MIDA
User Testing & Analysis Guide
Clinical Opinion About Intelligent Agents

Francisco Maria Calisto
francisco.calisto@tecnico.ulisboa.pt
Instituto Superior Técnico
University of Lisbon
Portugal

21/11/2019

BreastScreening:	breastscreening.github.io
Meta:	github.com/BreastScreening/meta
Datasets:	github.com/BreastScreening/meta/wiki/Datasets
MIDA:	mida-project.github.io
Meta:	github.com/mida-project/meta
Datasets:	github.com/mida-project/meta/wiki/Datasets
MIMBCD-UI:	mimbcd-ui.github.io
Meta:	github.com/MIMBCD-UI/meta
Datasets:	github.com/MIMBCD-UI/meta/wiki/Datasets

1 Introduction

During the breast cancer screening, missing cancers may not be identified until they are more advanced and less agreeable to treatment [2]. Artificial Intelligence (AI) in the medical workflows may help with this challenge [3]. Studies have demonstrated the ability of AI to meet the human's performance on various clinical tasks [6, 7]. As a lack of medical professionals threatens the adequacy and availability of clinical services worldwide [4, 5], the scalability of AI could improve to higher care.

The role of Human-AI Interaction (HAII) in healthcare delivery the appropriate settings in which it can be applied, and its impact on the quality of care have yet to be evaluated [8]. There have been several attempts at addressing the effects of HAII across multiple workflows and different levels of clinical expertise [1, 9]. However, the use case of breast cancer diagnosis to address the effects from varied representations of AI-based supported by intelligent agents is still scarce. This explains why it is an open topic research, and the motivation behind the proposed research of this User Testing and Analysis (UTA) guide.

2 Description

3 Methodology

4 Roles

The roles involved in our user tests are as follows. An individual may play multiple roles, as well as the test may not require all roles.

4.1 Trainer

- Provide training overview prior to user testing phases;

4.2 Facilitator

- Provides overview of study to participants;
- Defines tasks and purpose of the user testing to participants;
- Assists in conduct of participant and observer debriefing sessions;
- Responds to participant's requests for assistance;

4.3 Data Logger

- Records participant's actions and comments;

4.4 Test Observers

- Silent observer;
- Assists the data logger in identifying problems, concerns, coding bugs and procedural errors;
- Serve as note takers;

4.5 Ethics

All persons involved with the Usability (Usa.) test are required to adhere to the following ethical guidelines:

- The performance of any test participant must not be individually attributable;
- Individual participant's name should not be used in reference outside the testing session;
- A description of the participant's performance should not be reported to his or her superior;

5 Apparatus

6 Evaluation

7 Tasks

8 Metrics

9 Goals

10 Challenges

11 Results

12 Acknowledgements

A special thanks for the support and revisions provided by Hugo Lencastre and Nádia Mourão. We would like to thank Doctor Clara Aleluia, Doctor Gisela Andrade, Dr. Willian Schmitt, Dr. Ana Sofia Germano and Dr. Pedro Marques from the HFF for the generous support and medical expertise. Also, an immense thank for Doctor Cristina Ribeiro da Fonseca. My appreciation goes

also to Bruno Cardoso and Bruno Dias for help and above all for the good companionship. Thanks to Professor Daniel Gonçalves, Professor Daniel Simões Lopes and Daniel Mendes for the technical inputs and network. Last but not least, thank to my advisors Professor Jacinto C. Nascimento and Professor Nuno Jardim Nunes. We also want to provide a special acknowledgment to Professor Ramtin Zargari Marandi who, among others, gave us important information and comments regarding the presented report. This work was partially supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2013 and Instituto Superior Técnico (IST-ID) through the FCT/UID/EEA/50009/2013 project, BL89/2017-IST-ID grant. We would like to convey Hospital Fernando Fonseca (HFF) for the collaboration.

Acronyms

AI Artificial Intelligence.

Assis. Assistant.

BIRADS Breast Imaging Reporting and Data System.

CC CranioCaudal.

DICOM Digital Imaging and Communications in Medicine.

DOTS Dimensions Of Trust Scale.

HAII Human-AI Interaction.

MG MammoGraphy.

MI Medical Imaging.

MIDA Medical Imaging Diagnosis Assistant.

MIMBCD-UI MI Multimodality Breast Cancer Diagnosis UI.

MLO MedioLateral Oblique.

MM Multi-Modality.

MRI Magnetic Resonance Imaging.

NASA-TLX NASA Task Load Index.

SS Single-Modality.

SUS System Usability Scale.

UI User Interface.

US UltraSound.

UTA User Testing and Analysis.

References

[1] Krzysztof J. Geras, Ritse M. Mann, and Linda Moy. Artificial intelligence for mammography and digital breast tomosynthesis: Current concepts and future perspectives. *Radiology*, 293(2):246–259, 2019. PMID: 31549948.

[2] Nehmat Houssami and Kylie Hunter. The epidemiology, radiology and biological characteristics of interval breast cancers in population mammography screening. *npj Breast Cancer*, 3(1):12, Apr 2017.

[3] Scott Mayer McKinney, Marcin Sieniek, Varun Godbole, Jonathan Godwin, Natasha Antropova, Hutan Ashrafian, Trevor Back, Mary Chesus, Greg S. Corrado, Ara Darzi, Mozziyar Etemadi, Florencia Garcia-Vicente, Fiona J. Gilbert, Mark Halling-Brown, Demis Hassabis, Sunny Jansen, Alan Karthikesalingam, Christopher J. Kelly, Dominic King, Joseph R. Ledsam, David Melnick, Hormuz Mostofi, Lily Peng, Joshua Jay Reicher, Bernardino Romera-Paredes, Richard Sidebottom, Mustafa Suleyman, Daniel Tse, Kenneth C. Young, Jeffrey De Fauw, and Shravya Shetty. International evaluation of an ai system for breast cancer screening. *Nature*, 577(7788):89–94, Jan 2020.

[4] S Moran and H Warren-Forward. The australian breastscreen workforce: a snapshot. *Radiographer*, 59(1):26–30, 2012.

[5] Abi Rimmer. Radiologist shortage leaves patient care at risk, warns royal college. *BMJ: British Medical Journal (Online)*, 359, 2017.

[6] Jiayi Shen, Casper J P Zhang, Bangsheng Jiang, Jiebin Chen, Jian Song, Zherui Liu, Zonglin He, Sum Yi Wong, Po-Han Fang, and Wai-Kit Ming. Artificial intelligence versus clinicians in disease diagnosis: Systematic review. *JMIR Med Inform*, 7(3):e10010, Aug 2019.

[7] Eric J. Topol. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1):44–56, Jan 2019.

[8] Philipp Tschandl, Christoph Rinner, Zoe Apalla, Giuseppe Argenziano, Noel Codella, Allan Halpern, Monika Janda, Aimilios Lallas, Caterina Longo, Josep Malvehy, John Paoli, Susana Puig, Cliff Rosendahl, H. Peter Soyer, Iris Zalaudek, and Harald Kittler. Human–computer collaboration for skin cancer recognition. *Nature Medicine*, 26(8):1229–1234, Aug 2020.

[9] Walter F. Wiggins, M. Travis Caton, Kirti Magudia, Sha-har A. Glomski, Elizabeth George, Michael H. Rosenthal, Glenn C. Gaviola, and Katherine P. Andriole. Preparing radiologists to lead in the era of artificial intelligence: Designing and implementing a focused data science pathway for senior radiology residents. *Radiology: Artificial Intelligence*, 0(ja):e200057, 0.