

Low-Code Application for Ground Truth Data Acquisition in Mammogram Retrieval

Cátia Roriz¹
a21270590@isec.pt

Verónica Vasconcelos¹
veronica@isec.pt

Inês C. Moreira²
icm@ess.ipp.pt

Inês Domingues^{1,3}
ines.domingues@isec.pt

¹ Instituto Politécnico de Coimbra, Instituto Superior de Engenharia, Rua Pedro Nunes Quinta da Nora, 3030-199 Coimbra, Portugal

² Centro Hospitalar Universitário de São João Escola Superior de Saúde do Politécnico do Porto Cintesis-FMUP

³ Centro de Investigação do Instituto Português de Oncologia do Porto (CI-IPOP): Grupo de Física Médica, Radiobiologia e Protecção Radiológica

Abstract

Breast cancer is a significant global health concern, affecting thousands of individuals, primarily women, with estimated cases expected to climb by 2040. This paper describes the creation of an application to collect ground truth data to aid engineers in the development of a mammography retrieval system. The application is built upon OutSystems, a low-code application platform. Key features of the application include allowing experts to view probe images and associate them with relevant images from the database. Additionally, the platform allows image filtering based on eight mammogram dimensions. While the ultimate goal is to create a system for medical specialists, the current platform represents a step in the process, facilitating the acquisition of ground truth.

1 Introduction

Breast cancer is a malignant disease that affects thousands of individuals worldwide, primarily women. By the year 2040 [1], a significant increase in the number of breast cancer cases is estimated, with more than 3 million new cases annually, representing a 40% increase. In addition, the number of deaths related to this disease is expected to increase by 50%, exceeding 1 million deaths per year. Although various approaches have been developed to deal with this problem, much more effort has to be done to fight this disease.

The main objective of this work is to develop an initial version of a platform that will serve as a tool to gather data that will facilitate the development of a mammogram retrieval system [4]. The OutSystems platform was used to develop this application [2].

The remainder of this document is organised as follows. This section presents an introductory reflection on the subject. Section 2 provides a brief overview of existing software. Section 3 describes the application requirements and use cases, while Section 5 presents the performed implementation. Finally, Section 6 addresses the final considerations and future work.

2 Available Mammogram Retrieval Software

The retrieval of relevant mammograms is the ultimate goal of the software that will be created. Table 1 shows some competing software. In contrast to the software provided in Table 1, the created software thus far allows experts to submit ground truth data for the later building of retrieval models [4]. To the best of our knowledge, no tools are currently available for this purpose.

Table 1: Competing Software

	PenFetch [3]	Mammogram Image Retrieval System [5]
Capabilities	Automates fetching previous medical studies from PACS ¹	Users can send RF ² to improve system understanding of information needs
Methods	Uses DICOM ³ worklists to identify patients and images, requesting transfers based on predefined rules	Uses text-based methods for initial search and augments outcomes with visual aids
Advantages	Streamlines workflow by sending pertinent images to radiologists' stations ahead of time, minimising delays	Applications include computer-aided diagnosis, medical education, and research
Disadvantages	Dependency on DICOM Worklists. Limited to PACS environments	Specific focus on certain domains may limit broader applicability

¹ PACS: Picture Archiving and Communication System

² RF: Relevance Feedback

³ DICOM: Digital Imaging and Communications in Medicine

3 Requirements analysis

To be able to develop the application, an initial list of requirements was identified:

ID01- View probe image: The software must allow the visualisation of the probe image that will later be analysed and associated with a relevant image from the database.

ID02- View images in the database: The software shall allow the visualisation of the various images of the database, images that are tagged according to the several dimensions of the mammograms.

ID03- Filter images based on dimensions: The software should allow the filtering of the images in the database according to their dimensions such as breast density, existence or not of asymmetries, Breast Imaging-Reporting and Data System (BI-RADS), existence or not of calcifications, existence or not of distortions, laterality (right or left breast), existence or not of masses and image incidence (Cranio-Caudal (CC) or Medio-Lateral Oblique (MLO)).

ID04- Select an image from the database: For each probe image, the software must allow the selection of one image from the database, the one considered by the expert as the most relevant.

ID05- Visualise registrations: The software must allow the visualisation of all the registrations made in the application.

During the development of the software, several improvements were suggested by the expert. These include:

ID06- The probe image cannot be associated with it itself: The software must ensure that the probe image does not appear in the database images to be selected.

ID07- The probe image must always be visible: The software should allow the probe image to be visible at all times, i.e. it should always remain visible while selecting the image from the database.

ID08- Zoom for Enhanced Viewing: The software should allow zooming in on images, both probe images and database images.

ID09- Navigation to Next Probe Image: The software should allow to skip the current probe image and move on to the next probe image while still having access to the prior probe image.

ID10- Filter selection confirmation: The software should include a functionality to generate a warning, allowing for verification of whether the filter selection has been finalized or not.

The requirements are pictured in the Use Case Diagram of Figure 1.

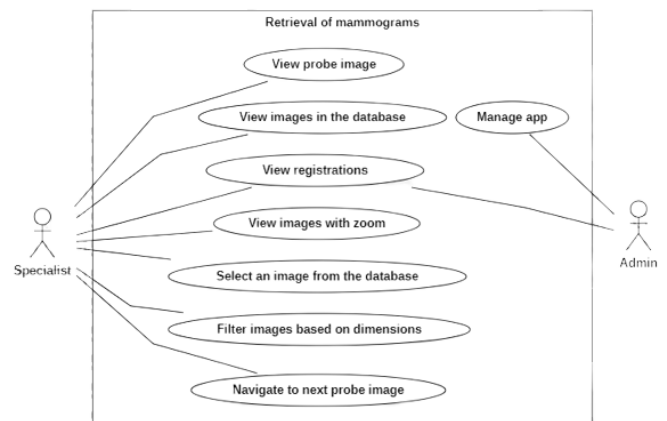


Figure 1: Use case diagram

4 Database

The data model, shown in Figure 2, was created to represent the information that is necessary for the application to meet its objectives.

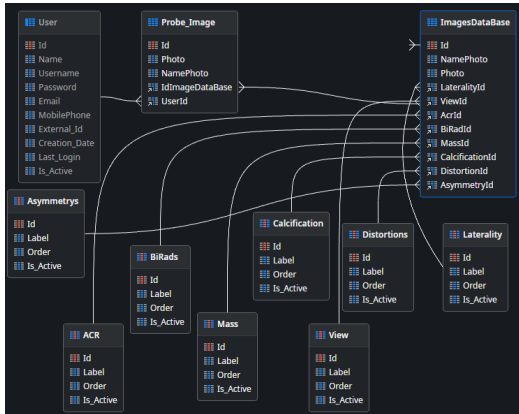


Figure 2: Database

The Database was built with three entities (Probe_Image, ImagesDataBase and User) and eight static entities (Distortions, ACR, BiRads, Mass, Calcification, Asymmetrys, View and Laterality). The Probe_Image entity represents all the probe images that will be associated with an image in the database. The ImagesDataBase entity represents all the images of the database, where only one of these, in each interaction, will be selected as the relevant image. The User entity is a default entity of the OutSystems platform and represents the registration of users on the platform. Static entities are entities that have a predefined set of possible values and in this context represent the dimensions of the mammograms. Static entities have a set of values ("Records") as shown in Figure 3.

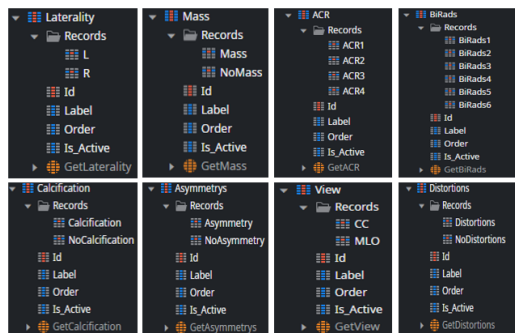


Figure 3: Static Entities Records

5 Implementation

The application starts with a login page (Figure 4) where the expert can fill in the form with their credentials to access the software.

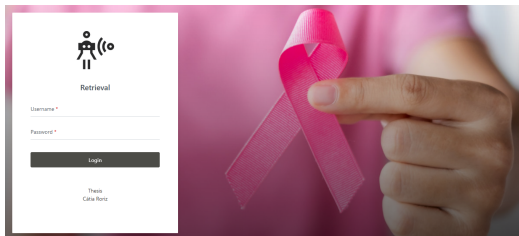


Figure 4: Login page

If the provided software access credentials are valid, the user will be directed to the homepage. There, the user only needs to click the "Start" button, which will then take them to a new page (Figure 5). This newly accessed page enables image retrieval, the specialist associates the probe image (highlighted in blue in Figure 5) with the most relevant image on the right (highlighted in orange in Figure 5). To assist the expert in their selection process, the images on the right (highlighted in orange in Figure 5) can be filtered using available filters (highlighted in green in Figure 5).

These filters allow the selection of images with specific dimensions of the mammogram (as seen in Figure 3) exclusively. After conducting a thorough analysis, the expert can select the most relevant image by selecting the respective checkbox, followed by clicking the "Save" button to save the record in the database.

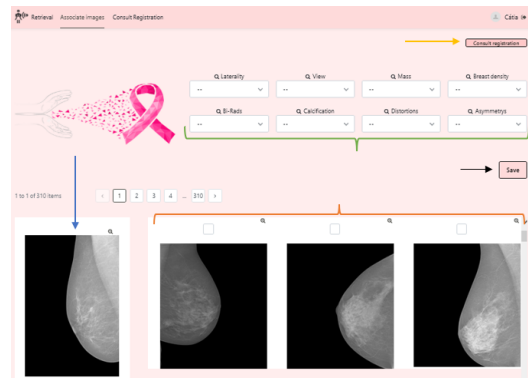


Figure 5: Mammogram retrieval page

After the association procedure is done, the expert is led to a new page (see Figure 6), where all records made can be viewed in detail. The probe image and the image selected by the expert as most relevant, along with their dimensions, are displayed on this page.



Figure 6: Registry example

6 Conclusions and directions for Future Work

Breast cancer is a widespread illness that affects a large number of individuals, with women being at a higher risk of developing it. Despite advancements in medical knowledge and treatment options, there remains a crucial need for ongoing research in this field. This urgency arises from the fact that breast cancer continues to claim the lives of thousands of individuals each year. By investing in further research efforts, we can strive to improve prevention, early detection, and treatment strategies, ultimately reducing the devastating impact of this disease on countless lives. The present project is related to the aforementioned concerns since early detection of breast cancer is essential in achieving a favourable prognosis. A software tool was developed to gather data, thereby facilitating the creation of a mammography retrieval system [4]. This software was developed according to initial requirements, and subsequently refined, to have a final quality version to facilitate the specialist's work. The disadvantages of the software are that it is not constantly fed with new mammogram images and has not yet been tested in a clinical environment. In terms of future prospects, the primary focus centres on software enhancement, specifically the creation of a version suitable for clinical application.

References

- [1] M. Arnold, E. Morgan, H. Rumgay, A. Mafra, D. Singh, M. Laversanne, J. Vignat, JR Galow, F. Cardoso, S. Siesling, and I Soerjomataram. Current and future burden of breast cancer: Global statistics for 2020 and 2040. *The Breast*, 66:15–23, 2022.
- [2] OutSystem. Low-code de alto desempenho para desenvolvimento de aplicativos, 2018. URL <https://www.outsystems.com/pt-br/>. Access date: 2023-07-27.
- [3] PenRad. PenFetch Automatic Smart Image Transfer and More. Eliminate Waste., 2008. URL http://penrad.comcastbiz.net/pdf/PENFETCH_2%20Page.pdf. Access date: 2023-08-07.
- [4] Cátia Roriz. Retrieval de mamografias. Master's thesis, Polytechnic University of Coimbra, 2023 (to be submitted).
- [5] V. Sharma. Mammogram Image Retrieval System Using Texture and Semantic Features. *J of Physics: Conf Ser*, 2267(1):012071, 2022.