SYSTEM FOR SEGMENTATION AND CLASSIFICATION OF DENSE TISSUE IN MAMMOGRAPHIC IMAGES USING MACHINE LEARNING

MARIA ANGELICA BRAVO BRAVO
NATALIA JOHANA CABEZA GUTIÉRREZ

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECÁNICAS
ESCUELA DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA Y DE TELECOMUNICACIONES
BUCARAMANGA
2023

SYSTEM FOR SEGMENTATION AND CLASSIFICATION OF DENSE TISSUE IN MAMMOGRAPHIC IMAGES USING MACHINE LEARNING

MARIA ANGELICA BRAVO BRAVO
NATALIA JOHANA CABEZA GUTIÉRREZ

A dissertation submitted in partial fulfillment of the requirements for the degree of electronic engineer

Advisor:
SAID DAVID PERTUZ ARROYO
INGENIERO ELECTRÓNICO. PhD.

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECÁNICAS
ESCUELA DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA Y DE TELECOMUNICACIONES
BUCARAMANGA
2023

ACKNOWLEDGEMENTS

I would like to thank God for allowing me to have arrived here. For giving me the necessary strength to give my best every day in everything I do. I am grateful to my family for their support, especially my mother, who was always there for me with kind words that kept me going through tough times. To my grandfather, who was like a father to me, despite his ailments and his years, every time he asked me how I had performed with each of my activities. I will never forget all the times he sat with me and heard everything I had to tell him. Thanks to the rest of my family members for believing in me and giving me the confidence to fight for my dreams. I also want to express my gratitude to my classmates, without them the university experience would not have been the same. To all those who passed through my life for a short or long time, I will always remember you. Thanks for making the way more enjoyable, for being there in the good and bad moments, for the laughs and for the unconditional support. I learned many things thanks to each one of you, who left a trace in my heart. To my professors, who made me feel a strong interest in their classes because they really enjoyed their job, my absolute admiration and respect. Thanks for teaching me that life is challenging, but with effort and dedication, everything is possible. Many thanks to all for allowing me to have a unique experience in each semester as I climbed into my university career. Of course, I also want to say thanks to my colleague and dear friend Natalia, whom I had the pleasure of sharing many experiences within this career. Even before we started our first semester, I knew she would be a support to me. For all the classes we took together and the ones we didn't, we completed them in the best way. And, here we are today, concluding our undergraduate project together. I feel so proud of all that I have achieved over these years, and of all that I have grown personally. My most sincere gratitude to all those who accompanied me during this great adventure.

-Maria Angelica Bravo

I would like to thanks my parents, without them it could not have been possible to achieve my goals and finish my career. They are my support and the people who inspire me the most. My mom, Hersilia, has been there every moment, and she is the person that understands me the most. She was there in my ups and downs boosting me with confidence. My dad, Victor, even if he did not understand what I studied, he gave me all his support. I am grateful to my classmates, they made all this experience more comfortable and enjoyable. I would like to thanks Jeison for being understanding, loving and patient during all the difficult times. I am also grateful with Maria Angelica, we both make a great team and obtained good results on our project. We met before starting our career and have been friends during all this time. I am happy we both got to finish this path together obtaining better results than expected. Finally, I would also like to thanks my siblings Victor Manuel, Arbelay, Deisy and Sebastian for being there for me. They took part of their times for taking me to the university, for explaining me things that I did not understand and for giving me more confidence during this journey. Thanks to all these beautiful people, family and friends, for their love and support.

-Natalia Cabeza

Dedicated to our families and friends.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

RESUMEN

TÍTULO: SISTEMA PARA LA SEGMENTACIÓN Y CLASIFICACIÓN DEL TEJIDO DENSO EN IMÁGENES MAMOGRÁFICAS USANDO APRENDIZAJE AUTOMÁTICO *

AUTORES: MARIA ANGELICA BRAVO BRAVO, NATALIA JOHANA CABEZA GUTIÉRREZ **

PALABRAS CLAVE: APRENDIZAJE AUTOMÁTICO, CÁNCER, CLASIFICACIÓN, DENSIDAD DEL SENO, INTERFAZ DE SOFTWARE, SEGMENTACIÓN.

DESCRIPCIÓN:
La densidad mamaria es uno de los factores de riesgo más importantes para estimar el cáncer de seno. Una mayor densidad indica una mayor probabilidad de desarrollar esta enfermedad y además, hace que sea más propenso que un radiólogo pase por alto lesiones pequeñas. Actualmente, existen técnicas basadas en aprendizaje automático y aprendizaje profundo para estimar la densidad mamaria. Sin embargo, la mayoría de estas técnicas se encuentran en etapa de desarrollo, lo que dificulta su uso por no expertos en el área de análisis computacional de imágenes. El objetivo de este trabajo es integrar desarrollos recientes encontrados en el estado del arte a través de una interfaz de software independiente que pueda ser usada por médicos para la segmentación y estimación de la densidad mamaria en imágenes de mamografía digital de campo completo. La interfaz de software se diseñó usando *Qt Designer* para aspectos gráficos y el lenguaje de programación Python, específicamente su módulo *PyQt5*, para el manejo de estas herramientas. Se realizó un estudio en lectores no expertos con cuarenta participantes para evaluar el desempeño de la interfaz de software diseñada. El análisis cuantitativo de los resultados de segmentación obtenido por los usuarios con las segmentaciones manuales hechas por lectores expertos obtuvo PD-errors entre 7.5% y 10.1%. Además, se evaluaron aspectos cualitativos de la interfaz, tales como disposición, eficiencia y satisfacción del usuario, concluyendo que el software desarrollado es considerado intuitivo, fácil de usar, completo y se desempeña de acuerdo a las expectativas del usuario. Con este trabajo, se pretende servir de base para una futura validación clínica de biomarcadores basados en densidad para la evaluación del riesgo de cáncer de seno.

---

* Tesis de Pregrado

** Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones. Director: Said David Pertuz Arroyo. PhD.

ABSTRACT

TITLE: SYSTEM FOR SEGMENTATION AND CLASSIFICATION OF DENSE TISSUE IN MAMMOGRAPHIC IMAGES USING MACHINE LEARNING *

AUTHORS: MARIA ANGELICA BRAVO BRAVO, NATALIA JOHANA CABEZA GUTIÉRREZ **

KEYWORDS: BREAST DENSITY, CANCER, CLASSIFICATION, MACHINE LEARNING, SEGMENTATION, SOFTWARE INTERFACE.

DESCRIPTION:
Breast density is one of the strongest risk factors for breast cancer. Higher density indicates a major probability of developing the disease and also increases the likelihood of radiologists missing small lesions. Currently, there are techniques based on machine learning and deep learning for estimating breast density. However, most of these techniques are in the development stage, which hinders their utilization by clinicians without experience in computational image analysis. The aim of this work is to integrate recent developments in the state of the art through a standalone software interface that can be used by clinicians for breast density segmentation of full field digital mammograms. We designed this software interface using *Qt Designer* for graphical aspects and Python programming language, specifically its module *PyQt5*, for managing these tools. We conduct a non-expert reader study with forty subjects for evaluating the performance of the designed software interface. Quantitative analysis comparing the segmentation results obtained by the users with ground-truth segmentations obtained from experts readers yielded PD-errors between 7.5% and 10.1%. We also evaluated our interface based on qualitative aspects such as disposition, efficiency and user satisfaction, and concluded that the developed software is considered intuitive, easy to use, complete and performed as users expected. This work aims to serve as a baseline for further clinical validation of density-based biomarkers for breast cancer risk assessment.

---

* Undergraduate Thesis

** Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones. Advisor: Said David Pertuz Arroyo. PhD.

1. INTRODUCTION

Cancer is the leading cause of death worldwide. According to the World Health Organization, the most common types of cancer worldwide are breast, lung, colorectal, prostate, skin and gastric cancer [1]. In the last few years, about 7.8 million of women were diagnosed with breast cancer, making it the world's most prevalent cancer [2]. Early detection is important for an opportune treatment and increasing the possibilities of survival [3].

One of the most common and cost-effective techniques for detecting breast cancer is the mammography; approximately 80% and 85% of cases are detected employing this technology [4]. Also, this imaging method has some advantages such as its low cost, easy acquisition and its widespread use as a primary diagnosis and screening tool. Nowadays, one of the most robust and widely used mammography modalities is Full Field Digital Mammography (FFDM).

Currently, the density of the breast tissue is one of the most important biomarkers used for assessing the risk of developing breast cancer [5]. This measurement, illustrated in

---

[1] WORLD HEALTH ORGANIZATION. *Cancer*. 2022. URL: `https://www.who.int/news-room/fact-sheets/detail/cancer`.

[2] WORLD HEALTH ORGANIZATION. *Breast cancer*. 2023. URL: `https://www.who.int/news-room/fact-sheets/detail/breast-cancer`.

[3] Osama M KORIECH. "Breast cancer and early detection". In: *Journal of family & community medicine* 3.1 (1996), p. 7.

[4] Arnau OLIVER et al. "A review of automatic mass detection and segmentation in mammographic images". In: *Medical image analysis* 14.2 (2010), pp. 87–110.

[5] Jeff WANG et al. "Automatic estimation of volumetric breast density using artificial neural network-

Fig. 1, is the ratio between the proportion of the parenchymal tissue and the breast's total area. In practice, the estimation of this biomarker is often performed using the BI-RADS (Breast Imaging Reporting and Data System) rating scale which provides a semi qualitative result based on the radiologist's experience [6]. Nevertheless, the emergence of Computer Aided Diagnosis techniques (CAD) has allowed the estimation of this biomarker, providing automatic or semiautomatic quantitative results. This brings some advantages in terms of time and reproducibility since the results are more consistent, independently of the reading radiologist. According to previous studies, most of the techniques used for the estimation of breast density are based on machine learning algorithms, which have shown promising results [7].

Although there are clinically validated algorithms for the identification of dense tissue, the adoption of such systems is difficult in clinical practice. Due to this, the aim of this work is to facilitate the use of these machine learning algorithms through a software interface easy to use for anyone interested in estimating breast density.

---

based calibration of full-field digital mammography: feasibility on Japanese women with and without breast cancer". In: *Journal of digital imaging* 30 (2017), pp. 215–227.

[6] Mary C SPAYNE et al. "Reproducibility of BI-RADS breast density measures among community radiologists: a prospective cohort study". In: *The breast journal* 18.4 (2012), pp. 326–333.

[7] Brad M KELLER et al. "Preliminary evaluation of the publicly available Laboratory for Breast Radiodensity Assessment (LIBRA) software tool: comparison of fully automated area and volumetric density measures in a case–control study with digital mammography". In: *Breast cancer research* 17 (2015), pp. 1–17.

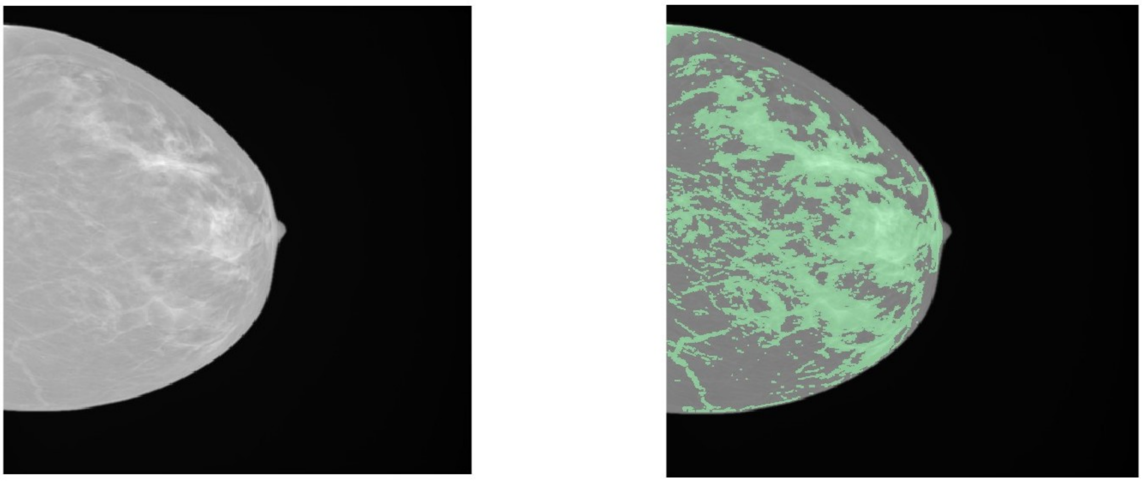

Figure 1. Breast density. (a) Original FFDM image. (b) Segmented image. The green area corresponds to dense tissue.

2. PREVIOUS RELATED WORK

Before the appearance of algorithms based on deep learning, the estimation of density was mainly performed through feature-based machine learning techniques, such as Gabor Filters and oriented gradient histogram analysis [8].

One of the alternatives for obtaining breast density results based on machine learning is proposed in Rampun et. al [9] approach. This algorithm employs Local Septenary Patterns (LSP) for making a feature extraction of the region of interest, then feature selection is approached through multi-topologies, and the classification process is executed by a Support Vector Machine (SVM). This LSP technique showed an accuracy of 86.9%, which can be considered as a good result.

Tlusty et. al in [10] proposed an unsupervised analysis for unlabeled medical images which consists of feature learning by Stacked Auto-Encoders, K-means clustering for building a data model and patch clustering for semantic classification.

Other techniques involve image processing, such as the one proposed in Patra et. al

---

[8] Hector LOPEZ-ALMAZAN et al. "A deep learning framework to classify breast density with noisy labels regularization". In: *Computer Methods and Programs in Biomedicine* 221 (2022), p. 106885.

[9] Andrik RAMPUN et al. "Breast density classification using local septenary patterns: a multi-resolution and multi-topology approach". In: *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE. 2019, pp. 646–651.

[10] Tal TLUSTY, Guy AMIT, and Rami BEN-ARI. "Unsupervised clustering of mammograms for outlier detection and breast density estimation". In: *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE. 2018, pp. 3808–3813.

approach [11]. This algorithm performed successfully even on the mediolateral oblique (MLO) view of mammograms, which is one of the two standard mammographic views and provides visualization of the greatest amount of breast tissue, including the posterior breast and axillary region [12]. The proposed algorithm employs thresholding for obtaining a binarized image, morphological operators, and area filtering for segmenting the dense tissue. The algorithm developed in [13] uses histograms, gray-level co-occurrence matrix (GLCM) and gray-level run length matrix (GLRLM) for feature extraction and a double SVM for breast density classification.

More recently, with the rise of deep learning, convolutional networks have been adopted for identifying the dense tissue within mammograms, leading to promising results. In the literature, the following architectures have been evaluated: U-net [14], DualViewNet [15], MobileNetV2 [15], VGG-like [16] and Inception-V3 [17]. Most of these architectures have

---

[11] Ankita PATRA et al. "Automated Breast Density Assessment using Image Processing Techniques". In: *2023 4th International Conference for Emerging Technology (INCET)*. IEEE. 2023, pp. 1–4.

[12] C. JU and L. DOEPKE. *Positioning and Technique*. URL: https://www.uclahealth.org/departments/radiology/education/breast-imaging-teaching-resources/screening-mammogram/positioning-and-technique.

[13] Yi-Chong ZENG. "Mammogram density classification using double support vector machines". In: *2018 IEEE 7th Global Conference on Consumer Electronics (GCCE)*. IEEE. 2018, pp. 547–550.

[14] Andrés LARROZA et al. "Breast Dense Tissue Segmentation with Noisy Labels: A Hybrid Threshold-Based and Mask-Based Approach". In: *Diagnostics* 12.8 (2022), p. 1822.

[15] Timothy COGAN and Lakshman TAMIL. "Deep understanding of breast density classification". In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE. 2020, pp. 1140–1143.

[16] Mickael TARDY, Bruno SCHEFFER, and Diana MATEUS. "Breast density quantification using weakly annotated dataset". In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE. 2019, pp. 1087–1091.

[17] Conrad TESTAGROSE et al. "Impact of Concatenation of Digital Craniocaudal Mammography Images on a Deep-Learning Breast-Density Classifier Using Inception-V3 and ViT". in: *2022 IEEE*

a common structure where noise filtering, background and pectoral muscle elimination is the first step. After this, the FFDM image passes through convolutional layers for a feature extraction and the output of these layers goes into a classification model where the prediction results are obtained.

Unfortunately, there are limitations with the algorithms mentioned above. Most of them are not clinically validated, and are also not publicly available. Besides, the results of these algorithms are not displayed in a format easy to understand for non-experts in the computer science area, such as the medical personnel which is the real target of these developments.

International Conference on Bioinformatics and Biomedicine (BIBM). IEEE. 2022, pp. 3399–3406.

3. METHODS AND MATERIALS

3.1. Dataset

In this work, we used a dataset of 582 craniocaudal view FFDM images from the screening program of Tampere University Hospital, in Finland. These images correspond to mammograms from women aged between 50 and 68 years old and have been collected between years 2014 and 2017 [18]. A total of 244 mammograms were captured with a Philips MicroDose SI system (Philips Digital Mammography, Sweden) and 338 were captured with a GE Senographe Essential (General Electric Medical Systems, USA) mammography system [19]. Moreover, this set of images was manually segmented by expert radiologists in order to obtain ground-truth masks of the dense tissue.

3.2. Selection of algorithms

We performed a review of the state of the art in breast density estimation algorithms for digital mammography.

We selected algorithms according to the following criteria: i) availability of source code and licenses allowing their use for research purposes, and ii) being clinically validated with annotations from human experts. With these criteria, we considered three algo-

---

[18] German F TORRES et al. "Morphological area gradient: System-independent dense tissue segmentation in mammography images". In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2019, pp. 4855–4858.

[19] Carlos S BENITEZ et al. "Automatic Dense Tissue Segmentation in Digital Mammography Images Based on Fully Convolutional Network and Intensity-Based Clustering". In: *2022 IEEE Colombian Conference on Applications of Computational Intelligence (ColCACI)*. IEEE. 2022, pp. 1–4.

rithms based on deep learning (Deep-LIBRA [20], MammoDL [21], and MULTILABEL [22]), two algorithms based on image processing and machine learning (MAG [18] and LIBRA [23]), and one algorithm combining both deep learning and image processing (HYBRID [19]). These algorithms are summarized in Table 1.

Among the algorithms listed in Table 1, three algorithms were discarded from further analysis: Deep-LIBRA [20], MammoDL [21] and MULTILABEL [22]. The first one was not included because it did not achieve proper segmentation of dense tissue on the FFDM images of the selected database. In some images this algorithm only identified the breast contour and not parenchymal tissue. MammoDL did not include the trained models to perform such tests in its repository. Finally, MULTILABEL [22] did not perform properly on FFDM images, arguably because this algorithm was designed for digitized mammograms. However, it is incorporated as an extra in our software interface, but is not included in the analysis. Selected algorithms are described in more detail in Appendix section 5.

---

[20] Omid Haji MAGHSOUDI et al. “Deep-LIBRA: An artificial-intelligence method for robust quantification of breast density with independent validation in breast cancer risk assessment”. In: *Medical image analysis* 73 (2021), p. 102138.

[21] Ramya MUTHUKRISHNAN et al. “MammoDL: mammographic breast density estimation using federated learning”. In: *arXiv preprint arXiv:2206.05575* (2022).

[22] VM TIRYAKI and V KAPLANOĞLU. “Deep learning-based multi-label tissue segmentation and density assessment from mammograms”. In: *IRBM* 43.6 (2022), pp. 538–548.

[23] Brad M KELLER et al. “Estimation of breast percent density in raw and processed full field digital mammography images via adaptive fuzzy c-means clustering and support vector machine segmentation”. In: *Medical physics* 39.8 (2012), pp. 4903–4917.

Table 1. Algorithms complying with our inclusion criteria.

Acronym	Title
Deep-LIBRA	Deep-LIBRA: An artificial-intelligence method for robust quantification of breast density with independent validation in breast cancer risk assessment
MammoDL	MammoDL: Mammographic Breast Density Estimation using Federated Learning
MULTILABEL	Deep Learning-Based Multi-Label Tissue Segmentation and Assessment from Mammograms
MAG	Morphological Area Gradient: System independent dense tissue segmentation in mammography images
LIBRA	Laboratory for Individualized Breast Radiodensity Assessment
HYBRID	Automatic Dense Tissue Segmentation in Digital Mammography Images Based on Fully Convolutional Network and Intensity-Based Clustering

3.3. Graphical user interface

For the design of the software interface, we decided to use Python programming language and make it a desktop application, due to its advantages in terms of performance and speed [24], offline access, licensing, more control over the installation by users and privacy regarding local storage of data.

[24] M. CATALÁN. *METODOLOGÍAS DE EVALUACIÓN DE INTERFACES GRÁFICAS DE USUARIO*. URL: http://eprints.rclis.org/6732/1/Metodologias_de_evaluaci%C3%B3n_de_interfaces_graficas_de_usuario.pdf.

Graphical aspects of the interface were done with *Qt Designer* which is framework widely used for the multiplatform development of programs with user graphic interfaces. It is characterized by its high speed and works on different operative systems. In order to access the tools designed and available in *Qt Designer*, we used *PyQt5* which is a bridge from the Qt graphical library to the Python programming language, since it allows us to use all the tools provided by Qt through this language.

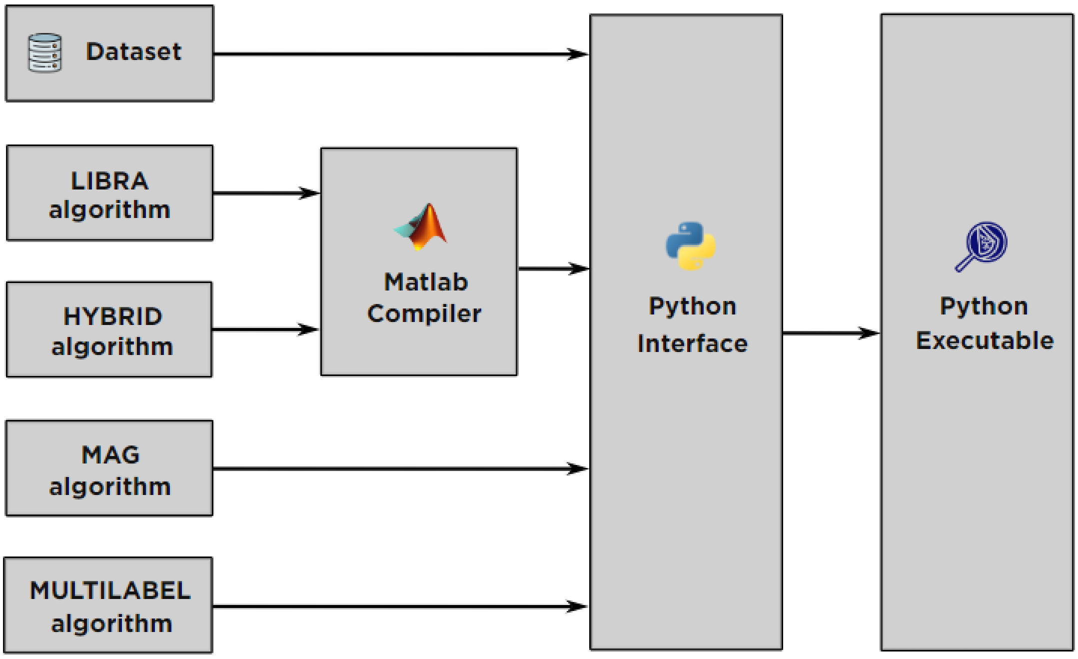

Figure 2. System interconnection block diagram.

For incorporating the algorithms to the software interface, we adapted three of these for using them in Python programming language. As shown in Fig. 2, for LIBRA [23] and HYBRID [19] algorithms, we had to use Matlab compiler application since both were originally implemented using Matlab coding. Initially, MAG [18] algorithm was coded in Matlab, thus we performed a conversion to Python. Consequently, MAG [18] and MUL-

TILABEL [22] algorithms were incorporated directly in the software interface since these were already coded in Python programming language.

As design criteria, we considered aspects such as utility [24] which focuses on whether the final user thinks if the tool is appropriate for estimating breast density. We also took into account effectiveness, in terms of computing resources and execution time, and learning ease. The software interface was designed so that the user sees a menu that allows to select the algorithm to be used. The user then loads the image and clicks a button to perform the density estimation using the selected algorithm as shown in Fig. 3.

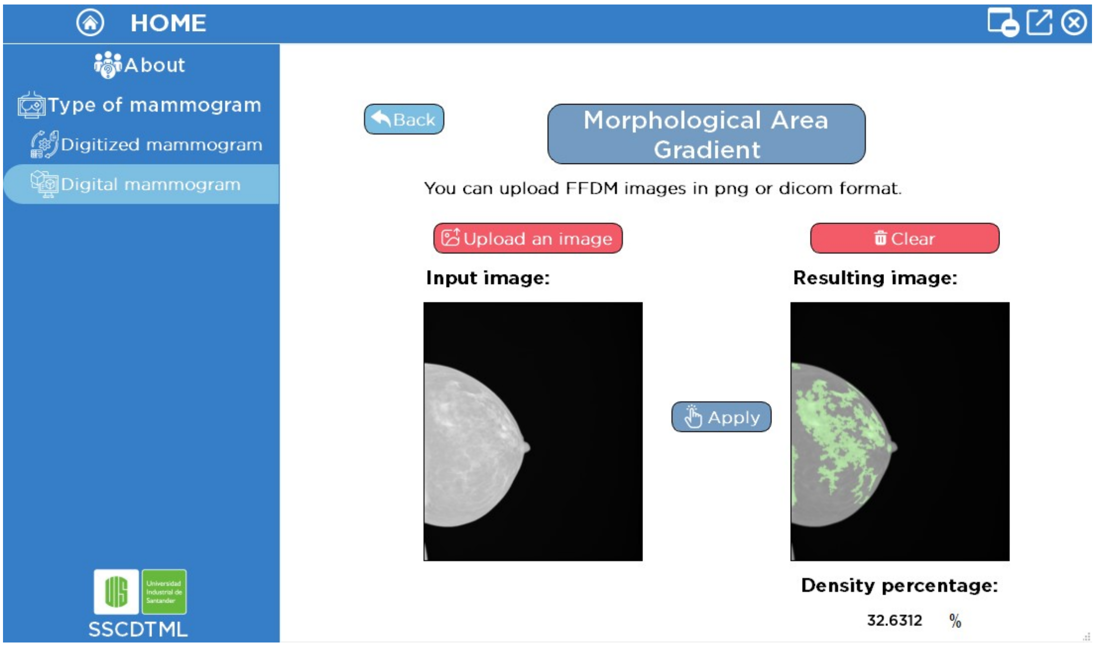

Figure 3. Software interface design.

3.4. Quantitative analysis

We performed two stages of quantitative analysis, the first is a *fully-automatic evaluation* of the performance of the algorithms in all the images available in the selected dataset. As evaluation criteria, we considered two measures: Dice Similarity Coefficient (DSC) and PD-error. Both of these measures are used for evaluating the performance of the segmentation task, comparing the obtained results with their respective reference or ground-truth values [25].

The DSC focuses on measuring the similarity between the two binary masks, it ranges from 0, indicating no overlap, to 1, indicating perfect overlap [26]. The DSC is defined as:

$$\text{DSC}(A, B) = \frac{2(A \cap B)}{A + B} \tag{1}$$

Where A and B are the target regions and $\cap$ is the intersection between both segmentations.

The PD-error is the median difference between the density percentage of the observations (ground-truth) and model output (predictions) [27]. Mathematically, the PD-error is expressed as:

$$\text{PD-error} = \text{median}(|Y_{gt} - Y_{pred}|) \tag{2}$$

Where Y_{gt} is the ground-truth density percentage value and Y_{pred} is the model prediction.

[25] Kelly H ZOU et al. "Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports". In: *Academic radiology* 11.2 (2004), pp. 178–189.

[26] *Dice Score*. URL: https://oecd.ai/en/catalogue/metrics/dice-score.

[27] INSIDE LEARNING MACHINES. *Median Absolute Error - Inside Learning Machines*. 2023. URL: https://insidelearningmachines.com/median_absolute_error/.

For the second stage, we performed an *interactive evaluation* for analysing quantitative aspects on our software interface. This evaluation was conducted as follows: twenty randomly selected FFDM images were provided to non-expert readers, they estimated the density percentage with the algorithm of their choice and registered the obtained value. These outcomes were subsequently studied to analyze the variability between the algorithms and the density percentage values given in the ground-truth, computing the PD-error per algorithm. We also registered the algorithm used by the non-expert reader per image, for determining which algorithm was the most used.

3.5. Qualitative analysis

For these qualitative aspects, we applied a methodology for the evaluation of graphical user interface [24]. It contains twenty questions that include aspects such as utility, interface consistency, effectiveness, learning ease, reliability and user satisfaction.

With the utility criteria, we meant to evaluate if our software interface allows the user to achieve their objective, which in this case is to segment and calculate the density percentage in a FFDM. This criteria also involves other aspects such as effectiveness, efficiency, consistency and satisfaction. For evaluating the effectiveness, we include questions about software and hardware, verifying that the user completes the task with minimum time and computational resources [28], and ensuring that no errors occur during this process.

Efficiency and satisfaction aspects are more related to the user perception. For these criteria, we evaluated whether it was easy and intuitive for the user to adapt to the

[28] N. MADRID. *Métricas de usabilidad y experiencia de usuario (UX)*. 2020. URL: https://www.nachomadrid.com/2020/01/metricas-de-usabilidad-y-experiencia-de-usuario.

software interface, and whether learning to use it took a large amount of time. Finally, for consistency, we considered whether the icons, menus and structure really reflected the purpose of each of the elements.

4. RESULTS

4.1. Quantitative results

For a total of forty participants in the non-expert reader survey, we registered the PD-error per algorithm. The results are shown in Table 2 with 95% confidence intervals. Fully-automatic values correspond to the results obtained in the evaluation using 582 images of the dataset, and interactive ones correspond to the quantitative results obtained using the twenty images provided to the non-expert readers.

Table 2. Fully-automatic and interactive evaluation results. Performance measures with 95% confidence intervals in parenthesis.

Algorithm	**DSC**		**PD-error (%)**	
	Fully automatic	**Interactive**	**Fully automatic**	**Interactive**
LIBRA	0.740 (0.725, 0.755)	-	8.4 (7.6, 9.2)	7.5 (5.7, 9.2)
HYBRID	0.734 (0.724, 0.745)	-	9.3 (8.7, 9.9)	8.4 (7.1, 9.7)
MAG	0.738 (0.722, 0.754)	-	7.4 (6.8, 8.0)	10.1 (9.6, 10.6)

According to the results obtained in the objective evaluation, the performance of these algorithms was similar and there were variations with the validation results reported by the authors. LIBRA [23] yielded the highest DSC value. However, MAG [18] algorithm had the lowest PD-error and was the fastest to run.

As reported in Table 2, subjective results shown that LIBRA [23] has the lowest PD-error value. These outcomes showed that there is a variability between the algorithms and the ground-truth values, and also between each algorithm. By comparing the objective and subjective results, there are some variations in the PD-error values since only a small subset of the dataset was selected for the second stage of quantitative analysis.

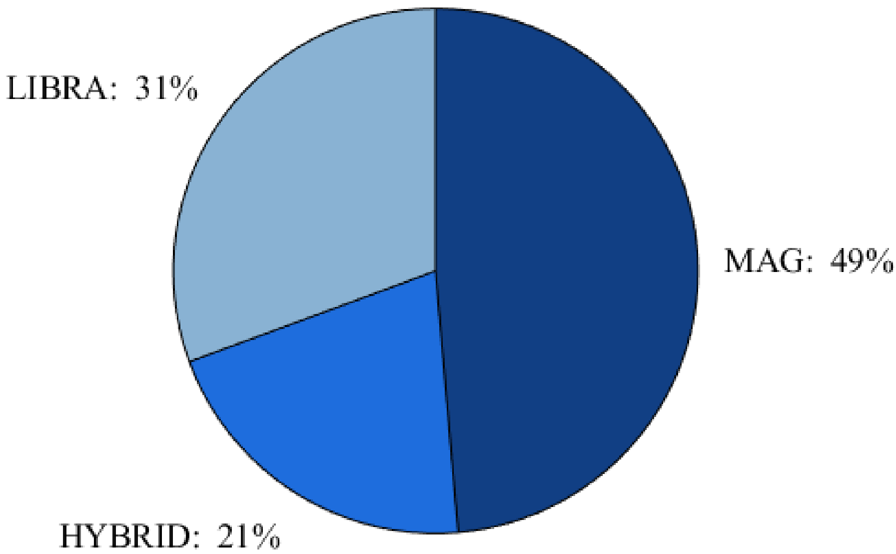

Figure 4. Algorithms preferred by users.

After the user navigated through the software interface and made use of each of the algorithms, we recorded the algorithm used for the density estimation of each of the images. With this we determined the algorithm most preferred by the users, being MAG the most used, preceded by LIBRA [23] and finally HYBRID [19]. Fig. 4 shows these results with their respective percentages.

4.2. Qualitative results

In Fig. 5, we show qualitative results related to the software interface disposition, more exactly to the understanding of the information and menu of the designed tool. We obtained good results in this aspect since most of the users found our software interface easy to understand and use.

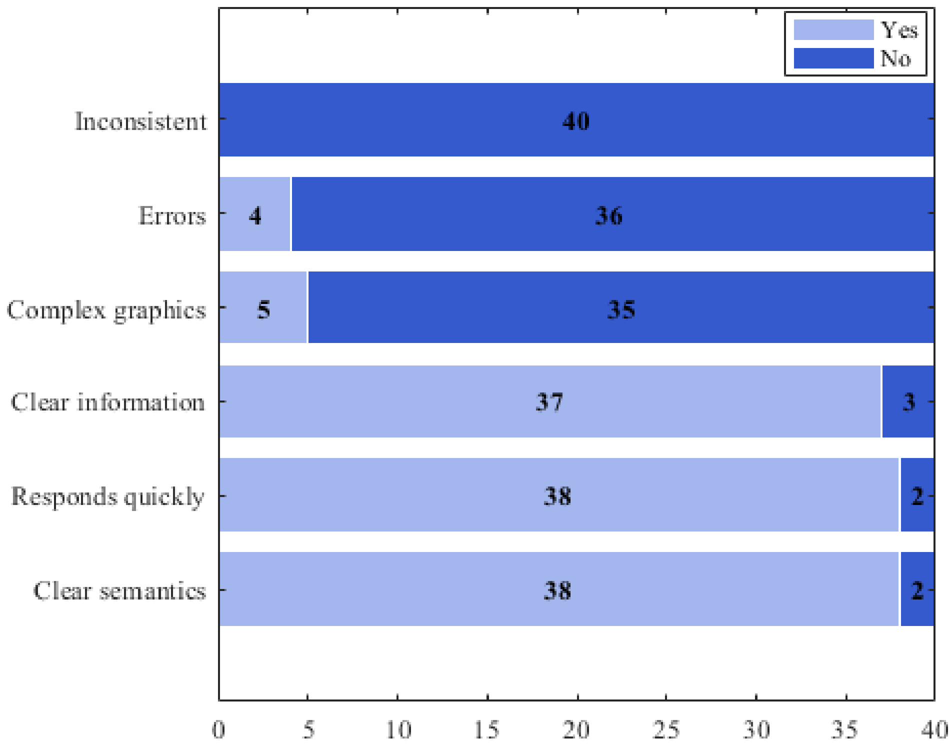

Figure 5. Results of aspects related to the disposition and effectiveness of the software interface.

Fig. 5 also shows the results obtained for the effectiveness. This includes errors during execution, software interface consistency, response time and whether the interface has complex graphics. As shown, all users reported that our software interface is consistent and most of them think that it responds quickly to actions and did not found complex graphics and errors during the execution.

For aspects of user perception, we reported results of satisfaction and efficiency. In Fig. 6, we illustrated as a linear scale from 1 to 5 how satisfactory and intuitive users found working with the software interface. More than half of all users considered that it was a great experience, since its design was complete and it could be used without instructions.

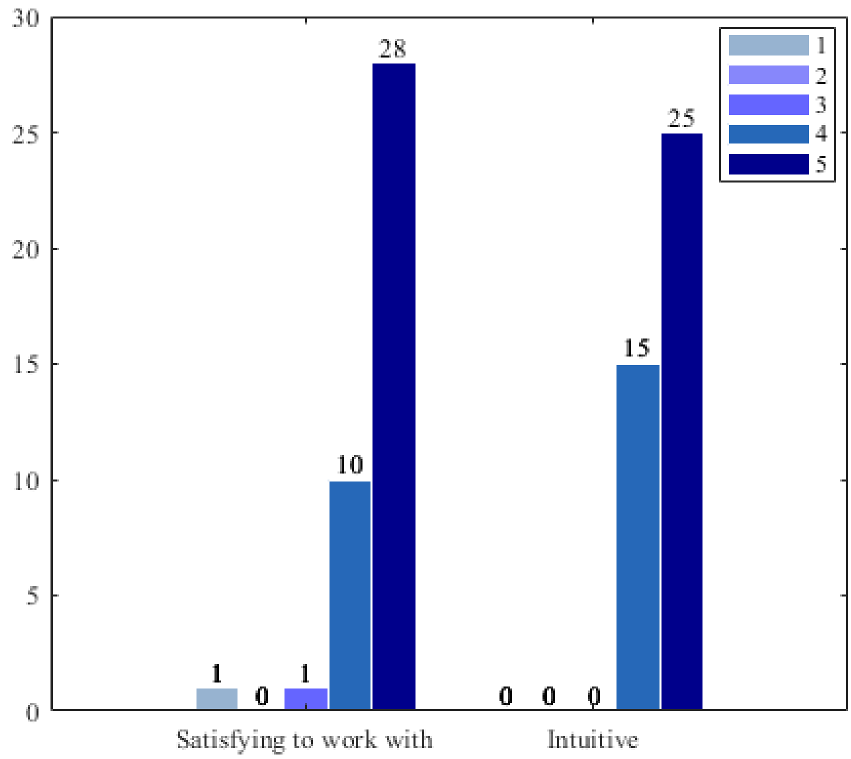

Figure 6. Results of aspects related to the satisfaction of the software interface.

Fig. 7 shows that about 25 users learn how to use the interface in less than five minutes, showing that the software interface is easy to use. All of them considered that it performed the estimation of breast density as expected and the majority did not find it a frustrating experience. On the other hand, more than half perceived that the software interface takes some time to execute the algorithms.

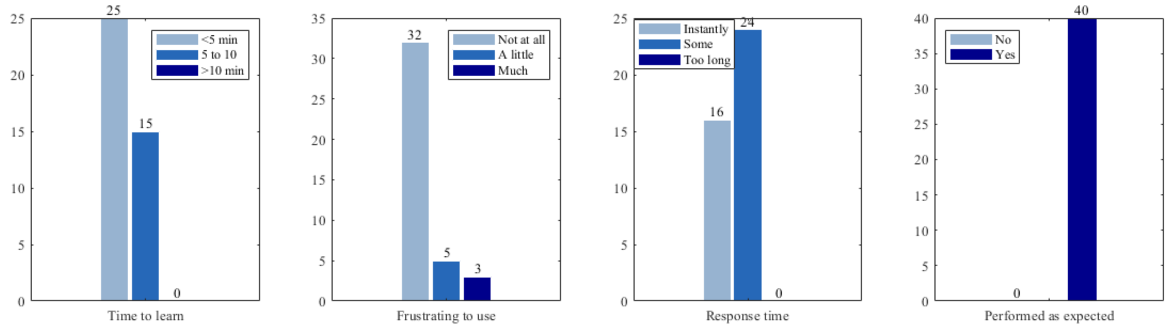

Figure 7. Results of aspects related to the efficiency of the software interface.

Table 3. Performance results in terms of percentage.

Performance	Aspect
100%	Information presented clearly, menu semantics clear, perform as expected, software consistency
80 - 99.9%	Software interface did not make any error, understanding of options, icons and buttons meanings, software responds quickly to actions, use without instructions, software is not a frustrating experience, software does not have complex graphics
$<$ 80%	Learn to use in less than five minutes, software interface takes some time to execute the algorithms

In Table 3 we summarize the performance of the questions we asked for evaluating qualitative aspects in terms of percentage. The survey can be seen in Appendix section 5.

5. DISCUSSIONS AND CONCLUSIONS

We implemented a software interface designed in Python programming language through its *PyQt5* module and *Qt Designer*, for assisting human readers in the estimation of breast density via computarized analysis of full-field mammographic images. The source codes for the interface are publicly available at `https://github.com/SSCDTML/SSCDTML_CODES.git`. We incorporated algorithms based on machine learning and deep learning techniques found in the state of art.

The selection of these algorithms was carried out according to their code availability and their performance in the selected database. We performed a preliminary evaluation of each algorithm in 582 mammograms and computed two performance measures: DSC and PD-error. From these tests, LIBRA [23] obtained the highest DSC score 0.740 (95% CI: 0.725-0.755). However, HYBRID [19] algorithm showed a smaller variability range. The best PD-error was obtained using MAG [18] algorithm: 7.4% (95% CI: 6.8-8.0).

For evaluating the performance of the proposed software interface, we conducted a non-reader expert study supported by a survey and took into account quantitative and qualitative aspects. For the quantitative ones, we computed the PD-error of each algorithm for quantifying the discrepancy between the ground-truth and the predicted values. LIBRA [23] algorithm obtained the lowest PD-error: 7.5% (95% CI: 5.7-9.2). On the other hand, we registered the algorithms selected by the users for the estimation of breast density and from all forty participants in the survey, 49% preferred using MAG [18] algorithm.

According to the results obtained from the qualitative test, we conclude that the designed software interface is an useful tool because the information and menu are clearly arranged and most users considered it intuitive. Besides, users found it efficient since it did not take much time to perform the task and has not complex graphics that may slow down computing time. Also, the majority of users considered that our tool was consistent and performed the estimation of breast density as expected. Based on that, we achieved our goal: non-experts in the computer science area can perform this task properly using our software interface.

The implementation of algorithms for the estimation of breast density from mammographic images involves a series of challenges that sometimes affect the success of the results. Among the main challenges is the need to build algorithms that combine the existing differences between image acquisition technologies and the lack of objective clinical validation for this task. This limited objective validation is due to the lack of a standard reference, since most of the datasets used for training models contain estimations performed based on the radiologist's experience.

With our software interface we aim to contribute to the access of different researchers to breast density segmentation tools. This is mainly because implementing algorithms by themselves can be time-consuming and requires knowledge in computer science area. With this work we are attempting to simplify and speed up the process of implementing algorithms for the segmentation and classification of dense tissue.

As future work, we plan to take into account the recommendations made by some of the users, such as the implementation of an option that allows selecting the language, English or Spanish, for navigating in our software. On the other hand, we will study the possibility of adding a tab that allows visualizing a direct comparison between the results obtained using the three algorithms available for FFDM images.

BIBLIOGRAPHY

BENITEZ, Carlos S et al. “Automatic Dense Tissue Segmentation in Digital Mammography Images Based on Fully Convolutional Network and Intensity-Based Clustering”. In: *2022 IEEE Colombian Conference on Applications of Computational Intelligence (ColCACI)*. IEEE. 2022, pp. 1–4 (cit. on pp. 18, 19, 21, 27, 31, 37).

CATALÁN, M. *METODOLOGÍAS DE EVALUACIÓN DE INTERFACES GRÁFICAS DE USUARIO*. URL: http://eprints.rclis.org/6732/1/Metodologias_de_evaluaci%C3%B3n_de_interfaces_graficas_de_usuario.pdf (cit. on pp. 20, 22, 24).

COGAN, Timothy and Lakshman TAMIL. “Deep understanding of breast density classification”. In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE. 2020, pp. 1140–1143 (cit. on p. 16).

Dice Score. URL: https://oecd.ai/en/catalogue/metrics/dice-score (cit. on p. 23).

INSIDE LEARNING MACHINES. *Median Absolute Error - Inside Learning Machines*. 2023. URL: https://insidelearningmachines.com/median_absolute_error/ (cit. on p. 23).

JU, C. and L. DOEPKE. *Positioning and Technique*. URL: https://www.uclahealth.org/departments/radiology/education/breast-imaging-teaching-resources/screening-mammogram/positioning-and-technique (cit. on p. 16).

KELLER, Brad M et al. “Estimation of breast percent density in raw and processed full field digital mammography images via adaptive fuzzy c-means clustering and support vector machine segmentation”. In: *Medical physics* 39.8 (2012), pp. 4903–4917 (cit. on pp. 19, 21, 26, 27, 31, 38).

KELLER, Brad M et al. “Preliminary evaluation of the publicly available Laboratory for Breast Radiodensity Assessment (LIBRA) software tool: comparison of fully automated area and volumetric density measures in a case–control study with digital mammography”. In: *Breast cancer research* 17 (2015), pp. 1–17 (cit. on p. 13).

KORIECH, Osama M. “Breast cancer and early detection”. In: *Journal of family & community medicine* 3.1 (1996), p. 7 (cit. on p. 12).

LARROZA, Andrés et al. “Breast Dense Tissue Segmentation with Noisy Labels: A Hybrid Threshold-Based and Mask-Based Approach”. In: *Diagnostics* 12.8 (2022), p. 1822 (cit. on p. 16).

LOPEZ-ALMAZAN, Hector et al. “A deep learning framework to classify breast density with noisy labels regularization”. In: *Computer Methods and Programs in Biomedicine* 221 (2022), p. 106885 (cit. on p. 15).

MADRID, N. *Métricas de usabilidad y experiencia de usuario (UX)*. 2020. URL: `https://www.nachomadrid.com/2020/01/metricas-de-usabilidad-y-experiencia-de-usuario` (cit. on p. 24).

MAGHSOUDI, Omid Haji et al. “Deep-LIBRA: An artificial-intelligence method for robust quantification of breast density with independent validation in breast cancer risk assessment”. In: *Medical image analysis* 73 (2021), p. 102138 (cit. on p. 19).

MUTHUKRISHNAN, Ramya et al. “MammoDL: mammographic breast density estimation using federated learning”. In: *arXiv preprint arXiv:2206.05575* (2022) (cit. on p. 19).

OLIVER, Arnau et al. “A review of automatic mass detection and segmentation in mammographic images”. In: *Medical image analysis* 14.2 (2010), pp. 87–110 (cit. on p. 12).

ORGANIZATION, WORLD HEALTH. *Breast cancer*. 2023. URL: `https://www.who.int/news-room/fact-sheets/detail/breast-cancer` (cit. on p. 12).

— *Cancer*. 2022. URL: `https://www.who.int/news-room/fact-sheets/detail/cancer` (cit. on p. 12).

PATRA, Ankita et al. “Automated Breast Density Assessment using Image Processing Techniques”. In: *2023 4th International Conference for Emerging Technology (INCET)*. IEEE. 2023, pp. 1–4 (cit. on p. 16).

RAMPUN, Andrik et al. “Breast density classification using local septenary patterns: a multi-resolution and multi-topology approach”. In: *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE. 2019, pp. 646–651 (cit. on p. 15).

SPAYNE, Mary C et al. “Reproducibility of BI-RADS breast density measures among community radiologists: a prospective cohort study”. In: *The breast journal* 18.4 (2012), pp. 326–333 (cit. on p. 13).

TARDY, Mickael, Bruno SCHEFFER, and Diana MATEUS. “Breast density quantification using weakly annotated dataset”. In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE. 2019, pp. 1087–1091 (cit. on p. 16).

TESTAGROSE, Conrad et al. “Impact of Concatenation of Digital Craniocaudal Mammography Images on a Deep-Learning Breast-Density Classifier Using Inception-V3 and ViT”. In: *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE. 2022, pp. 3399–3406 (cit. on p. 16).

TIRYAKI, VM and V KAPLANOĞLU. “Deep learning-based multi-label tissue segmentation and density assessment from mammograms”. In: *IRBM* 43.6 (2022), pp. 538–548 (cit. on pp. 19, 22, 38).

TLUSTY, Tal, Guy AMIT, and Rami BEN-ARI. “Unsupervised clustering of mammograms for outlier detection and breast density estimation”. In: *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE. 2018, pp. 3808–3813 (cit. on p. 15).

TORRES, German F et al. “Morphological area gradient: System-independent dense tissue segmentation in mammography images”. In: *2019 41st Annual International*

Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE. 2019, pp. 4855–4858 (cit. on pp. 18, 19, 21, 26, 31, 37).

WANG, Jeff et al. “Automatic estimation of volumetric breast density using artificial neural network-based calibration of full-field digital mammography: feasibility on Japanese women with and without breast cancer”. In: *Journal of digital imaging* 30 (2017), pp. 215–227 (cit. on p. 12).

ZENG, Yi-Chong. “Mammogram density classification using double support vector machines”. In: *2018 IEEE 7th Global Conference on Consumer Electronics (GCCE)*. IEEE. 2018, pp. 547–550 (cit. on p. 16).

ZOU, Kelly H et al. “Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports”. In: *Academic radiology* 11.2 (2004), pp. 178–189 (cit. on p. 23).

APPENDICES

APPENDIX A. Selected algorithms

• Automatic Dense Tissue Segmentation in Digital Mammography Images Based on Fully Convolutional Network and Intensity-Based Clustering (HYBRID) [19]:

This algorithm combines a U-Net Convolutional Neural Network (CNN), which is commonly used in biomedical image segmentation, and a filtering stage based on intensity clustering. The output of the CNN is a binary classification mask, where dense pixels are assigned a value of 1. This binary mask is the input of a clustering process where the segmentation is refined, selecting those regions that actually correspond to dense tissue. This step ensures that the segmentation does not contain spurious disconnected pixels caused mainly by the high heterogeneity of the fibroglandular tissue.

Originally in its design, this algorithm was validated in a dataset of 99 mammograms and authors reported a DSC of 0.795 and PD-error of 0.077.

• Morphological Area Gradient (MAG) [18]:

This algorithm quantifies the change in the segmented area as a function of the gray-level intensity. Firstly, an automatic detection of the background is applied for extracting the breast region. Then, the morphological area function is computed depending of the amount of segmented tissue which is obtained based on a gray-level threshold. For computing the optimal gray-level value for the segmentation, the Morphological Area Gradient function is obtained from the first derivative of the morphological area function and its maximum is the appropriate segmentation threshold. Finally, morphological operators are applied as a post processing step.

For validating this algorithm, authors conducted experiments in a dataset of 566 FFDM images and reported an absolute error of 7.6% and a DSC of 0.831.

• **Laboratory for Individualized Breast Radiodensity Assessment (LIBRA)** [23]**:**

This is a clinically validated and fully automated algorithm that delineates the breast region by using edge-detection techniques and applies fuzzy c-means clustering for partitioning the breast area into gray-level intensity clusters that are later used for the final dense tissue segmentation. LIBRA provides an estimate of the breast density percentage by computing the area of dense tissue and normalizing it by the total breast area results.

Authors validated this algorithm in a dataset of 324 images divided into raw and processed mammograms. For raw images, obtained a Pearson product-moment correlation coefficient, r, of 0.82 (CI: 0.76 - 0.86) and for processed images, r of 0.85 (CI: 0.80 - 0.89).

• **Deep Learning-Based Multi-Label Tissue Segmentation and Density Assessment from Mammograms (MULTILABEL)** [22]**:**

This method proposes a new fully-automated deep learning-based cascaded model which involves two main stages. In the first one, the segmentation of adipose, fibroglandular, and pectoral muscle tissues from digitized film mammograms was performed using a ResNet50-U-net CNN. The second one, a segmentation error elimination (SEE) algorithm consisting of connected component labeling (CCL) and morphological image processing operators was developed to improve the breast tissue segmentation results.

The quality of the mammogram segmentation labels, authors performed a validation in 296 mammograms achieving a DSC of 0.73.

APPENDIX B. Qualitative test: Survey

The designed survey contained the following questions:

1. Is the information shown clearly?

   (a) Yes

   (b) No

2. Are menu semantics clear?

   (a) Yes

   (b) No

3. Were you able to perform the task the way you expected?

   (a) Yes

   (b) No

4. Has the software made any errors or taken any actions that you did not expect?

   (a) Yes

   (b) No

5. Is there an option that you do not know what it means?

   (a) Yes

   (b) No

6. Have you found an icon that did not represent the action you expected?

   (a) Yes

   (b) No

7. Do you think the software interface is inconsistent?

(a) Yes

(b) No

8. Do you know the difference between the algorithm options for the two types of mammographic imaging?

(a) Yes

(b) No

9. Do you know the difference between the ”upload”, ”clear” and ”apply” buttons?

(a) Yes

(b) No

10. How satisfying is it to work with the software interface? (1 is the lowest and 5 is the highest)

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

11. Does the software interface respond quickly and appropriately to your actions?

(a) Yes

(b) No

12. How intuitive is the software interface to use? (1 is the lowest and 5 is the highest)

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

13. Do you think the software interface can be used without the need of instructions?

(a) Yes

(b) No

14. How long did it take you to learn how to use the software interface?

(a) Less than five minutes

(b) Between five and ten minutes

(c) More than ten minutes

15. Interacting with the software interface was a frustrating experience?

(a) Much

(b) Not at all

(c) A little

16. Did you have any problems running and using the software interface?

(a) Yes

(b) No

17. Does the software interface respond instantly to the action you provide?

(a) Responds instantly

(b) Takes some time

(c) Takes too long

18. Does the software interface have very complex graphics that may slow down its operation?

(a) Yes

(b) No

19. Rate the interface according to your satisfaction scale:

(a) User-friendly:

○ Low ○ Sufficient ○ Satisfactory ○ Good ○ Excellent

(b) Complete:

○ Low ○ Sufficient ○ Satisfactory ○ Good ○ Excellent

(c) Easy to use:

○ Low ○ Sufficient ○ Satisfactory ○ Good ○ Excellent

(d) Design:

○ Low ○ Sufficient ○ Satisfactory ○ Good ○ Excellent

20. What recommendations do you have to improve the user experience with the software interface?

____

____

____