Comparison between machine learning and deep learning for the classification of mammograms in BI-RADS

Moragues, Ignacio[1]; Benítez, Raul[1]; Mata, Christian[1]

[1]Automatic Control Department, Universitat Politècnica de Catalunya - EEBE, Barcelona, Spain.

***Abstract*—Epidemiological statistics portray the fact that breast cancer is a significant health concern and economic burden, undoubtedly justifying the need for breast cancer screening. Nevertheless, how the current diagnosis is made in clinical practice is prone to errors. Hence, there is a necessity for a tool to assist physicians when classifying mammographies into the four categories of BI-RADS. In this project, two approaches are presented: one based on machine learning and the other one based on deep learning. Mainly, beyond the comparison of the results, what is intended is to analyze and discuss in-depth the process followed to achieve their respective developments and subsequent implementation. Thus, the difficulties and drawbacks found when evaluating and comparing the two models are shown. Consequently, three mammography databases are used that experts have already classified following the BI-RADS guidelines.**

***Keywords*—BI-RADS, Texture, GLCM, LAWS, LBP, machine learning, deep learning, mammograms, breast cancer.**

I. Introduction

Female breast cancer has surpassed lung cancer as the most frequently diagnosed cancer. It is the most diagnosed cancer and the leading cause of cancer death in women [1]. Therefore, breast cancer is a substantial public health concern with a significant medical and economic burden [2]. Furthermore, the current diagnosis is error-prone. Thus, a justified need to develop a computer-aided diagnosis (CAD) system capable of assisting physicians in mammograms' classification exists. This project pretends to find the best pipeline to follow for developing this CAD tool in the Python language.

During the last years, different approaches have been proposed to deal with the classification of radiological breast images. These works exploit both machine learning and deep learning techniques. However, not many of them attempt to classify in the BI-RADS scale. In fact, the vast majority of works are focused on finding regions of interest, the likelihood of being cancerous, or simply whether mammograms are malignant or not.

The novelty and uniqueness of this project is the direct comparison between two different approaches of machine learning and deep learning for mammogram classification in BI-RADS.

II. Project framework

Different methodologies have been explored and addressed. However, only two approaches were fully developed: one based on machine learning and the other one on deep learning. The development strategy around other possibilities that were not fully implemented is also explained, with reference to the motives behind these decisions.

In the case of the machine learning model, texture is extracted using algorithms such as Grey-level co-occurrence matrix (GLCM), Local binary patterns (LBP) and Law's masks (LAWS). The next figure, represents the idea of texture extraction which is iterating on the image's pixels to obtain texture images:

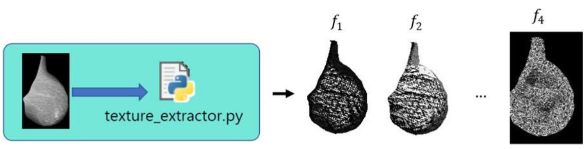

Figure II.1. Texture extractor concept.

Following the extraction, mammograms and texture images are stored in a data frame, structured as the following way:

$$pixel_{[i]} = \left[\, f_{1[i]}\ f_{2[i]}\ f_{3[i]}\ \dots\ f_{n[i]} \,\right] \qquad \textbf{Eq. II.1}$$

Where $pixel_{[i]}$ is the i^{th} pixel of a mammogram and $f_{n[i]}$ is the image's texture pixel that, by definition, characterize and describe each of the mammogram's pixel and its local neighbors. The dense area of the breast is then segmented, with the information obtained from texture, using Fuzzy C-means (an unsupervised soft clustering technique):

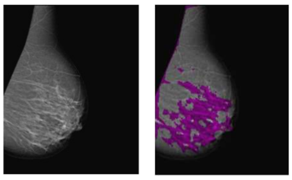

Figure II.2. Fuzzy C-means segmentation.

Subsequently, a feature selection process is carried out. The classification of the dense areas is performed using a k-nearest neighbors algorithm (k-NN), with a good trade-off between performance and computational time

In the case of deep learning, different architectures were tested, including Alexnet, Inception, Resnet50, but VGG-16 was the one that gave the best results. Grad-CAM [3] was used to understand and see the regions to which the neural network was giving more importance:

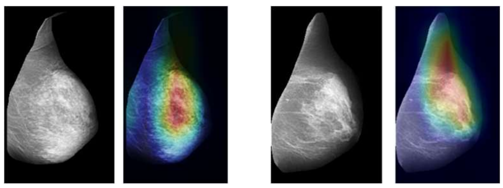

Figure II.3. Mammogram and Grad-CAM

Finally, the results obtained are presented and an exhaustive discussion is performed, demonstrating that the machine learning model requires great effort and experience to obtain acceptable results. In contrast, the deep learning model shows a much higher accuracy and can be considered as key for future work or research in this area.

References

[1] H. Sung *et al.*, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA. Cancer J. Clin.*, p. caac.21660, Feb. 2021, doi: 10.3322/caac.21660.

[2] O. Wyman, "El impacto económico y social del cáncer en España," 2020. [Online]. Available: https://www.aecc.es/sites/default/files/content-file/Informe-Los-costes-cancer.pdf.

[3] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 336–359, Oct. 2016, doi: 10.1007/s11263-019-01228-7.