23 – 25 July 2018

Lviv Data Science Summer School 2018

Machine Learning for Medical Applications:

Introduction

Igor Koval

PhD Student in Applied Mathematics

Brain and Spine Institute, Pitié Salpétrière Hospital, Paris, France
& Mathematical Laboratory of Ecole Polytechnique

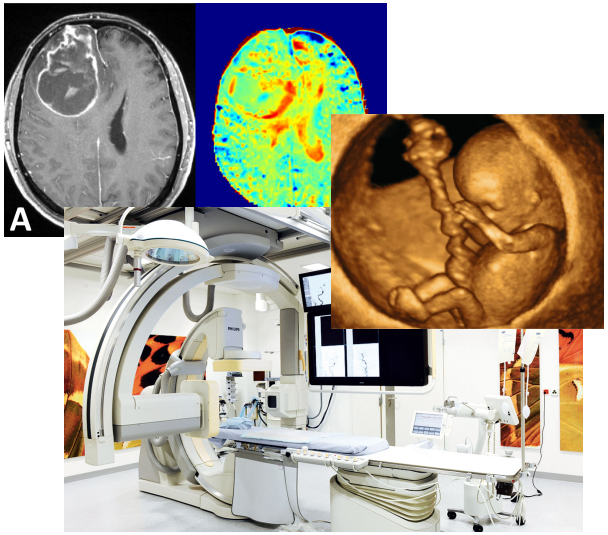
igor.koval@icm-institute.org

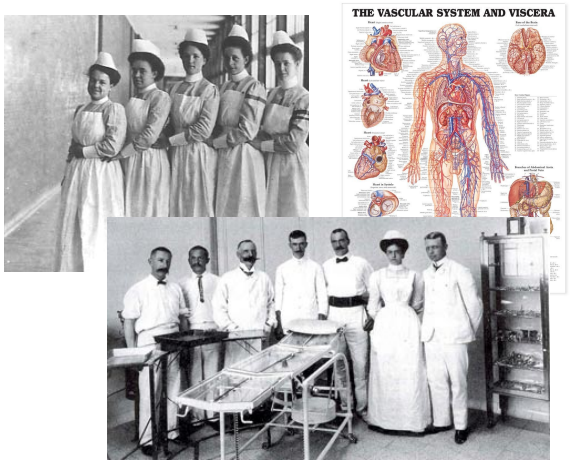
INTRODUCTION

History of medecine

ARAMIS LAB
BRAIN DATA SCIENCE

UKRAINIAN CATHOLIC UNIVERSITY
FACULTY OF APPLIED SCIENCES



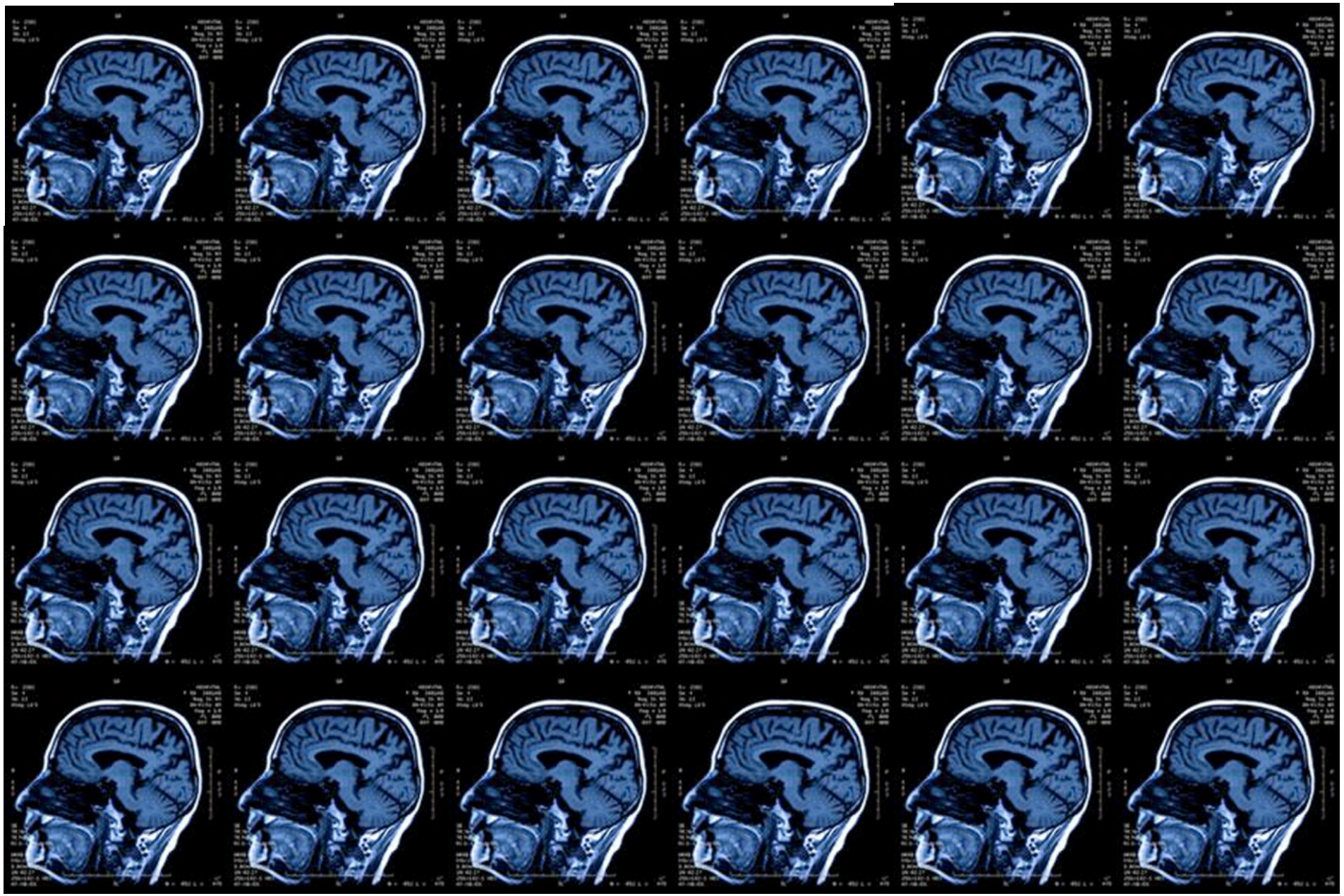


Increase of complexity, “time to response”, variety, scale, ...

Machine Learning for Medical Applications

ARAMIS LAB
BRAIN DATA SCIENCE

UKRAINIAN CATHOLIC UNIVERSITY
FACULTY OF APPLIED SCIENCES

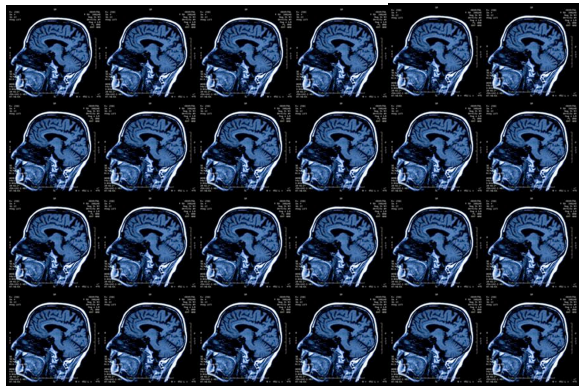


Machine Learning for Medical Applications

ARAMIS LAB
BRAIN DATA SCIENCE

UKRAINIAN CATHOLIC UNIVERSITY
FACULTY OF APPLIED SCIENCES

The accuracy of the tide phenomenon (at a local scale) does not come from its physical model ..



... but from its repetition

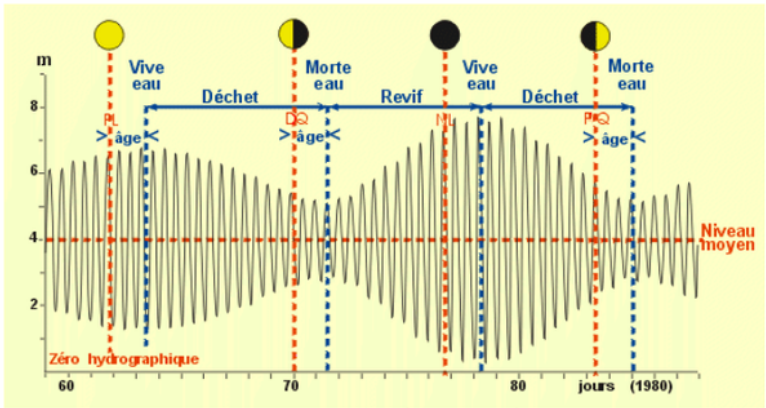


Fig. 1. Amplitude et âge de la marée

Machine Learning for Medical Applications

ARAMIS LAB BRAIN DATA SCIENCE

UKRAINIAN CATHOLIC UNIVERSITY | FACULTY OF APPLIED SCIENCES

Medecine

- Needs few examples to understand a correlation or causality
- Mechanistics problems
- Treat the disease afterwards

- Increasing number of data
- Complex pathologies (Cancer, AIDS, cancer, ...)
- Prevent / Predict the disease

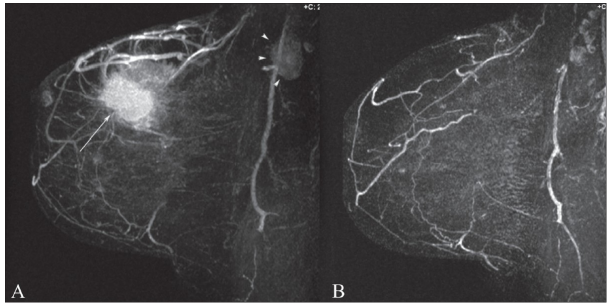
Machine Learning

Examples (1/3)

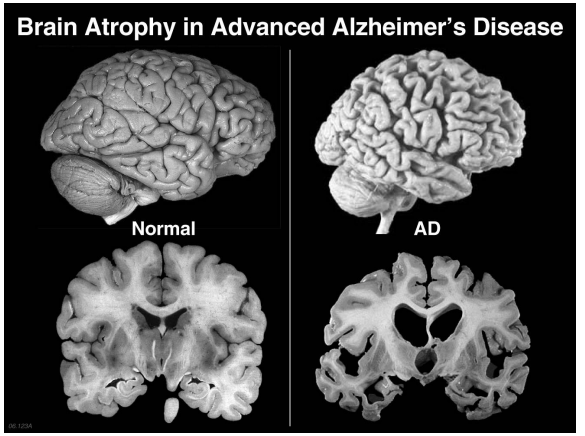
ARAMIS LAB BRAIN DATA SCIENCE

UKRAINIAN CATHOLIC UNIVERSITY | FACULTY OF APPLIED SCIENCES

- Detection of breast tumors and diabetes based on images (Deep Learning ; Kaggle)



- Diagnose the particular step of a disease (cancer tumor, Alzheimer's Disease, ...)



- Test drugs on homogeneous cluster of people to assess the consistency

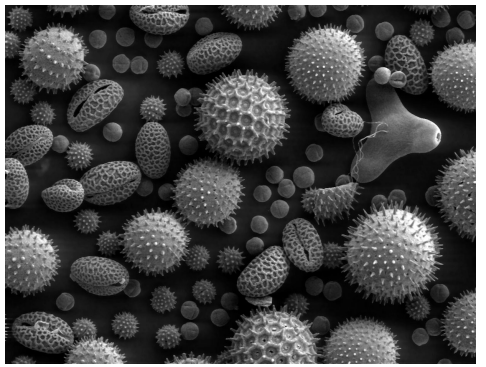
Examples (2/3)

ARAMIS LAB
BRAIN DATA SCIENCE

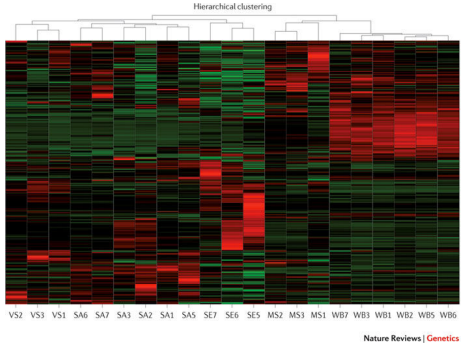
UKRAINIAN CATHOLIC UNIVERSITY
FACULTY OF APPLIED SCIENCES

- Reduce the cost of a diagnosis : cognitive scores instead of PET or MRI

- Better microscopy with Deep Learning



- Exhibition of genes involved in Alzheimer's Disease

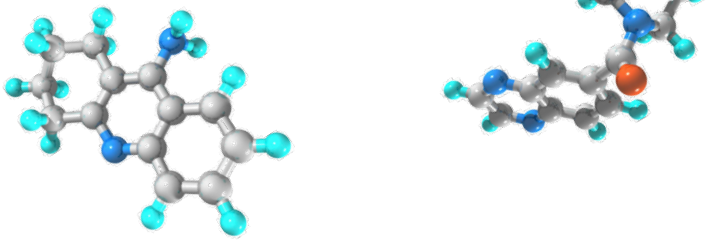


Examples (3/3)

ARAMIS LAB BRAIN DATA SCIENCE

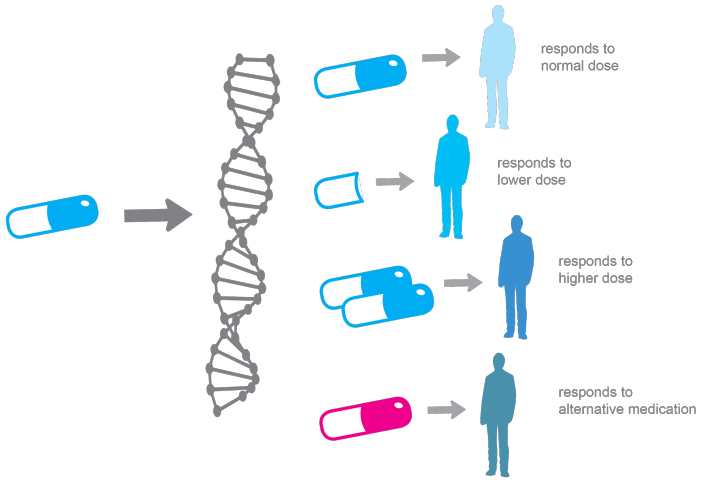
UKRAINIAN CATHOLIC UNIVERSITY | FACULTY OF APPLIED SCIENCES

- Kernels specially designed to built / create consistent molecules (considered as graphs)



- And many more to come…

- Personalized medecine : diabete people use a scale everyday to predict an upcoming risk



PLAN

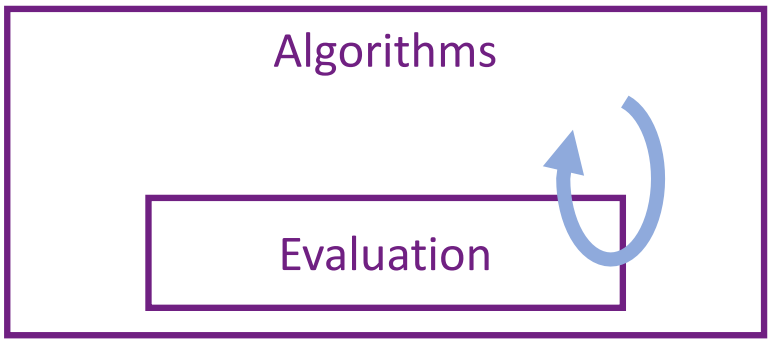
The pipeline

- Prediction
- Exhibit groups
- Diagnosis
- Pattern detection
- …

- Biomarkers
- Images
- Genomics
- Text
- Graph, network, mesh
- Time-series
- …

- Regression
- Clustering
- Classification
- Distribution of the population
- …

- Missing values
- Outliers
- Unbalanced classes
- Categorical data : ordered or not
- …



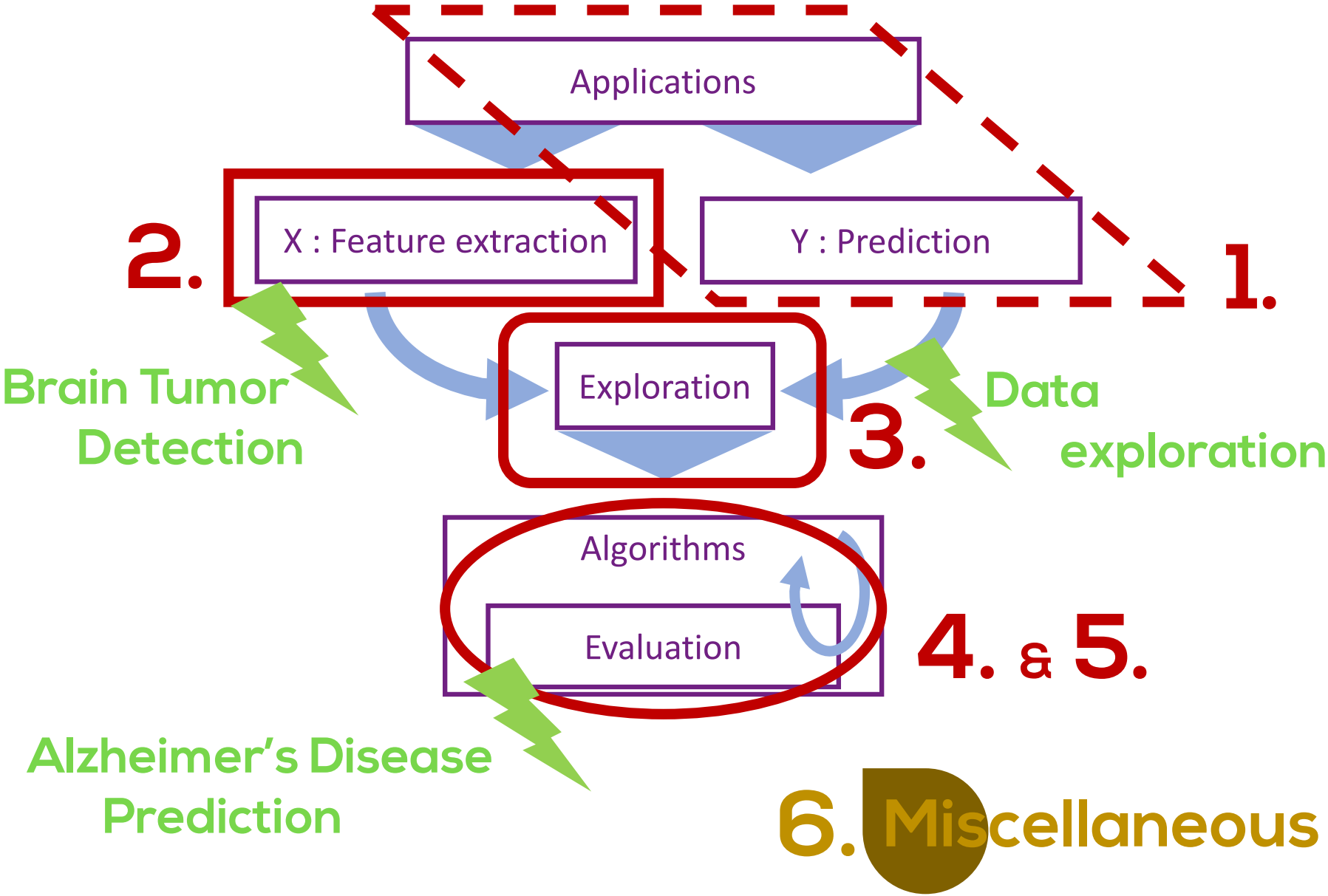
- Linear regression
- k-NN
- Log-Reg
- SVM
- Trees
- Deep Learning
- Other methods : Assembling, boosting, bagging, …

- Proper cross-validation
- Metrics / Loss to optimize
- Hyper-parameters tuning
- …

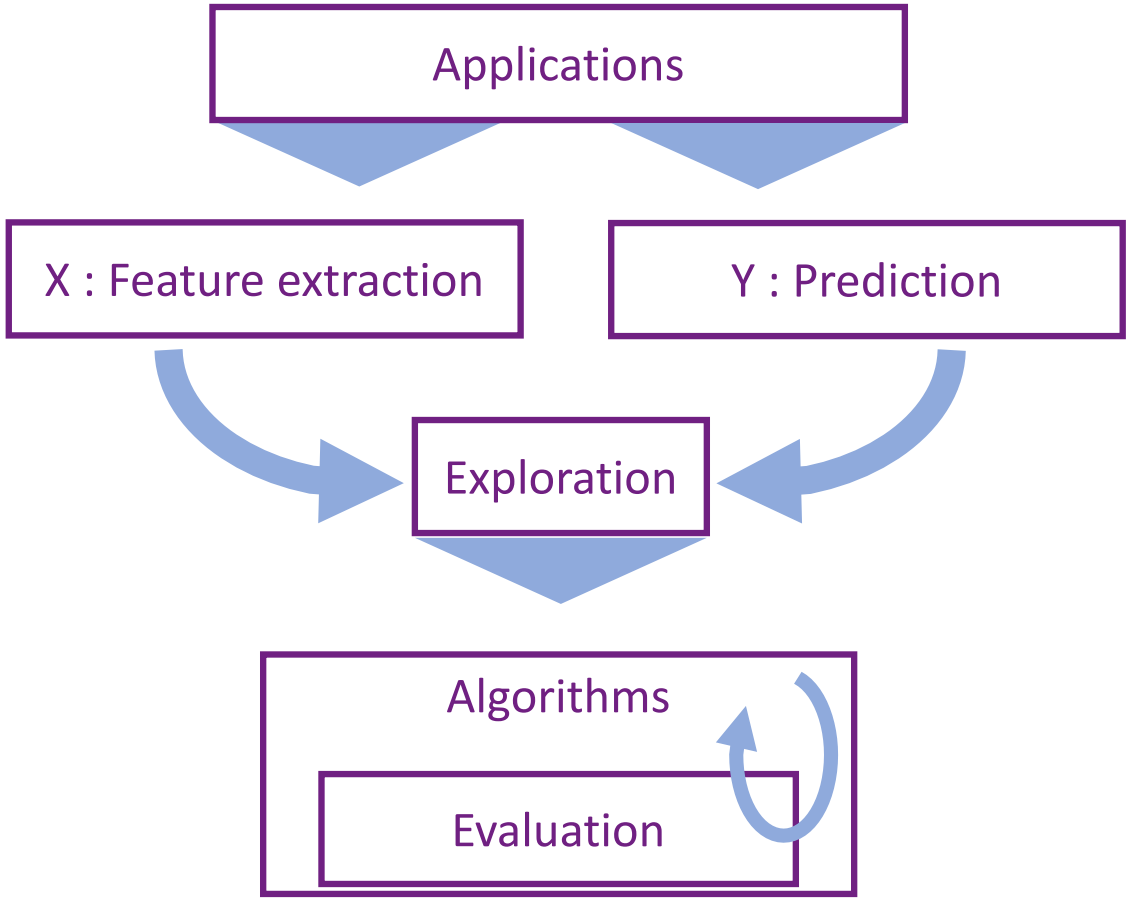
Schedule & Objectives

ARAMIS LAB BRAIN DATA SCIENCE

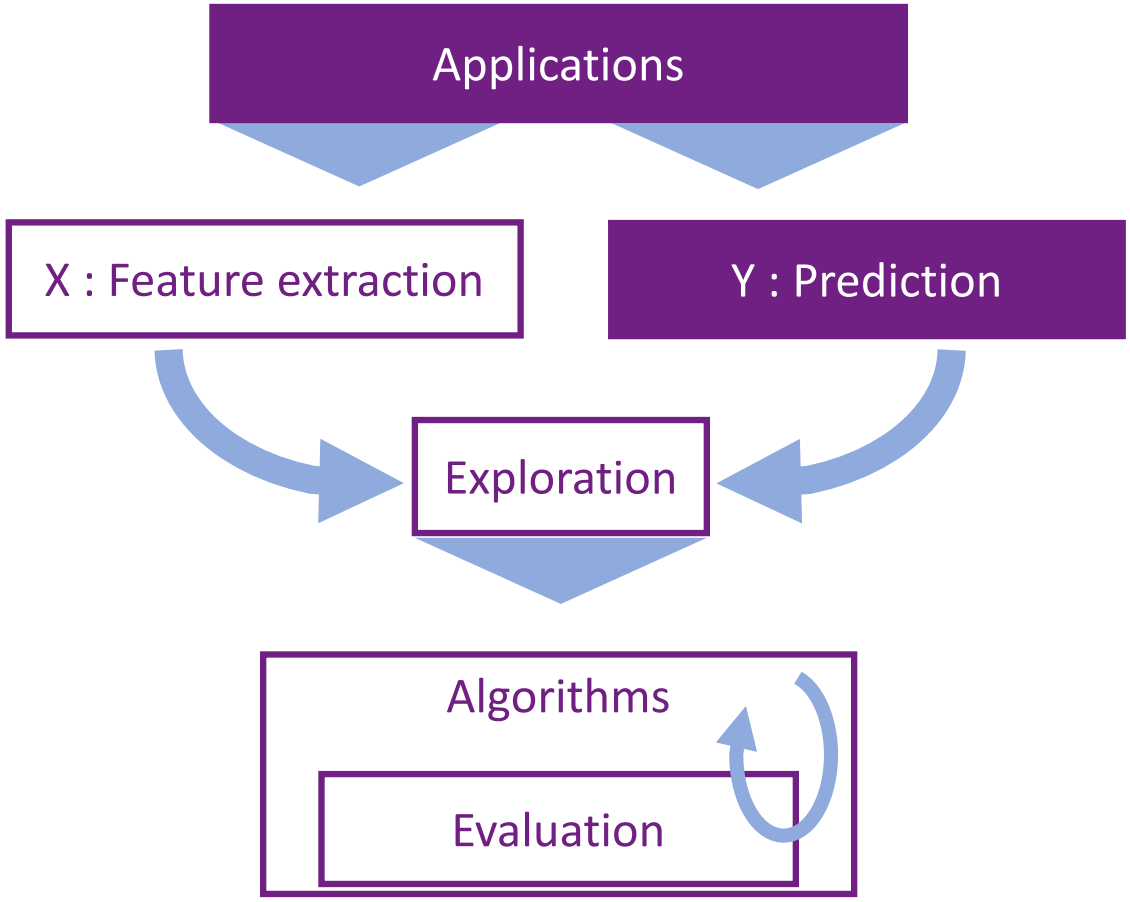
UKRAINIAN CATHOLIC UNIVERSITY | FACULTY OF APPLIED SCIENCES



Plan


Applications
X : Feature extraction
Y : Prediction
Exploration
Algorithms
Evaluation

Plan


Applications
X : Feature extraction
Y : Prediction
Exploration
Algorithms
Evaluation

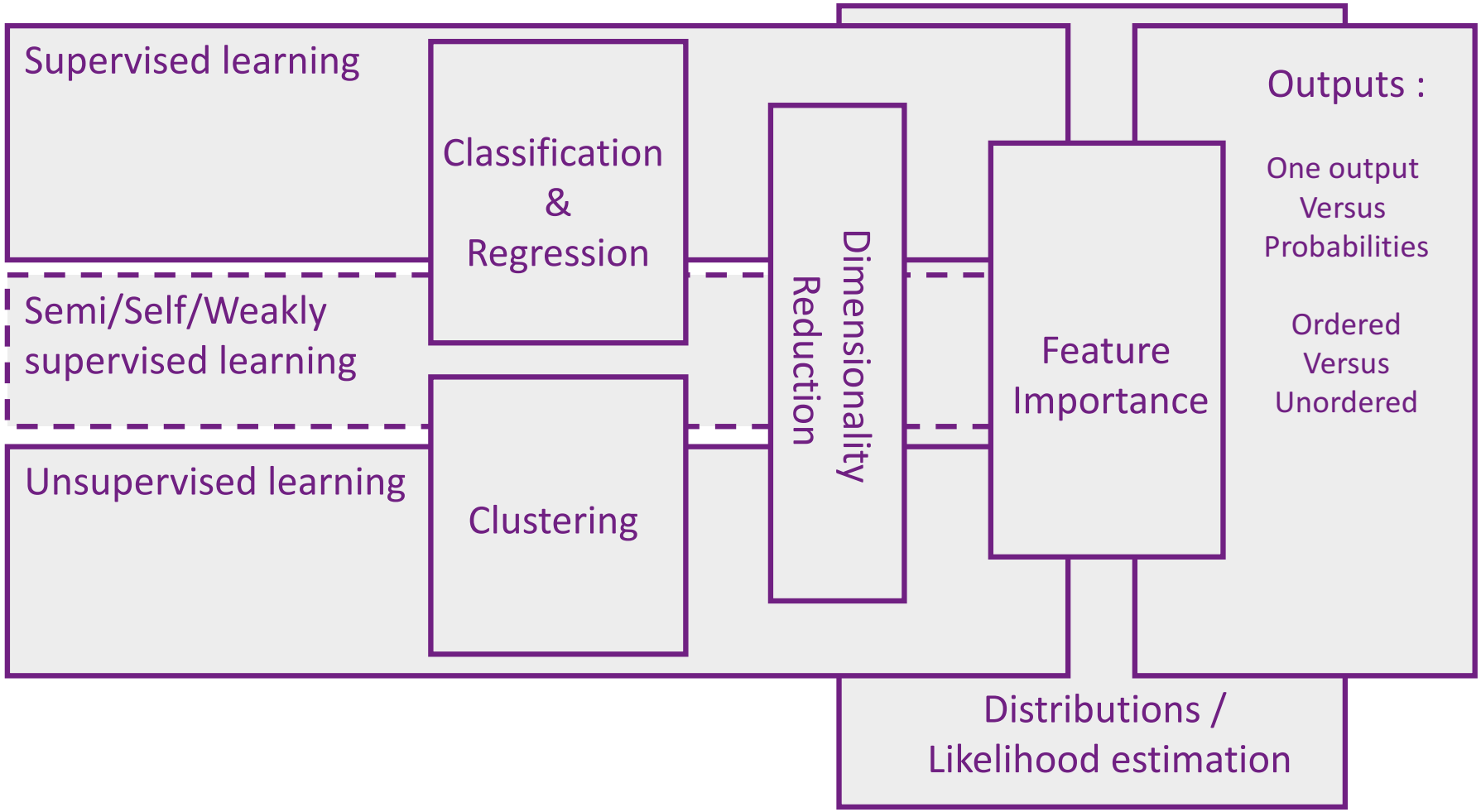
Application & Corresponding prediction

ARAMIS LAB BRAIN DATA SCIENCE

UKRAINIAN CATHOLIC UNIVERSITY | FACULTY OF APPLIED SCIENCES

Application ▶ Mathematical prediction

Convert the medical problem, into a mathematical prediction Y = f(X) ?



Application & Corresponding prediction

ARAMIS LAB BRAIN DATA SCIENCE

UKRAINIAN CATHOLIC UNIVERSITY | FACULTY OF APPLIED SCIENCES

Application ▶ Mathematical prediction

1. Given a population whose individuals have a particular disease, we want to if there are subgroups.

2. An individual was diagnosed with cognitive impairments. The doctors want to know which degenerative disease the subject has.

3. We have individuals with the same disease. Given all their features, we want to know which are the most informative.

4. Personalized medecine; we want to adapt the treatment to a new individual.

Supervised / Unsupervised / Semi-supervised

Clustering

Regression

Classification (Ordered / Unordered)

One output / Probabilities

Feature importance (Algorithm-based / Dimensionality-reduction based)

Application & Corresponding prediction

ARAMIS LAB BRAIN DATA SCIENCE

UKRAINIAN CATHOLIC UNIVERSITY | FACULTY OF APPLIED SCIENCES

Application ▶ Mathematical prediction

1. An individual was diagnosed with a cancer. We want to assess the stage of the cancer.

2. A pharmaceutical group has a potential new preventive drug. It wants to maximize the chances of passing through the tests, i.e. having the best clinical impact.

3. Someone caries a gene that will express a disease at some point. We want to know when (s)he going to convert from normal state to abnormal.

4. Given a set of genes, we want to know which one(s) are related to a particular disease.

Supervised / Unsupervised / Semi-supervised

Clustering & Mixture of Models

Regression

Classification (Ordered / Unordered)

One output / Probabilities

Feature importance (Algorithm-based / Dimensionality-reduction based)

Other issues & Challenges

ARAMIS LAB BRAIN DATA SCIENCE

UKRAINIAN CATHOLIC UNIVERSITY | FACULTY OF APPLIED SCIENCES

- How to compare patients? For instance, big brain versus small brain
- How to know if someone is really representative of a healthy patient?
- How to know if the truth is true?
- Dimensionality curse -> Medecine is always cursed
- Doctors know the Linear Regression well … The role of explanation is key
- Pharmaceutical : how to know if the change is related to the treatment or not?

  (Multi-armed bandit algorithms – game theory related)
- Overlapping and/or continuous classes, labeled by the doctors