Welcome to the age of individualized medicine and machine (deep) learning for medical imaging applications. With the development of medical imaging innovations, quantitative imaging can be obtained, which has the ability to capture characteristics of tumors in a non-invasive way. Radiomics refers to extraction and analysis of large amounts of advanced quantitative imaging features with high throughput from radiological images, such as computed tomography (CT), positron emission tomography (PET) or magnetic resonance imaging (MRI). Radiomics data, which can include biological or medical data, are in a mineable form that can be used to build descriptive and predictive models relating image features to provide valuable diagnostic, prognostic or predictive information. The goal of this thesis is to build prognostic or predictive models from Radiomics data using machine leaning in order to help oncologists to improve individualized therapeutic treatment and monitoring.

The first step of the thesis is to design descriptors to extract relevant imaging features which can describe the characteristics of tumors. Information about patient state will be also used to construct Radiomics data. The second step is to develop predictive models based on the Radiomics data using deep learning and kernel methods. In medical field, we are usually face to the problems of small data size comparing to the quantitative imaging features and imbalanced data. For the first problem, we propose to develop methods of deep learning using HPC, dimensionality reduction and selection of task-specific features. Machine learning from imbalanced data sets is an important problem, both practically and for research. The thesis will focus on machine learning methods especially to deal with imbalanced data. The proposed methods for analysis within radiomics must be reliable and reproducible so that they could potentially be employed within a clinical context.

The thesis work is in collaboration with "Regional Centre Henri Becquerel for the Fight Against Cancer " that provides radiological images and medical expertise.

This thesis will be carried out in the LITIS laboratory (http://www.litislab.eu/), and supervised by the Professor Stéphane Canu of INSA de Rouen and the Professor Su Ruan of the Rouen University. Note that Stéthane Canu is the PI of deep in Normandy the NVIDIA award research center (https://developer.nvidia.com/academia/centers/normastic).
Machine learning modeling challenges in medical images includes:

- Deep learning: sparsity and generalization using GPU
- Noise, data quality and unbalanced classes
- Use data to innovate new representations
- The evaluation issues: how to measure the quality and the effect of a model?

This thesis is at the crossroad of individualized medicine, medical image processing and machine learning. Individualized medicine define the problems, medical image processing specify it and machine learning provide tools to solve it.

The main phases of the work are the following:

1. literature survey
   - machine learning for medical imaging
   - deep learning and medical image processing
   - unbalanced classes
2. experimental setup
   - data acquisition
   - validation protocol
3. sparsity and generalization for
   - deep learning
   - kernel approaches to feature extraction
   - machine learning and prediction

The work will be done in the LITIS lab building at the INSA engineering school in Rouen.