



# Synthetic data for unleashed machine learning

Project Management and Software Development  
for Medical Applications

## General Info

Contact Person: Mickael Tardy, CSO, Hera-MI

Contact Email: [Mickael.tardy@hera-mi.com](mailto:Mickael.tardy@hera-mi.com)

## Project Abstract

In this project we will focus on the issue of lack of annotated data for training of deep learning algorithms. More precisely, we will work on the techniques of data synthesizing aiming for improving the performances of the algorithms without sourcing more data.

## Background and Motivation

In this project we will focus on the issue of lack of malignant data for training of deep learning algorithms. More precisely, we will work on the techniques of creation of synthetic data allowing to improve the performances of the algorithms without sourcing more data.

Deep learning algorithms require data for optimization of the thousands of parameters of neural networks. Sourcing of data is a challenge, especially in medical imaging, where abnormal and malignant cases are usually hard to get. That is, the prevalence of the diseases are usually low, leading to the imbalance of datasets and, by consequence, the difficulty to learn the malignancy. Various techniques of data augmentation exist, and the data synthesizing is amongst them.

In this work we will focus on the creation of the synthetic masses for mammograms [1], that has been successfully applied in the context of deep learning task [2].

## Student's Tasks Description

We will depart from existing public (<https://breastmass.readthedocs.io>) and private codebase to improve both the generation of the masses and their blending in mammograms. The following tasks are to be addressed:

- Thoroughly exploring the configuration opportunities of the proposed tool;
- Extending the limits of the proposed model;
- Improving the techniques of blending of the generated masses in real mammography images.

This project offers an opportunity mixing mathematical modeling and coding, by working on bio-medical simulation task while relying on and improving existing code.

## Technical Prerequisites

The following skills are needed:

- C++ (required), CMake (optional), Python (required), Docker (strongly desired)

## References

- [1] L. de Sisternes, J. G. Brankov, A. M. Zysk, R. A. Schmidt, R. M. Nishikawa, and M. N. Wernick, "A computational model to generate simulated three-dimensional breast masses," *Med. Phys.*, vol. 42, no. 2, pp. 1098–1118, 2015.
- [2] M. Tardy and D. Mateus, "Looking for Abnormalities in Mammograms with Self-And Weakly Supervised Reconstruction," *IEEE Trans. Med. Imaging*, vol. 40, no. 10, pp. 2711–2722, Jan. 2021.