MammoInsight: Innovating Early Breast Cancer Detection through Artificial Intelligence

Ernesto Rafael Perez[1]; Emilio Angelina[1]; José Leonardo Gómez Chávez[1]; German Conti[1]; Ramon Torres[2]; Nélida Peruchena[1]

[1] Laboratorio de Estructura Molecular y Propiedades, Instituto de Química Básica y Aplicada del Nordeste Argentino (IQUIBA-NEA; CONICET UNNE); [2] Universidad Nacional del Nordeste. rafaelperez@conicet.gov.ar

Abstract. The MammoInsight project aims to revolutionize the interpretation of digital mammographic images through the integration of artificial intelligence (AI) models. Facing the challenge of early and accurate breast cancer detection, this web platform seeks to overcome the subjectivity and heavy workload of specialists, significantly improving survival rates and accelerating the diagnostic process. Through the development and implementation of AI modules for the automatic categorization of breast density, the classification of mammograms, and the detection and segmentation of anomalies, this represents a crucial advancement in the diagnosis of breast pathologies and has a positive impact on the field of radiology and public health.

Keywords: Artificial Intelligence; Breast Cancer Diagnosis; Mammographic Images; Health Technology; Radiology.

1 Introduction

Breast cancer accounts for 16% of female cancers worldwide and is the leading cause of death from malignant tumors in women, according to the World Health Organization [1].

In Argentina, breast cancer was the most frequent in women with 22,024 cases (32.1% of the total) in 2020. The International Agency for Research on Cancer (IARC) estimated an age-adjusted incidence rate of 73.1 cases per 100,000 women for breast cancer in Argentina in 2020 [2].

Accurate mammogram interpretation, essential for the early detection of breast cancer, faces challenges such as erroneous diagnoses, estimated between 10-30%, and limitations in time and precision, exacerbated by the shortage of radiologists [3]. Recent studies in *Lancet Oncology* and *Radiology* in 2023 highlight how AI can match specialist detection rates, reduce errors, and lighten their workload, demonstrating its potential to transform mammographic diagnosis [4][5]. In Argentina, projects like the Artemisia network at the Italian Hospital of Buenos Aires show advances in the use of AI for breast density assessment, underscoring the importance of adopting these technologies for more efficient and early diagnosis [6]. It is important to note that, currently, there are no free or easily accessible tools in our country for health personnel that integrate functionalities to analyze different characteristics in mammograms

automatically. This gap in the availability of advanced medical interpretation technologies poses a significant challenge in healthcare.

The MammoInsight project aims to develop a web-based platform for analyzing digital mammographic images, assisted by artificial intelligence (AI) models, to achieve early and accurate breast cancer detection, thereby improving survival rates and accelerating the diagnostic process in the context of the high prevalence of this disease.

This article aims to present the current progress in the development of MammoInsight, highlighting its main components, methodologies used, and preliminary results obtained to date.

2 Methodology

The development platform focuses on analyzing three key aspects for breast cancer diagnosis, implemented in the following three modules:

1. **Breast Density Categorization:** Implementation and optimization of the convolutional neural network (CNN) “breast-density classification” pre-trained by the Mayo Clinic.
2. **Mammography Classification:** Development and improvement of classifiers based on CNN architectures such as DenseNet121 and ResNet50 to distinguish between benign and malignant mammograms, including Grad-CAM techniques to generate heatmaps on the image highlighting relevant sections and patterns. Through our participation in an AI competition on Kaggle [7] sponsored by the Radiological Society of North America, we gained access to DICOM data for training and validating the classification models.
3. **Anomaly Detection:** Creation of algorithms to identify microcalcifications using segmentation techniques like U-Net, currently being optimized and integrated into our platform. The dataset used: "Breast Micro-Calcifications". The goal is to expand the model's capabilities to automatically segment regions of interest in the image and classify them according to BI-RADS categories.

The architectures and techniques were chosen for their high precision and effectiveness in medical image analysis, surpassing traditional methods and significantly reducing processing time in training and implementation.

For model training, the Google Colab Pro infrastructure with GPUs is used. To avoid overfitting, the models are monitored using cross-validation and regularization techniques, employing metrics such as accuracy, sensitivity, specificity, and AUC-ROC.

The project is currently in the AI model development phase, with initial versions of the mentioned modules. The web platform under development aims to be intuitive and secure, focusing on professional interactivity and data protection through advanced AES-256 and OpenPGP encryption and two-factor authentication.

3 Preliminary Results

The project is currently in a development phase, focusing on creating AI models in virtual environments to be later integrated modularly into the web platform.

Figure 1 shows the main interface of the web platform, with access to the 3 modules containing the AI models being developed.

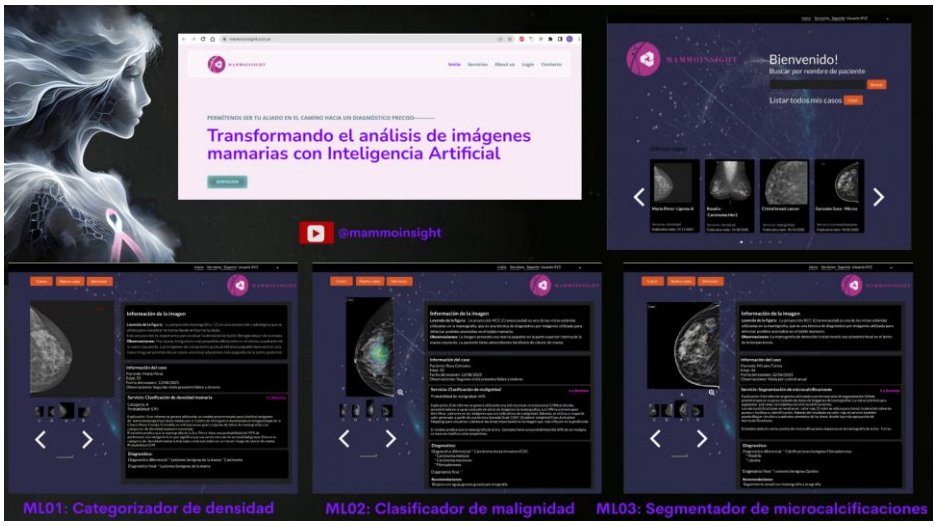

Fig. 1. Conceptual main interface of the MammoInsight platform with deployment of its categorization, classification, and segmentation modules.

The pre-trained model from Mayo Clinic was tested with the RSNA dataset, achieving results similar to the sensitivity presented in the official paper [8], with an ROC curve greater than 90%.

For malignancy classification, the model uses a large-volume dataset, which is processed in small partitions/mini-batches due to computational limitations. Although initial results are promising, the process is ongoing to avoid overfitting, conducting multiple full iterations that require time and resources.

Regarding anomaly segmentation, the U-Net based model is being optimized and integrated for microcalcification identification.

A layered architecture is being established with modules running on training, inference, web, and user database servers. Communication is carried out via FastAPI, with end-to-end encryption systems.

The potential impact is notable, with AI models showing high accuracy, sensitivity, and specificity. We seek early feedback for refinements and collaborations for clinical expansion and validation. With funding and premium feature monetization, the project aims for sustainability and future expansion.

The project's recognition as a finalist in the Ideatón Salud 2023 by CAEME highlights our approach and reinforces our commitment to innovations that improve medical diagnostics, inspiring us to overcome challenges and continuously improve.

4 Conclusions

Breast cancer, being the most common among women, presents challenges in early detection and diagnostic accuracy due to the limitations of manual mammogram interpretation. An innovative web platform using artificial intelligence for mammogram analysis offers advantages such as early anomaly detection, greater diagnostic accuracy, reduced interpretation times, and improved experience for patients and doctors. Despite challenges like the need for large datasets and clinical precision, this solution promises to advance towards more precise AI-driven breast imaging, significantly improving the diagnosis and treatment of breast cancer.

References

1. Instituto Nacional del Cáncer. Página principal de Argentina.gov. [Online]. Available: https://www.argentina.gob.ar/salud/instituto-nacional-del-cancer/estadisticas/incidencia (2023)
2. International Agency for Research on Cancer. Argentina Fact Sheet, Globocan 2020. [Online]. Available: https://gco.iarc.fr/today/data/factsheets/populations/32-argentina-fact-sheets.pdf (2020)
3. American College of Radiology. Breast Imaging Reporting and Data System (BI-RADS) Atlas, 7th ed. [Online]. Available: https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Bi-Rads (2022)
4. Lancet Oncology. "Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): A clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study." Lancet Oncology, vol. 24, no. 8, pp. 936-944, [Online]. Available: https://doi.org/10.1016/S1470-2045(23)00298-X (2023)
5. Radiology. "Comparison of Mammography AI Algorithms with a Clinical Risk Model for 5-year Breast Cancer Risk Prediction: An Observational Study." Radiology, vol. 307, no. 5. [Online]. Available: https://doi.org/10.1148/radiol.222733 (2023)
6. Pesce, K. "Inteligencia Artificial como Herramienta en el Diagnóstico Mamario." Revista Argentina de Mastología, vol. 39, no. 143, pp. 5-11. [Online]. Available: https://www.revistasamas.org.ar/revistas/2020_v39_n143/03.pdf (2020)
7. Kaggle. RSNA Screening Mammography Breast Cancer Detection. [Online]. Available: https://www.kaggle.com/competitions/rsna-breast-cancer-detection/data (2023)
8. A multi-reconstruction study of breast density estimation using Deep Learning. [Online]. Available: https://arxiv.org/pdf/2202.08238