

# A Machine Learning Approach for Cancer Detection

## A-Introduction:

### • Cancer : a challenging disease

Cancer involves a **abnormal proliferation of cells** that tend to damage surrounding tissues and spread to other parts of the organism.

France :

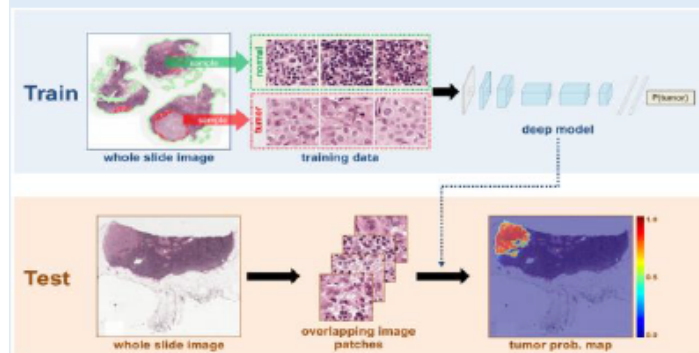
2018 : **First cause of Death**, 382,000 New Diagnosed cases, 157,400 Deaths.

1980-2005: +35% for Men +43% for Women

### • Machine Learning : a new perspective

**Machine Learning** is a field of study in mathematics and statistics that aims at making **machines** able to **learn** automatically from **data**.

By detecting **new patterns** in data sets, it provides new criterias for cancer detection.



Example of a Machine Learning Process Applied to Breast Cancer Detection

## B-Objectives :

Provided a data set containing information about DNA methylation, Gene Expression and Exon Expression we aim at detecting liver, colon and cervical cancer using machine learning tools.

## C-Methods:

### • Logistic Regression :

The goal is to **separate** our set of data into three distinct areas corresponding to each type of cancer by **minimising a cost function**.

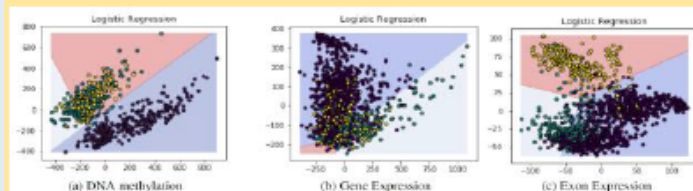


Figure 1: Logistic Regression using PCA reduced data-set

### • Reducing Dimension :

The goal is to reduce the number of variables we consider to obtain a more **compact model** and so **regroup data**.

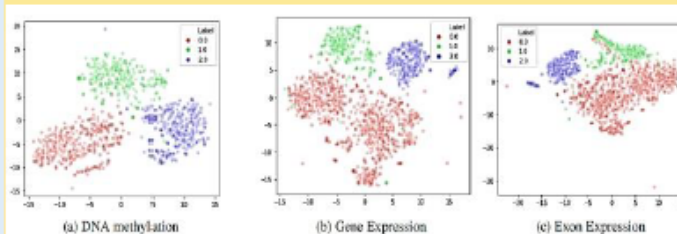


Figure 2: t-SNE dimensionality reduction results

### • Decision Tree :

The goal is to **construct a tree** in order to **classify** a potential patient by successively **subdividing** the data. These subdivisions give us criterias to determine at each step what kind of cancer the patient is most likely to have.

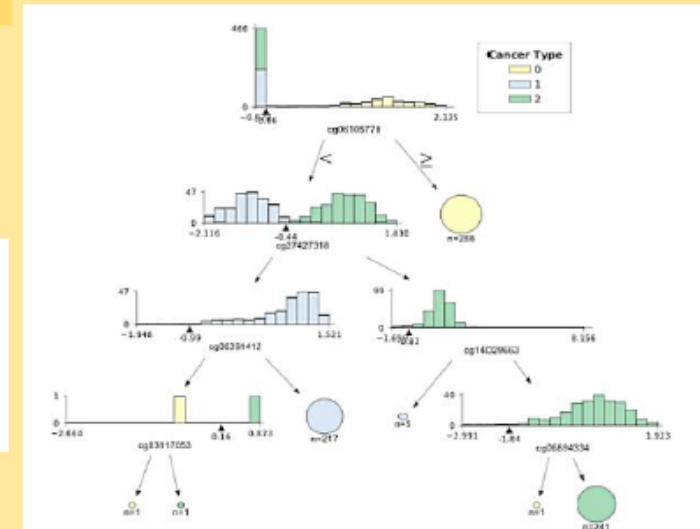


Figure 3: DNA methylation Decision Tree main features

## D-Results :

Algorithm	DNA methylation (%)	Gene Expression (%)	Exon Expression (%)
Logistic Regression	100	99	98
Random Forest	100	98	99
Support Vector Machine	100	99	100
Decision Trees	98	97	98
Linear Discriminant Analysis	100	96	42
Naive Bayes Classifier	99	83	60

Table 1: Classification accuracy between the 3 different features of the 3 considered cancers

## E-Conclusion:

**Machine Learning** gives us extremely accurate methods to identify different types of cancer. Therefore everything leads us to believe that this technology, in the field of **medicine**, has a very **promising future**.