BreastScreening: On the Use of Multi-Modality in Medical Imaging Diagnosis

Francisco Maria Calisto
francisco.calisto@tecnico.ulisboa.pt
Interactive Technologies Institute
Funchal, Madeira, Portugal

Nuno Nunes
nunojnunes@tecnico.ulisboa.pt
Interactive Technologies Institute
Funchal, Madeira, Portugal

Jacinto C. Nascimento
jan@isr.tecnico.ulisboa.pt
Institute for Systems and Robotics
Lisbon, Estremadura, Portugal

ABSTRACT
This paper describes the field research, design and comparative testing of a multimodal medical imaging user interface for breast screening. The main contributions described here are threefold: (1) The design of an advanced visual interface for multimodal diagnosis of breast cancer (BreastScreening); (2) Insights from the field comparison of Single-Modality vs Multi-Modality screening of breast cancer diagnosis with 31 clinicians and 566 images; and (3) The visualization of the two main types of breast lesions in the following image modalities: (i) MammoGraphy (MG) in both Craniocaudal (CC) and Mediolateral oblique (MLO) views; (ii) UltraSound (US); and (iii) Magnetic Resonance Imaging (MRI). We summarize our work with recommendations from the radiologists for guiding the future design of medical imaging interfaces.

CCS CONCEPTS
• Human-centered computing → User studies; Usability testing; Interaction techniques; User centered design; User interface design.

KEYWORDS
human-computer interaction, user-centered design, multimodality, healthcare systems, medical imaging, breast cancer, annotations

1 INTRODUCTION
Breast cancer is the most common cancer in women worldwide [12]. Screening plays a fundamental role in the reduction of patient mortality rate. The most widely employed image modality for breast screening is MammoGraphy (MG). However, high-risk or dense breast patients require UltraSound (US) or Magnetic Resonance Imaging (MRI) for proper examination [18]. Therefore, it is quite rare to conduct screening using a Single-Modality.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

AVI ’20, September 28 – October 2, 2020, Island of Ischia, Italy
© 2020 Association for Computing Machinery.

In this paper, we describe the design and comparative testing of BreastScreening integrating information from several and different image modalities. We tested the design of BreastScreening with 31 clinicians noting that the time spent per each image on a Multi-Modality strategy is reduced when compared with the Single-Modality scenario. In addition, the lesion classification (e.g., Breast Imaging Reporting and Data System - BI-RADS [29]) is also reduced from our Multi-Modality proposed approach.

1.1 BreastScreening Challenges
Overall the system involves the following functionalities: (1) an interface for identifying (and annotating ground truth) of two types of lesions (i.e., masses and calcifications) across image modalities; (2) support for categorization of the breast tissues (dense vs non-dense); (3) a classification (and recommendation) schema for lesion severity using BI-RADS [1, 29]; (4) prompt access to clinical covariables, such as personal and familiar records; and (5) proper visualizations for a follow-up diagnosis of the patients.

1.2 Design Process
The following topics summarize the process we conducted: (1) findings from a formative study with 31 clinicians, comprising Radiology Room (RR) observations and interviews, which are relevant for both Health Informatics (HI) and Human-Computer Interaction (HCI) fields of research. This leads us to explore the design goals (see Section 3); (2) findings from an evaluation study [6] of BreastScreening, a prototype we developed for the generation of a breast dataset with expert annotations (see Section 4); and (3) design recommendations for the use of visualizations to support medical imaging diagnosis (see Sections 4 and 5).

1.3 Contributions
In BreastScreening we provide several new insights, following novel interaction and visualization paradigms [23] in the context of breast cancer screening: (i) multimodal interaction; (ii) indistinct visualization of cluttered lesions; (iii) big data management platform; and (iv) clinicians’ multi-screen, multi-environment interaction.

2 RELATED WORK
This section addresses related work in the HCI field, describing several medical imaging applications. Our approach covers the limitations of the works following described. More specifically, we are able to deal with non-homogeneous data. Comprising multimodal images [38], classification (i.e., BI-RADS scores) and annotations (i.e., delineation of the lesion contours).
2.1 Data Visualization
To our knowledge, few papers [15, 16, 24] have focused purely on supporting the image search user experience through novel UIs. These authors described several techniques for presenting all images within a collection in a short time. Moreover, authors asked users to think and perform browsing an image gallery and selecting an image from the gallery. These studies, showed us refinement techniques as complements in image systems with relevant user feedback. However, the presented works are limited to non-clinical users, making it impossible to do a generalization to our research.

2.2 Clinical Workflow
In medical imaging, diagnostic tools enable clinicians to manage patient data, better attend to ongoing tasks and view critical information. For the diagnostic, understanding the clinical workflow is of chief importance while introducing novel tools and interaction techniques. Other authors [11, 30] present many considerations for collaborative healthcare technology design and discuss the implications of their findings on the current clinical workflow for the development of more effective care interventions. Supported by the above literature, our goal is to introduce a new tool with several novel interaction techniques, which will improve the final medical imaging diagnosis.

2.3 Medical Imaging
From current medical imaging technologies, several issues were identified in the HCI design [5, 7, 13]. Some works [2, 25] show the current medical imaging identification techniques for other clinical domains, where most of available systems fail to address the visual nature of the task. In these two works [2, 25], the authors create a visual approach to support the Mental Model development of the user. Medical imaging technologies are used to support physicians on the examination, diagnosis, and (in some cases) report [37]. Others [10, 28, 33], study the effectiveness and performance of medical imaging systems, demonstrating how to design a user study for medical imaging experts. Further, van Schooten et al. [33] measured user performance in terms of time taken and error rate, while interacting with the provided system. Executing it with several medical users, in this work, the authors show an experiment where their users have similar characteristics as ours.

2.4 Diagnostic Systems
Medical imaging has also been extensively studied under the topic of Computer-Aided Diagnosis (CADx), which refers to systems that assist radiologists in image interpretation [4, 22]. Wilcox et al. [35] propose a design for in-room, patient-centric information displays, based on iterative design with clinicians. However, these systems are not contemplating the design of an advanced visual interface for multimodal diagnosis on breast cancer disease. In the above works, we still lack on empirical studies regarding how clinicians can contribute with information contextualization about their clinical workflow, and general medical imaging diagnosis. Having said that, we also want to add contribution with a study of how medical imaging technologies can play a role in this contextualization.

3 DESIGN OF BREASTSCREENING
The design of BreastScreening started with a qualitative study to understand radiology practices and workflow in the context of breast screening. Our study involved 31 clinicians, recruited on a volunteer basis from a large range of clinical scenarios (distinct health institutions in Portugal): 8 clinicians from Hospital Fernando Fonseca; 12 clinicians from IPO-Lisboa; 1 clinician from Hospital de Santa Maria; 8 clinicians from IPO-Coimbra; 1 clinician from Madeira Medical Center; and 1 clinician from SAMS. Clinicians’ experience ranged from 5 to 30 years of medical practice. The recruited specialists are in advanced career positions and were observed and interviewed in a semi-structured fashion. Each session took approximately 30 minutes.

3.1 Standard Clinical Environments
BreastScreening works with the standard formats supported by medical imaging [21], including the MG, US and MRI modalities. These modalities are available in a standard Digital Imaging and Communications in Medicine (DICOM) format and supported in Single-Modality by existing systems [12]. Moreover, most systems are general purpose and do not adapt to specific clinical domains (e.g., breast screening). Therefore they do not provide adequate support to the different clinical workflows [7].

3.2 Design Goals
Combining the clinical context and the technical design challenges lead to a set of design issues, including: medical imaging structure trade-offs, RR temporal awareness, image segmentation [20], and radiologists system trust. Based on these, we define five design goals:

Design around and for Medical Imaging (DMI): by taking into account the heterogeneous nature of medical imaging to leverage its contextual richness;

Temporal Awareness Support (TAS): by observing how the radiology workflow events, treatments, and problems progressed over time;

Image Segmentation Support (ISS): the overview of image details allowing a more accurate diagnostic. Namely, reducing the number of false-positives classification (BIRADS) of the lesion, as well as improving the number of clicks (Section 5) when performing the lesion delineation, i.e., segmentation;

Several Modalities Support (SMS): to enable the view and the process of diagnostic imaging studies, including MG, US and MRI medical imaging modalities;

Growing Trust Overview (GTO): by allowing an efficient triangulation via visualizations, image processing between medical images and available features, i.e., annotations of masses and calcifications;

4 BREASTSCREENING
To validate the proposed design goals, we created BreastScreening, as a Medical Imaging visualization proof-of-concept to be evaluated in a realistic clinical scenario. In our design explorations, we sought to integrate several image modalities and visualization to support insight.
4.1 User Interface

The User Interface (UI) consists of two main components: 4. List of Patient Views; and 5. Medical Imaging Diagnosis Views. These two main components (Figure 1) are also divided into several sections: 4.5. Study List Tabs; 5.1. Viewports; 5.2. Toolbars; and 5.3. Modality Selection. Concerning 5. Medical Imaging Diagnosis Views (Viewports, Toolbars and Modality Selection) this contributes for the temporal awareness (TAS). More specifically, the clinician can probe for lesion patterns [17] via the 5.1. Viewports, processing the image by using the 5.2. Toolbars features (GTO). The system 5.2. Toolbars are supporting our image segmentation (ISS). The 5.1. Viewports are displayed right after the 5.2. Toolbars, designing around and for medical images (DMI) what also improves the temporal awareness (TAS) of the task. On the same time, this design is supporting the way how to interact with several modalities (SMS). Regarding 5.3. Modality Selection, this allows to the clinician to find more different views (SMS) of the same lesion, allowing to perform a better severity classification (Section 5). Finally, the clinician may look for the lesion shape and contour irregularities (Figure 1) to focus on the segments of the image (ISS). After interacting with the system at the first time, the clinician is able to efficiently process (ISS) several images at a same time and use the various given modalities (SMS).

4.2 Implementation

BreastScreening was implemented using CornerstoneJS [32] with a NodeJS server. To populate the system, we selected image sets from HFF patients and upload them into an Orthanc server [14]. The images were pre-processed and anonymized on the Orthanc server and then consumed by the BreastScreening system. The BreastScreening core is developed in JavaScript with jQuery for HTML document manipulation, event handling and dicomParser for parsing DICOM files. The DICOM files can be loaded by drag-and-drop files into the browser window on the Orthanc view.

5 RESULTS

We conducted an evaluation of BreastScreening in real-world conditions. Our goal was to quantitatively and qualitatively assess the proposed design principles and to understand how these principles will play in practice [3]. We are particularly interested in understanding how the design goals and challenges (Section 3) are addressed [34]. Ultimately, we are focused on clinicians’ opinions how to improve diagnostic reliability. To accomplish this, the clinicians will have first to deal with: i) new mechanisms of multi-modal data visualization; ii) identification and delineation of lesions; and iii) classification of severity (i.e. BIRADS). The experimental setup aimed at testing two conditions: Cond. C1 - Single-Modality, and Cond. C2 - Multi-Modality. For each condition (i.e., Single-Modality or Multi-Modality) we collected complete imaging exams for three patients (P1, P2 and P3) on all possible modalities (MG, US and MRI). The MG and US comprise a single 2D image (i.e., static modality), whilst the MRI [19, 27] comprises a volume with N slices (i.e., dynamic modality [26]). The exams were previously annotated and classified with a BIRADS severity from an expert doctor who leads the HFF radiology department.
5.1 Participants
Our study involved 31 clinicians, recruited on a volunteer basis from a broad range of clinical scenarios, including six different health institutions (two public hospitals, two cancer institutes and two private clinics). From the demographic questionnaires: 16.10% of the clinicians have between 31 and 40 years of practical experience (Seniors), 45.20% have between 11 and 30 years of experience (Middle), 9.70% have between 6 and 10 years of experience (Juniors), and 29% have limited experience (Interns). Interviews were conducted in a semi-structured fashion taking about 30 minutes. Overall, 17 days were spent on the clinical institutions for the observation process and six months for the classification.

5.2 Quantitative Analysis
Four relations² emerged from our analysis: a) differences between SUS Scores and SUS Questions [31] among clinical experience (i.e., Intern, Junior, Middle, and Senior); b) the workload measurements of both Single-Modality and Multi-Modality views; c) the relation between Time and Number of Clicks, clustering by Patient (i.e., P1, P2 and P3). The expert classification for the patients used in this study are BIRADS(P1) = 2, BIRADS(P2) = 5 and BIRADS(P3) = 3 respectively, for both Single-Modality and Multi-Modality views; and, d) the distributions of the BIRADS variation (Figure 2).

5.2.1 SUS Scores vs SUS Questions. The ANOVA test² [36] yields a significant difference in both Single-Modality (F_SM = 11.79, p_SM = 0.001 < 0.05) and Multi-Modality (F_MM = 2.93, p_MM = 0.001 < 0.05) conditions among the various clinical experience of Clinicians. Participants adopting the Multi-Modality (M_MM = 2.9, SD_MM = 0.90) condition obtained higher SUS scores than those using the Single-Modality (M_SM = 2.7, SD_SM = 1.01) condition.

5.2.2 Workload. The results generated from the NASA-TLX [9] yields a significant main effect for the Physical Demand (F_SM = 5.81, p_SM = 0.003 < 0.05) and Temporal Demand (F_SM = 4.86, p_SM = 0.009 < 0.05). On the other hand, the Multi-Modality condition indicates that there exists a significant difference among Mental Demand (F_MM = 3.13, p_MM = 0.04 < 0.05), Physical Demand (F_MM = 4.61, p_MM = 0.009 < 0.05), and Temporal Demand (F_MM = 9.17, p_MM = 0.001 < 0.05). The NASA-TLX yields significant difference among groups for both Effort (F_SM = 3.74, p_SM = 0.02 < 0.05) and Frustration (F_MM = 3.93, p_MM = 0.01 < 0.05).

5.2.3 Time vs Number of Clicks. Results showing the amount of Time and Number of Clicks in each of the 566 images among the three patients are following described. The ANOVA test shows a non-significant interaction effect over the total Time from both Single-Modality (F_SM = 0.68, p_SM = 0.56 > 0.05) and Multi-Modality (F_MM = 0.28, p_MM = 0.83 > 0.05) regarding the clinical experience groups. In addition, our results show a non-significant interaction effect over the total amount of Number of Clicks from both Single-Modality (F_SM = 1.76, p_SM = 0.17 > 0.05) and Multi-Modality (F_MM = 0.57, p_MM = 0.63 > 0.05).

5.2.4 BIRADS Classification. The first and second order statistics of the BIRADS classification is shown in Figure 2. The mean values are referenced to the patient BIRADS (previously performed by the expert), that is, we have (from left to right) the patients P1, P2 and P3, with BIRADS\_real = 2, BIRADS\_real = 5 and BIRADS\_real = 3, respectively. From this figure, it is clear that the Multi-Modality performs better, since the most severe BIRADS exhibits the smallest mean and variance (BIRADS\_real - BIRADS\_provided) in the most of the cases. Also note that for the most problematic patient (in this case P2 scored with BIRADS = 5) the multi-modal largely outperforms the Single-Modality setting.

5.3 Qualitative Analysis
Clinicians were invited to give some feedback about the UI during the open interviews. We received several positive comments regarding our BreastScreening system. At the end, several clinicians (19/31) answered that the assistant will be an asset of an immense importance for the current RR situation: “The system will be a great asset for us” (C6). Another positive answer was the one related to the frequency of use (28/31) for this new assistant regarding the current system used by the clinicians on the daily practice: “I would like to frequently use your system on my daily practice” (C1).

6 CONCLUSION
Medical imaging systems provide a promising but challenging problem for HCI research. In this paper, we presented field research, design and comparative deployment of a multimodal user interface for breast screening. BreastScreening is a proof-of-concept prototype developed to embody the emerging design goals from the underlying clinical context. Our work and contributions included:

- identifying the main clinical workflow issues, the interaction cognitive load challenges [8] and the opportunities;
- establishing a set of design goals for medical imaging design;
- the design, reflections and in-situ evaluation of BreastScreening supporting the clinical translation; and
- the impact evidence of Multi-Modality in diagnosing and severity classification of breast lesions with 31 radiologists in six different clinical institutions. Our results show the system can lead to more efficient and accurate clinical diagnosis.

²Available datasets usability (mimbd-ui.github.io/dataset-uta4-user), workload (mimbd-ui.github.io/dataset-uta4-workload), time (mimbd-ui.github.io/dataset-uta4-time), severity rates (mimbd-ui.github.io/dataset-uta4-rates), and images (mimbd-ui.github.io/dataset-uta4-dicom).

³N: the number of users (Clinicians); F(var) the F-test used for comparing the factors of the total variation per each variable (var) categorized by clinical experience; M\_var Mean value of the variable (var); SD\_var the Standard Deviation (SD) per each variable (var).
