Usability Study of CARTIER-IA: A Platform for Medical Data and Imaging Management

Andrea Vázquez-Ingelmo[1(✉)] , Julia Alonso[1],
Alicia García-Holgado[1] , Francisco J. García-Peñalvo[1] ,
Jesús Sampedro-Gómez[2] , Antonio Sánchez-Puente[2] ,
Víctor Vicente-Palacios[3] , P. Ignacio Dorado-Díaz[2] ,
and Pedro L. Sánchez[2]

[1] GRIAL Research Group, Research Institute for Educational Sciences,
University of Salamanca, Salamanca, Spain
{andreavazquez,id00738609,aliciagh,fgarcia}@usal.es
[2] Cardiology Department, Hospital Universitario de Salamanca, SACyL.
IBSAL, Facultad de Medicina, University of Salamanca, and CIBERCV (ISCiii),
Salamanca, Spain
{jmsampedro,asanchezpu}@saludcastillayleon.es,
{acho,plsanchez}@usal.es
[3] Philips Healthcare, Salamanca, Spain
victor.vicente.palacios@philips.com

Abstract. Artificial Intelligence algorithms' application to medical data has gained relevance due to its powerful benefits among different research tasks. Nevertheless, medical data is heterogeneous and diverse, and these algorithms need technological support to tackle these data management challenges. The CARTIER-IA platform enables different roles (including principal researchers, IA developers and data collectors) to unify medical data, both structured data and DICOM images, and to apply Artificial Intelligence algorithms to them in a straightforward way through an online web application. However, given the diversity of roles involved in the platform, it is essential to account for its usability. It is necessary that users feel comfortable using the platform as relevant and complex tasks are carried out through its different services (such as the application of algorithms to the stored data, the manual edition of medical images or the visualization of structured data). This work presents a heuristic evaluation of the CARTIER-IA platform to improve its interaction mechanisms and get the most out of its functionalities.

Keywords: Data management · Structured medical data · Medical imaging · Health platform · Artificial intelligence · Usability · Heuristics

1 Introduction

Artificial intelligence algorithms are turning into a norm in the investigation of clinical information for research purposes, given their usefulness in identifying patterns in data and even images [1]. In fact, the application of deep learning models to medical

imaging has enabled the automation of complex tasks such as disease detection, segmentation of structures, or assessing organ functions [2] with similar performance compared to human skill [3].

However, these advances are generally limited to specific tasks and very specific datasets in which algorithms are trained and validated. That is why applying these technologies in real-world medical contexts is complex, due to the necessity of unifying several sources of image data (cohorts, machine brands, operators, etc.) without losing a reliable performance.

For these reasons, information systems are required to organize and gather all these data from different heterogeneous sources and to apply AI algorithms in a friendly, secure and anonymized manner.

This study presents a usability study of the CARTIER-IA platform. This platform is set to unify both structured data and medical images. Researchers and physicians can inspect data from different projects and also manipulate the images associated with different studies. The image editor also enables users to apply AI algorithms transparently, unburdening unskilled users of having to implement and run complex algorithms through command-line tools.

However, the complexity of the data collection processes, medical imaging edition and the great quantities of data that the platform holds, could make it difficult to use for novice users. That is why a heuristic evaluation of the platform has been carried out; to identify major and minor HCI issues with the goal of solving them and obtaining a friendly platform to get the most out of its features.

The rest of this paper is organized as follows. Section 2 outlines the platform and its main features. Section 3 describes the methodology followed throughout the study. Section 4 presents the results of the heuristic evaluation, which are discussed in Sect. 5. Finally, Sect. 6 concludes the work with a summary of the findings.

2   Platform Architecture

Two main tasks are supported by the platform: structured data and image data management. Different technologies and frameworks have been integrated through a client-server approach to obtain a web service that allows users to explore medical data and medical images without the necessity of installing external tools. Structured data and image data management are accomplished through different features.

The platform is organized into projects, which will have different anonymized patients. Each patient will have associated data, such as structured data, as well as image studies. Studies will be composed of sets of files and can also be characterized by adding more structured data at the study- and file-level.

Inside each projects' page, a data uploader allows users with enough permissions to upload both structured data and medical images. Both types of resources are differentiated, yielding two different uploaders. The structured data uploader accepts CSV and XLSX files containing the data.

Different tables or sheets can be created to further organize the structured data, and there are three levels in which data can be uploaded: patient-level, study-level and resource-level. Data can be explored through a tree-like architecture (Fig. 1) and can be filtered through the different specific variables created for each level (patient, study and file).
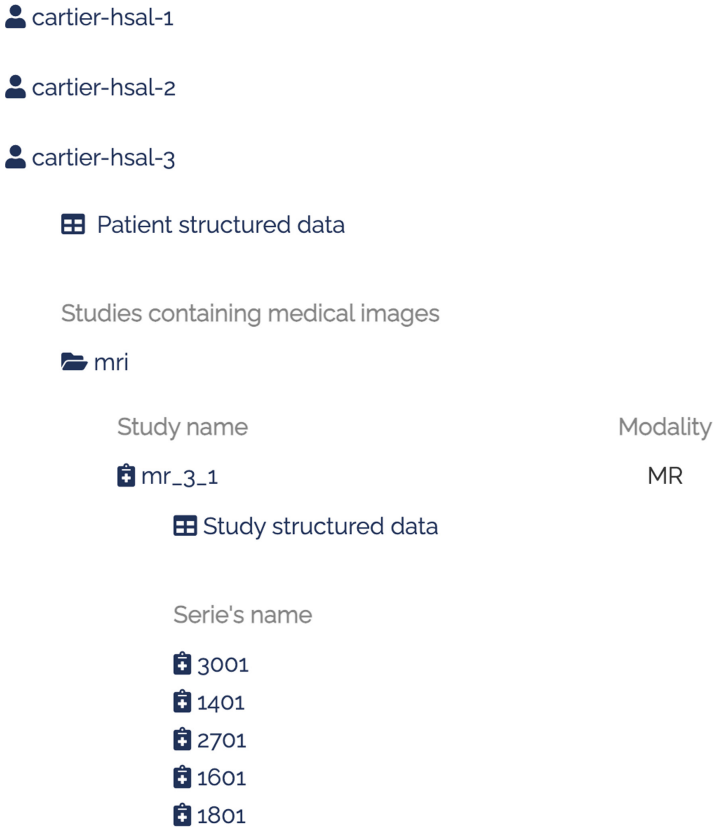
Fig. 1. Screenshot of the projects' navigation

Another main feature is image visualization and edition. The platform relies on an image editor in which users can draw, annotate, segmentize, crop, measure, etc., medical images (Fig. 2). Users can edit the images they are currently viewing. They can make annotations and modifications (such as cropping images). To make these changes permanent, the viewer has a button that sends the modifications to the server, storing all changes in the database.

In this case, annotations, segmentations, measurements, etc. are stores as raw data, so the image itself is never modified. This approach allows the storage of annotations from different users or different dates, thus enabling comparisons or even annotations' version control.

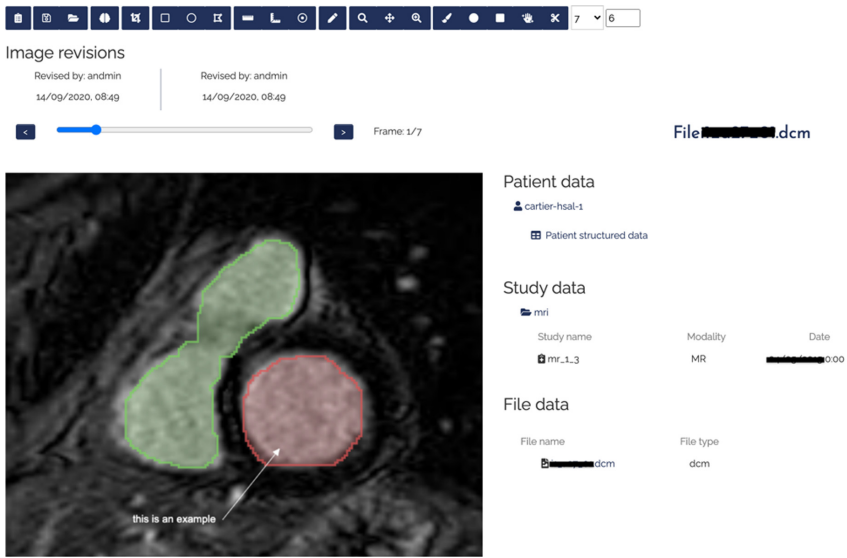**Fig. 2.** Screenshot of the image editor

The last main feature of the platform is the integration of Artificial Intelligence (AI) algorithms. This feature allows researchers to upload their AI scripts into the platform and make them available to other users transparently.

Once uploaded, algorithms are available at the viewer component. The application process is straightforward: a button provides information about the available scripts for the current image and the user only needs to select an algorithm and confirm their application.

The algorithm's outcomes can be structured data (which will be saved along with the current patient's structured data) or a new image (for example, a segmented image, which will be displayed along with the original image in the viewer).

3   Methodology

3.1   Participants

According to [4], the optimal number of experts involved in a heuristic evaluation varies among authors. Nielsen [5] considers that three to five evaluators identify 65–75% of the usability problems; meanwhile, other authors used eleven experts to identify 80% of the detectable problems. The review of methods of usability testing in

the development of eHealth applications [6] identifies that the studies which only used heuristic methods involve a maximum of five participants.

In this study, we have involved six experts, four single experts and two double experts. These double experts have expertise in user interface issues as well as the domain [7], in particular, they have used the CARTIER-IA platform as users before analyzing it (E2 and E6), the remaining had access for two weeks before completing the analysis (E1, E3, E4, E5). We looked for equitable distribution so that three men and three women participated, with ages ranging between 25 to 45 years old.

The experts were selected according to their profiles:

- E1: Web developer and researcher with eleven years of experience working on developing technological ecosystems for knowledge and learning management. Furthermore, the expert has experience in teaching human-computer interaction in a Computer Science degree.
- E2: A Ph.D. student whose doctoral dissertation deals with customizable dashboards to analyze and visualize any kind of data.
- E3: Student with experience in human-computer interaction and user experience.
- E4: Developer and researcher with more than ten years of experience focused on bioinformatics and information visualization, especially integrating different source data, analysis algorithms and representations for a better understanding of biological problems.
- E5: Researcher with more than ten years of experience in multimodal human-computer interaction.
- E6: Clinical Data Scientist and expert in the development of interfaces for the medical domain.

3.2   Instrumentation

There are numerous heuristics sets, although the most commonly used are the ten heuristics by Nielsen [5]. Despite there are heuristics specific to health care [8], those heuristics are mainly related to the evaluation of Electronic Health Records (EHR) [9]. These EHR heuristics include categories related to patients, collaborative team care or privacy [10]. CARTIER-IA platform is in the health context, but it is mainly focused on diagnosis and research tasks based on medical data, both structured data and DICOM images. Moreover, the medical data upload to the platform is anonymized, so privacy is important but not like in an EHR. Therefore, the selected set of heuristics was Nielsen's heuristics [5]:

- HR1: Visibility of system status.
- HR2: Match between system and the real world.
- HR3: User control and freedom.
- HR4: Consistency and standards.
- HR5: Error prevention.
- HR6: Recognition rather than recall.
- HR7: Flexibility and efficiency of use.
- HR8: Aesthetic and minimalist design.

- HR9: Help users recognize, diagnose, and recover from errors.
- HR10: Help and documentation.

We prepared a simple template and a set of guidelines to support the evaluation process and the reporting. The template was shared with each expert individually, in such a way that individual experts could not have access to the evaluation of the others.

The template has three fields to collect the evaluator's name, the name of the tool evaluated, and the browser used to access the CARTIER-IA platform. Furthermore, the template has a table with three columns - heuristic rule, points from 1 to 10, and problems detected – and one row per problem detected in each heuristic proposed by Nielsen.

3.3   Study Design and Data Collection

The first step was focused on selecting the set of heuristics, as described in the section above, as well as the desired number and characteristics of the experts. We selected the experts among our contact network, searching involving experts from different areas, expertise and non-direct relation with the project. The contact with the experts was made via email, where they were also provided with a brief contextualization of the CARTIER-IA platform.

Each expert applied the following indications:

- Navigate through the interface several times, looking at all screens, options, tasks. It is recommended to navigate at least once through the entire application to get to know the flow of interaction and what the application offers, and then a second review where you focus on specific details of the interface.
- Compare each screen/interface element with the ten usability principles proposed by Nielsen. For each detected problem:
- Give a value from 1 (non-relevant problems) to 10 (serious problems) to each problem detected.
- Brief description of the problem to justify the associated value.
- Indicate which heuristic is affected.
- The score for each heuristic will be an average of the values assigned to the detected problems.
- In addition to Nielsen's heuristics, the evaluator may also consider any additional usability principles or outcomes relevant to a specific interface element.

4   Heuristic Evaluation

Each expert was identified by a number (E1, E2, E3, E4, E5, E6) to show the heuristic evaluation results. The number of problems identified by each expert is small, but the combination of all of them provides an input to improve the CARTIER-IA platform. Table 1 summarizes the quantitative part of the heuristic evaluation providing by each expert. In particular, the table shows the average value assigned to the detected problems in each heuristic. This value is combined with the number of problems detected. Values near 1 indicate that the expert detected non-relevant problems, and

values near 10 mean that serious problems were identified. A zero value represents that the expert did not identify problems in that heuristic.

The total average of each heuristic rule was calculated in order to get a final value for each heuristic (Fig. 3). However, not only is the severity of the problems important, but we also wanted to analyze which of the heuristics has the greatest number of problems to solve (Fig. 4).

Most of the experts identified problems associated with all the heuristics, except E1, E3 and E5. In particular, those three experts did not detect any problem related to user control and freedom (HR3); meanwhile the rest of the experts detected few but serious problems such as:

- The deletion processes cannot be stopped (E2).
- Issues associated with the edition and deletion of the user account (E4).
- No support Undo and Redo of the actions available in the platform (E6).
- Moreover, E5 did not detect problems related to aesthetic and minimalist design (HR8). Regarding HR8, highlight the issues described by double experts (E2 and E6):
- On the project's home page, the way of showing information can be messy, as several patients-studies-files are shown using a tree-like architecture (Fig. 1). Also, several data are displayed alongside this structure, which makes the display very cluttered and difficult to read (E2).
- Clarity and simplicity by showing the relevant information of the clinical study (E6).
- Simplify the project creation design (E6).

The heuristic that presents the largest number of usability issues was HR4 (*Consistency and standards*) with 26 different problems (Fig. 4). Regarding the severity rating, the average is 6.84 (Fig. 3), most of them have a value over 5 (19 of 26 problems). The main problems detected are related to the translation of the interface (E1 and E4), the font size and responsive design (E5), the template download and upload data processes inside a project (E4), the different styles applied in links, buttons and icons (E2, E3 and E6), the differences between search or filtering (E3), the project's actions are located in different places inside the interface (E2) and the actions available in the DICOM viewer (E1). It should be noted that only expert E5 has detected problems related to accessibility standards, particularly in terms of the contrast between text and background.

The heuristic with a higher severity rating is help and documentation (HR10). All experts detected different problems mainly related to the lack of tutorials, legends and documentation support. Furthermore, there is no contact information to get technical support when the system breakdown or does not work as the user expects.

The heuristic HR9 (Help users recognize, diagnose, and recover from errors) has the second higher severity rating (Fig. 3). Experts identified problems related to:

- A reset button in the DICOM editor would be useful if the user wants to recover from an error or restart the process (E1).
- Sometimes, the DICOM editor's toolbar is blocked when the user clicks on IA algorithm or crop the image (E1).

- Errors information during the deletion and uploading processes are not displayed (E2, E3, E4, E5, E6).
- Not enough information, no error messages when an action is not permitted (E6).

Table 1. Heuristic evaluation summary by expert. N = number of errors

| Heuristic rule	E1		E2		E3		E4		E5		E6	
	avg	N	avg	N	avg	N	avg	N	avg	N	avg	N
HR1: Visibility of system status	5.75	4	5	4	8.5	2	9	1	3	2	9.75	4
HR2: Match between system and the real world	6	1	7.5	4	8	1	8.2	5	5	1	8.5	4
HR3: User control and freedom	0	0	8.5	2	0	0	6.33	6	0	0	9.25	4
HR4: Consistency and standards	7.13	8	4.67	3	7.75	4	7	3	5.25	4	9.25	4
HR5: Error prevention	5	1	7.67	3	6	2	10	2	8	1	9	4
HR6: Recognition rather than recall	4.33	3	8	2	7	4	8.5	2	3	1	9	4
HR7: Flexibility and efficiency of use	4	1	8.5	2	6	1	8.75	4	7	1	9.25	4
HR8: Aesthetic and minimalist design	2.5	4	6	3	7	3	7	3	0	0	9.5	4
HR9: Help users recognize, diagnose, and recover from errors	8.67	3	7.67	3	7	2	10	1	7	1	8.5	4
HR10: Help and documentation	9	2	9	2	8	2	9	2	9.5	2	9.25	4
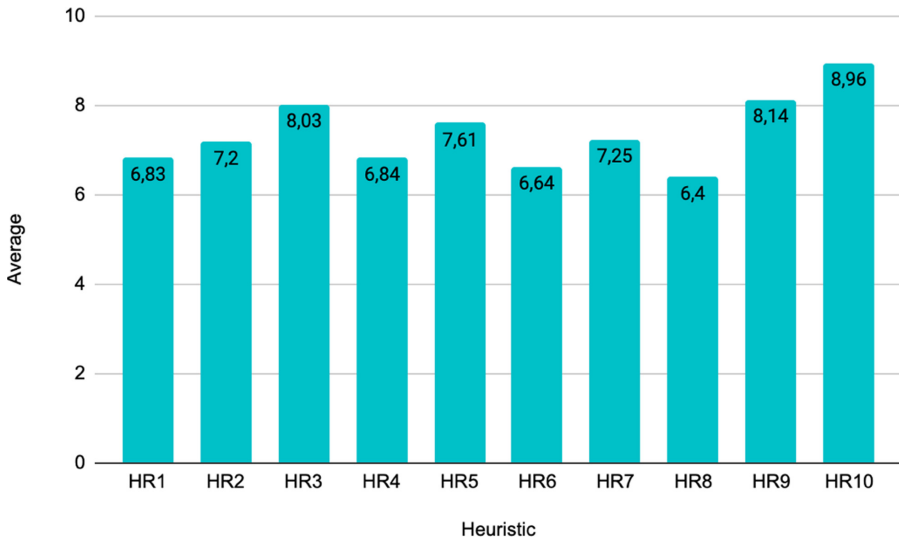
Fig. 3. The final average value for each heuristic rule
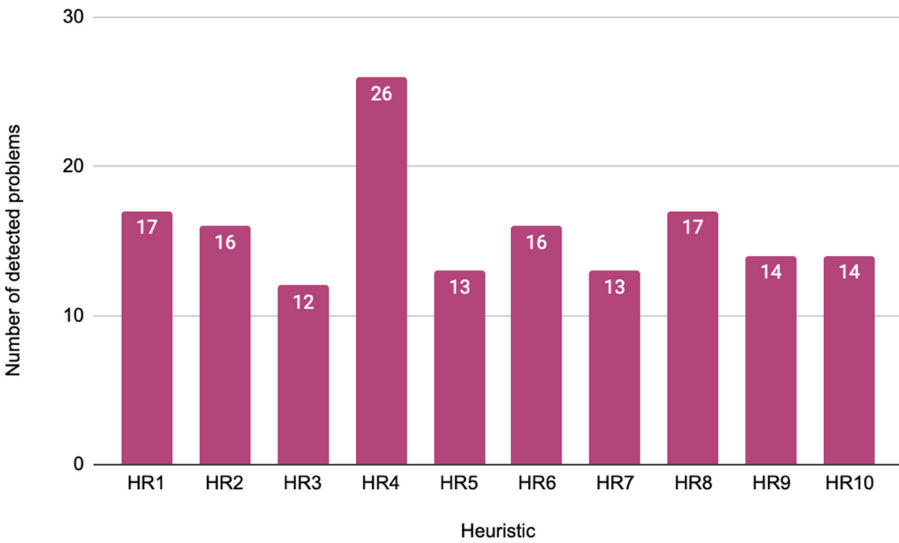
Fig. 4. The total number of detected problems per Nielsen's heuristic

On the other hand, the lowest number of usability problems was detected in HR3, HR5 and HR7. Regarding error prevention, each expert detected few errors; some of them are:

- Not enough information to define the password during the registration process (E1).
- The whole upload process involves different tasks that should be done in a specific order, but all actions are available from the beginning, which can be confusing and introduce errors. Some actions should not be accessible before performing the necessary preliminary steps and the process should be guided (E2, E4, E5).
- When select "download only selected data" and no data are selected, empty files are downloaded and the system does not advise the user he has not selected any data (E3).
- Users can apply erroneous filters when searching for patients in a project (E3).
- Clearly defined user privileges (E6).

Highlight some problems related to the remaining heuristics with high severity ratings. The main problems about the visibility of system status (HR1) are:

- The DICOM viewer does not show inside in which project is the image and you can come back to the project, only navigate through a set of images (E1).
- There is no progress bar indicating the deletion progress, though the process can be time-consuming when you delete several files (E2 and E6).
- No information about the user privileges and role (E4).

The problems related to matching between the system and the real world (HR2) are mainly related to the vocabulary used inside the projects. The language used is very

technical, such as table names, code reservation, 'upload' templates, etc. (E1, E2 and E4), and medical acronyms (E3 and E6).

Finally, although experts detected several problems related to recognition rather than recall (HR5), we want to emphasize the patients' section inside a project because it is at the bottom of the page and you have to scroll down to see them. According to E4, a reorganization of this screen will be useful to solve several usability problems.

5   Discussion and Conclusions

The CARTIER-IA platform unifies structure data and medical images, specifically DICOM images, to support researchers and physicians in the analysis associated with different studies, with a particular focus on supporting the application of AI algorithms in images.

The platform has been designed and developed using a user-centered approach, involving researchers and physicians to define the functionality and the different interfaces in an iterative process. Moreover, they were the data providers to test the platform. However, the complexity of the data collection processes, the edition of the DICOM images and the huge amount of data available in each project, whose can also include several studies, difficult the use of the platform for those users that are not directly involved in the design and development of the CARTIER-IA platform.

This study has served to identify the main usability issues to improve the platform with a particular focus on ensuring that novice users would be able to use it without a previous training. The group of experts involved in the study combined users with experience using the platform and users that discover the platform during the heuristic evaluation process. This approach has made it possible to achieve the study's main objective, which was to analyze whether a person joining the system for the first time can use it.

However, a heuristic evaluation does not ensure identifying all the problems that affect a real user of the platform. Some studies reveal whether the problems detected by the usability experts are problems that will affect or be encountered by the actual users of the system. In some areas, evaluators' perception in using this method is not consistent with the users' experience with a system [11]. One of the experts has extensive knowledge of the platform's domain to alleviate this problem, though further research might explore the user experience of the researchers and physicians.

Regarding the results of the evaluation, each expert identified a small number of usability problems. However, the combination of the results provides relevant information to develop a new version of the CARTIER-IA platform to solve the different identified problems. It would be interesting to assess this new version through a user testing and a second heuristic evaluation.

Even though the usability issues detected are related to most of the screens and functionality of the platform, the problems that most affect the use of the platform itself are those related to the information provided in each project and the DICOM viewer and editor. Several heuristics affect both issues. In particular, the DICOM viewer and editor appears in problems detected by three experts (E1, E2 and E6) and associated with seven heuristics (HR1, HR2, HR4, HR6, HR7, HR9 and HR10). Besides, after

receiving all the evaluations, we asked the other experts (E3, E4 and E5) about the image editor and all of them confirm that they did not notice the tool; they did not find that part of the platform. A reasonable approach to tackle this issue could be a user experience study with real users only focused on the DICOM viewer and editor.

References

1. Rajkomar, A., Dean, J., Kohane, I.: Machine learning in medicine. N. Engl. J. Med. **380**, 1347–1358 (2019)
2. Litjens, G., et al.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017)
3. Liu, X., et al.: A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digit. Health **1**, e271–e297 (2019)
4. Ssemugabi, S., De Villiers, M.R.: Effectiveness of heuristic evaluation in usability evaluation of elearning applications in higher education. South Afr. Comput. J. (SACJ) **45** (2010)
5. Nielsen, J.: Heuristic evaluation. In: Nielsen, J., Mack, R.L. (eds.) Usability inspection methods, vol. 17, pp. 25–62. John Wiley & Sons, Inc. (1994)
6. Maramba, I., Chatterjee, A., Newman, C.: Methods of usability testing in the development of eHealth applications: a scoping review. Int. J. Med. Inform. **126**, 95–104 (2019)
7. Nielsen, J.: Finding usability problems through heuristic evaluation. In: CHI 1992: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 373–380. ACM, New York, NY, USA (1992)
8. Dobre, J., et al.: Rapid heuristic evaluation: ensuring fast and reliable usability support. Proc. Hum. Factors Ergonomics Soc. Annu. Meeting **61**, 610–614 (2017)
9. Tarrell, A., Grabenbauer, L., McClay, J., Windle, J., Fruhling, A.L.: Toward improved heuristic evaluation of EHRs. Health Syst. **4**, 138–150 (2015)
10. Armijo, D., McDonnell, C., Werner, K.: Electronic health record usability: evaluation and use case framework. AHRQ Publication No. 09(10)-0091-1-EF. Agency for Healthcare Research and Quality, Rockville, MD (2009)
11. Khajouei, R., Ameri, A., Jahani, Y.: Evaluating the agreement of users with usability problems identified by heuristic evaluation. Int. J. Med. Inform. **117**, 13–18 (2018)