.nngroup.com/articles/search-visible-and-simple/ (accessed Mar. 10, 2023).

[47] L. Wiebelitz, P. Schmid, T. Maier, and M. Volkwein, "Designing User-friendly Medical AI Applications - Methodical Development of User-centered Design Guidelines," Proceedings - 2022 IEEE International Conference on Digital Health, ICDH 2022, pp. 23–28, 2022, doi: 10.1109/ICDH55609.2022.00011.

[48] "4 Reasons Healthcare should use Personalisation | Codehouse." https://www.codehousegroup.com/insight-and-inspiration/digital-strategy/4-reasons-healthcare-should-use-personalisation (accessed Mar. 10, 2023).

[49] S. M. Kelders, R. N. Kok, H. C. Ossebaard, and J. E. W. C. Van Gemert-Pijnen, "Persuasive System Design Does Matter: A Systematic Review of Adherence to Web-Based Interventions," J Med Internet Res 2012;14(6):e152 https://www.jmir.org/2012/6/e152, vol. 14, no. 6, p. e2104, Nov. 2012, doi: 10.2196/JMIR.2104.

[50] D. Bertsimas, N. Kallus, A. M. Weinstein, and Y. D. Zhuo, "Personalized Diabetes Management Using Electronic Medical Records," Diabetes Care, vol. 40, no. 2, pp. 210–217, Feb. 2017, doi: 10.2337/DC16-0826.

[51] J. ANDRES. "Exploring AI-based personalization of a mobile health intervention and its effects on behavior change, motivation, and adherence," 2021. http://reports-archive.adm.cs.cmu.edu/anon/hcii/CMU-HCII-21-104.pdf (accessed Mar. 20, 2023).

[52] Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, "Dissecting racial bias in an algorithm used to manage the health of populations," Science (1979), vol. 366, no. 6464, pp. 447–453, Oct. 2019, doi: 10.1126/SCIENCE.AAX2342/SUPPL_FILE/AAX2342_OBERMEYER_SM.PDF.

[53] "'Most healthcare apps not up to NHS standards' - BBC News." https://www.bbc.com/news/technology-56083231 (accessed Mar. 20, 2023).

[54] B. H. Kann, A. Hosny, and H. J. W. L. Aerts, "Artificial intelligence for clinical oncology," Cancer Cell, vol. 39, no. 7, pp. 916–927, Jul. 2021, doi: 10.1016/J.CCELL.2021.04.002.

[55] F. Manzoor, L. Wei, A. Hussain, M. Asif, and S. I. A. Shah, "Patient Satisfaction with Health Care Services; An Application of Physician's Behavior as a Moderator," Int J Environ Res Public Health, vol. 16, no. 18, Sep. 2019, doi: 10.3390/IJERPH16183318.

[56] W. J. Gordon, A. Landman, H. Zhang, and D. W. Bates, "Beyond validation: getting health apps into clinical practice," NPJ Digit Med, vol. 3, no. 1, Dec. 2020, doi: 10.1038/S41746-019-0212-Z.

[57] C. Lee Ventola, "Mobile Devices and Apps for Health Care Professionals: Uses and Benefits," Pharmacy and Therapeutics, vol. 39, no. 5, p. 356, 2014, Accessed: Mar. 16, 2023. [Online]. Available: /pmc/articles/PMC4029126/

[58] E. M. Grua, M. De Sanctis, I. Malavolta, M. Hoogendoorn, and P. Lago, "An evaluation of the effectiveness of personalization and self-adaptation for e-Health apps," Inf Softw Technol, vol. 146, p. 106841, Jun. 2022, doi: 10.1016/J.INFSOF.2022.106841.

[59] U. Josefsson, "Association for Information Systems AIS Electronic Library (AISeL) Exploring e-patients ' heterogeneity: towards personalized e-health applications," p. 2006, Accessed: Mar. 16, 2023. [Online]. Available: http://aisel.aisnet.org/ecis2006

[60] A. Saad, H. Fouad, and A. A. Mohamed, "Situation-aware recommendation system for personalized healthcare applications," J Ambient Intell Humaniz Comput, vol. 1, pp. 1–15, Feb. 2021, doi: 10.1007/S12652-021-02927-1/TABLES/4.

[61] X. Zhou, W. Liang, K. I. K. Wang, and S. Shimizu, "Multi-Modality Behavioral Influence Analysis for Personalized Recommendations in Health Social Media Environment," IEEE Trans Comput Soc Syst, vol. 6, no. 5, pp. 888–897, Oct. 2019, doi: 10.1109/TCSS.2019.2918285.

 ISBN: 978-1-68558-096-4