POLITECNICO DI TORINO
Commission of Computer, Cinema and Mechatronic
Master's Degree in Data Science and Engineering

Summary of Master's Degree Thesis (October 2022)

"Deep learning for breast cancer diagnosis in contrast-enhanced breast CT"

Candidate: Francesco Di Salvo

Supervisor: Prof. Filippo Molinari, PhD
External Supervisor: Dr. Marco Caballo, PhD

Background: Breast cancer is the second most common cause of death from cancer in women in the United States after lung cancer. Thanks to early detection and treatment improvements, the mortality rate has been steady decreasing in the last decades. Therefore, there is an increasing interest in finding new methodologies for improving the current state of the art. The current gold standard for screening is digital mammography. However, because it only produces a 2D image of the breast, the lesion and the tissue will be on the same plane, making it more challenging to distinguish benign from malignant tumors or even to detect the lesions on particularly dense breast tissues. The dedicated breast computed tomography (bCT) is a recent technology for breast imaging, which provides a 3D image of the breast, gathering from 300 to 500 projections across 360°. Additionally, the visibility and differentiation of the lesions are improved by the intravenous injection of a contrast material.

Several works validated the efficiency of Artificial Intelligence (AI) algorithms for cancer detection and diagnosis, but the application of uncertainty-based models, which have potential to enhance result interpretability and therefore clinical translation, remains to be investigated in depth.

Project: This thesis aims to develop and validate Deep Learning algorithms for tumor classification and segmentation on 3D contrast enhanced breast computed tomographic (CE-BCT) scans, exploring the mass-level uncertainty of the predictions through the Monte Carlo Dropout.

Methods: 542 biopsy-proven breast masses (181 benign, 343 malignant) from 409 patients were imaged with a clinical Breast CT system after iodinated contrast medium administration. A 3D volume of interest (VOI) of 3.5cm per side was placed around each mass, and all masses were manually annotated in 3D by a board-certified breast radiologist.

The raw images were first resized for computational reasons, halving the initial dimension. Afterwards, due to the abnormal voxel values (by means of Hounsfield units) representing the radiodensity of the air (<1000) or a foreign body (>1000), they were clipped in the range [-200,250] and then normalized in the range [0,1].

The mass VOIs and respective binary annotations were used to train (n = 262) and fine-tune (n = 88) a two-channel 3D Dense Convolutional Network (Figure 1) and a 3D Residual UNet (Figure 2) for mass classification and segmentation, respectively. Both networks were tested on an independent dataset of 192 biopsy-proven breast masses (89 benign, 103 malignant). The classification algorithm was evaluated with the area under the receiver operating characteristics curve (AUC), with 95% confidence interval (C.I.) calculated with

bootstrapping (2,000 bootstraps) whereas the segmentation architecture was evaluated with the Dice score.

Due to the prohibitive cost of Bayesian Deep Learning for the uncertainty estimation, we employed the Monte Carlo Dropout, which was proved to approximate the Bayesian inference. It leverages a set of N inferences with dropout activated, which yields to N different model configurations that are likely to produce slightly different results, called Monte Carlo samples. Finally, multiple mass-level uncertainty metrics were tested on both classification and segmentation Monte Carlo outcomes, analyzing the performance improvement obtained by rejecting the predictions at different uncertainty and sensitivity thresholds.

Results: On the independent test set, the two-channels 3D Dense Convolutional Network achieved an AUC of 0.84 (95% CI 0.78-0.90). Then, the 3D Residual UNet achieved an average DICE score of 0.79 ± 0.2. Moreover, low-performance classification was found to be correlated with a high variance, increasing the accuracy by 8% when 57 test masses with highest prediction uncertainty were excluded. Moreover, the uncertainty was also found to be correlated with the segmentation performances, observing a linear correlation coefficient (ρ) of 0.76 and 0.58 for the Intersection over Unions (IoUs) and the average Dice score over Monte Carlo samples, respectively. This allowed to increase the Dice score by 12% in both cases by removing 57 test masses based on their relative uncertainty metric.

Finally, all our models were stressed through a five-fold cross validation approach, making inference on all our available images, in a more complex and more generalizable scenario. As expected, the semantic segmentation task and its relative uncertainty metrics confirmed their effectiveness, achieving comparable results. However, even though the imbalanced introduced by the cross validation affected the classification task (reducing the AUC by 8%), the variance-based exclusion criteria improved and balanced the overall performances.

Conclusions: The AI methods developed and validated under this study achieved promising performances and the evaluation of the uncertainty for the exclusion of the masses might enhance the performances. This could possibly be valuable for facilitating the translation of AI into clinics. Moreover, to the best of our knowledge, there is no prior work on fully 3D classification or segmentation on contrast-enhanced breast CT. However, our results are comparable or slightly below to the ones obtained on MRI, which is less cost-effective and provides a 4D image of the breast (including temporal information). Due to the radiation dose received through a breast CT, we are not ready to provide temporal information too. Anyway, the radiation dose steady decreased over the years, hoping to being able to provide temporal information in a couple of years, therefore there might be the chance to outperform the current state of the art obtained on MRI with a significantly more cost-effective technology, which may be more accessible to our hospitals.

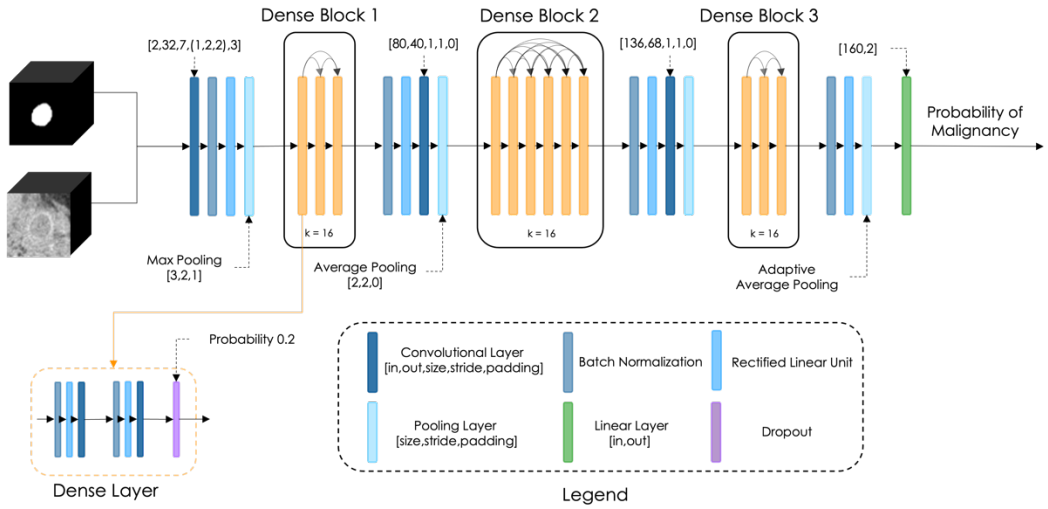

Figure 1. 3D DenseNet architecture implemented in this study. The network has three dense blocks, having 3,6 and 3 dense layers each, followed by dropout with probability of 0.2. The number of filters in the initial convolutional layer was set to 32, with a growth rate of 16, and all activation functions were rectified linear units (ReLUs). The network was inputted with the pre-processed mass VOIs (1.7 cm per side) and respective manual annotations, provided as a second channel, and it was trained with the Adam optimizer, minimizing the Cross Entropy Loss.

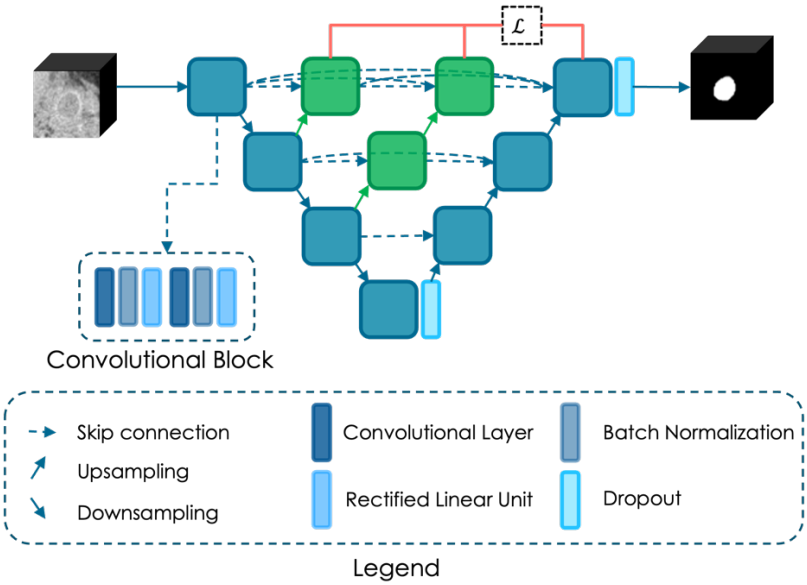

Figure 2. 3D UNet++ architecture implemented in this study. The network has three downsampling and upsampling blocks, and two dropout layers with probability of 0.2, placed after the bottleneck and after the final convolution. The residual blocks on the skip pathways (in green) aim to fill the semantic gap between the two submodules, improving the gradient flow. The network was inputted with the pre-processed mass VOIs (1.7 cm per side) and it was trained with the Adam optimizer, minimizing the Dice Loss.