

## Improving Breast Cancer Detection - Junior Data Scientist

**Internship description:** [Breast cancer](#) is the most frequently diagnosed solid cancer and second leading cause of cancer death among women. In most countries all the women beginning from a certain age should pass a mammography exam to detect the cancer as early as possible. Whereas this demands a large number of radiologists to analyze all the images, a lot of countries are missing experienced breast radiologists.



1 in 8 women affected during her lifetime



Every 1000 women screened, 5 will have breast cancer.  
But 100 will be recalled for further testing.

↓  
We can do better !

[Therapixel](#) is a spin-off Inria created in 2013 specialized in medical imaging with 12 employees. We are renowned for winning the [Digital Mammography DREAM Challenge](#) [1,2], a world-wide competition organized to improve the state-of-the-art in automatic mammography screening. Recently, we successfully closed a 5M€ series-A fundraise to accelerate the development of an algorithm for improving breast cancer screening.

We are now dealing with a tremendous amount of data. Preparing those to use them in practice is becoming a real challenge: raw data should be filtered to reject outliers (non-screening images), and anatomical landmarks (e.g., the nipple) have to be consistently extracted in each image. What was possible to do manually with a few thousands of data becomes impractical with hundreds of thousands or millions of images!

At the interface between the Data Science Team and the Data Management Team, this internship aims at developing algorithms to help us prepare and annotate data. The candidate will develop algorithm to:

- Automatically reject non-screening images (e.g., magnified images, images taken during biopsy, images with surgery artefacts). A sequence of deep networks with different purposes for each will surely be necessary. The candidate will have to find the best architecture for each objective along with the optimal working image resolution, and perform hyper-parameter search.
- Automatically detect landmarks by implementing a detection network such as retina-net, and pilot data annotation in collaboration with the Data Management team.

A server with 8 recent GPUs (NVidia 1080Ti / Titan X / P100) will be entirely dedicated to this project. The candidate is highly encouraged to take initiatives, to propose and test new ideas.

**Candidate description:**

- Motivated by medical challenges
- Good understanding of modern Deep Learning architectures (VGG, ResNet, RetinaNet)
- Some experience with at least one Deep Learning framework (ideally Tensorflow)
- Good coding skills (Python / C++ / C)
- Familiar with standard Python libs (Numpy, Scipy, matplotlib)
- High general scientific culture and research spirit

**Modalities :**

Internship length: 6 months  
Internship location: Paris, RER station Port Royal (pépinière Paris Santé-Cochin)  
Salary: 2000 € net / month  
Contact e-mail: [stage-dl@therapixel.com](mailto:stage-dl@therapixel.com)

Note that this internship could lead to a research scientist position at Therapixel and / or a PhD in partnership with a French Research Institute.

Please include your resume, motivation letter, and grades obtained so far.

*Join Therapixel – help in building the medicine of the future*

**References:**

[1] DM Challenge Yaroslav Nikulin (Therapixel) write up. Available:  
<https://www.synapse.org/#!Synapse:syn9773040/wiki/426908>.

[2] Therapixel Press Release <https://goo.gl/BSG6rZ>