



## Formation

“Data Science et image satellitaires”

*03 février 2023 et 08 mars 2023*

# Sommaire

1. Introduction à l'imagerie satellitaire
2. Comment récolter des données issues de l'imagerie spatiale ?
3. L'IA et les applications avec des données spatiales
  - a. USE CASE 1 : utilisation d'images vectorielles
  - b. USE CASE 2 : utilisation d'images rasters



1

# Introduction à l'imagerie satellitaire

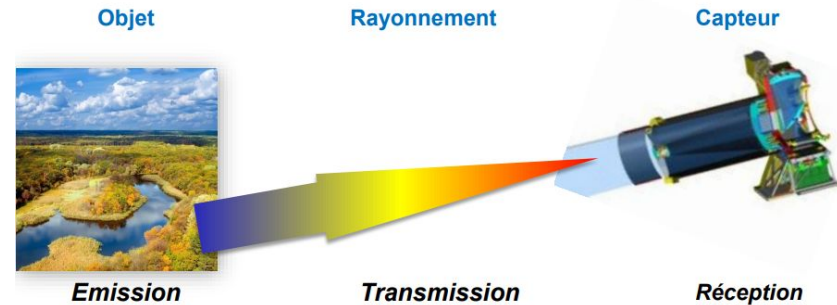
## A. La télédétection optique et radar

# Introduction à l'imagerie spatiale

Définition [Imagerie satellitaire] = désigne la prise d'images depuis l'espace, par des capteurs placés sur des satellites.



## *Qu'est-ce que la Télédétection ?*

C'est l'ensemble des **techniques** utilisées pour **déterminer à distance** les **propriétés** d'objets, naturels ou artificiels, à partir des **rayonnements** qu'ils émettent ou réfléchissent.



# Introduction à la télédétection optique et radar

On peut exploiter deux types de données :

- des ondes émises **par le soleil** puis réfléchies par la surface de la Terre,  **télédétection passive et images optiques**
- des ondes émises **par un émetteur artificiel** placé sur le satellite puis réfléchies par la surface de la Terre,  **télédétection active et images radars**

Différences points forts / faibles :

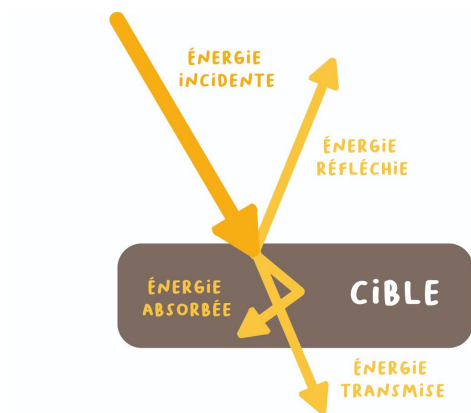
	Avantages	Inconvénients
<b>Image optique</b>	<ul style="list-style-type: none"> <li>• traitement intuitif (comme des photos en couleur)</li> </ul>	<ul style="list-style-type: none"> <li>• pas de données de nuit</li> <li>• perte d'information selon la couverture nuageuse</li> </ul>
<b>Image radar</b>	<ul style="list-style-type: none"> <li>• pas de contraintes de couverture : les ondes émises par les satellites traversent les nuages</li> <li>• pouvoir acquérir des images de jour comme de nuit</li> </ul>	<ul style="list-style-type: none"> <li>• exploitation moins intuitive</li> <li>• expertise nécessaire pour l'interprétation</li> </ul>

# Comment ça marche la télédétection ?

**Définition [Mesures d'un capteur]** = Lorsqu'une surface (la cible) reçoit un rayonnement (artificiel ou naturel), la cible va en partie :

- absorber l'énergie reçue,
- transmettre aux surfaces sous-jacentes,
- réfléchir vers le ciel → c'est cette part d'énergie qui est enregistrée par les capteurs des satellites

Les capteurs enregistrent le pourcentage de réflectance :

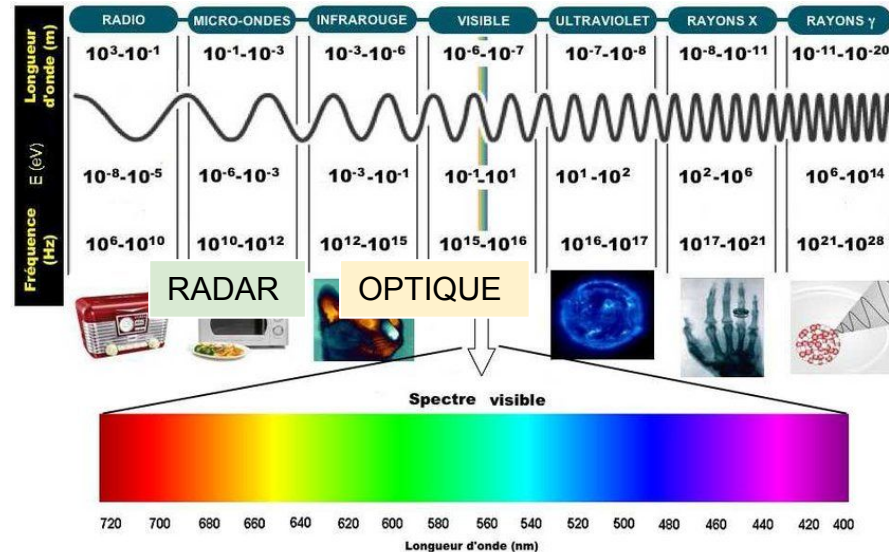


$$\text{RÉFLECTANCE (\%)} = \frac{\text{ÉNERGIE RÉFLÉCHIE}}{\text{ÉNERGIE INCIDENTE}}$$

# La mesure enregistrée par le satellite

**Rappels :** toute onde se caractérise par une longueur d'onde (ou une fréquence) et une amplitude.

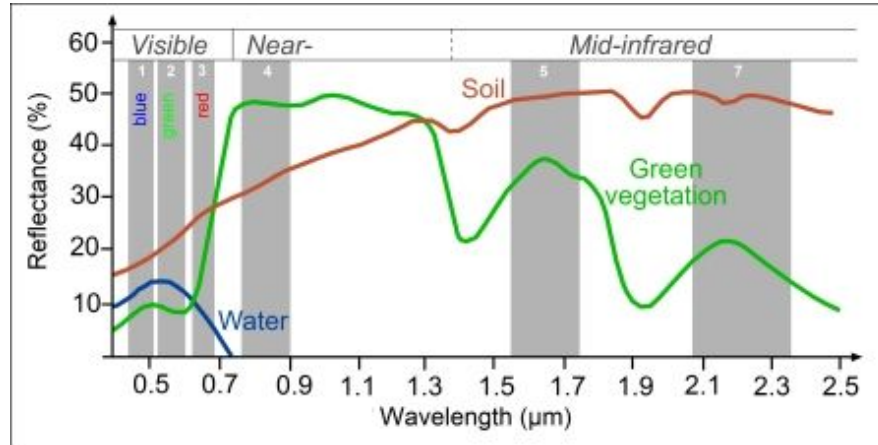
Le rayonnement qui est réfléchi par la cible, se propage sous forme d'ondes, représentées sous forme de spectre électromagnétique. Chaque partie du spectre est susceptible de fournir des informations sur l'objet.





# Exemple d'interprétation d'une mesure

Graphique avec l'information de la réflectance selon la surface :

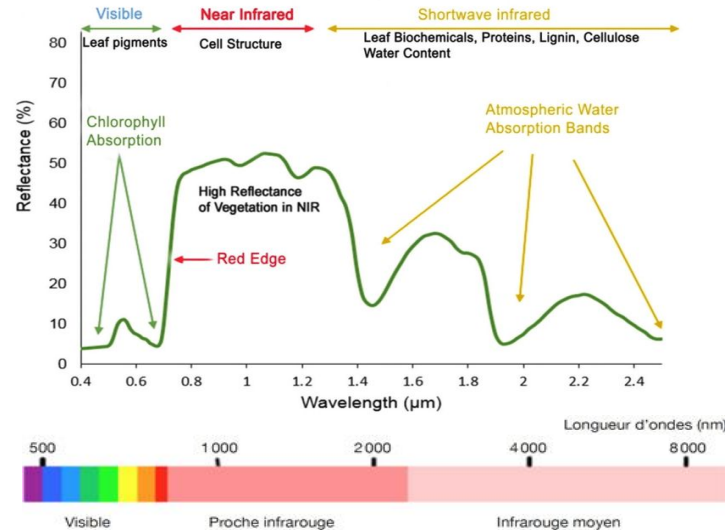


- Les trois courbes sont **relativement proches** dans le **spectre visible**, donc la réflectance de l'eau, de la végétation et du sol nu est similaire dans le spectre visible.
- Pour distinguer les trois types d'occupation du sol, on doit regarder **la réflectance sur d'autres longueurs d'ondes**, plus grandes, par exemple dans le **"proche infrarouge"**.

# Conclusion de l'exemple

## Exemple :

- la **végétation** absorbe les **radiations bleu et rouge**, et ne laisse passer que **les radiations jaunes et vertes**.
- Si les pigments chlorophylliens absorbent la majeure partie du rayonnement visible qui leur parvient, ils sont en revanche totalement transparents au rayonnement de plus grande longueur d'onde
- Il y a un très haut pic sur le début de la bande spectrale du **proche infrarouge**.





2

## Collecte des données satellitaires

## B. Les données satellitaires

# Choix des données

Présentation de différents satellites et propriétés d'acquisition des données :


	Sentinel 2	MODIS	Landsat	SPOT 617	WorldView	Pleiades
<b>Résolution</b>	10 m	250 m	30 m	1.5 m	30 cm	50 cm 30 cm
<b>Fréquence de visite</b>	5 jours	5/7 jours	2 semaines	N/A	N/A	N/A (à programmer)
<b>Coût</b>	Free	Free	Free	\$	\$	\$
<b>Couverture</b>	Partout	Partout	Partout	Acquisition	Acquisition	Acquisition
<b>Bande spectrale</b>	13 bandes	N/A	13 bandes	4 bandes (RGB NIR)	6 bandes	4 bandes 6 bandes



Surveillance de la  
Terre

# Téléchargement des données

Présentation de différents fournisseurs de données :

	Scihub	DIAS	PEPS	THEIA	AWS	Google	USGS
<b>Coût</b>	Free	Free	Free	Free	Request pay	Free	
<b>Profondeur historique</b>	Archive	Archive	Archive	Non	Non	Non	
<b>Couverture</b>	Partout	Partout	Partout	France / UE / 159 sites dans le monde	Partout	Partout	
<b>Recherche</b>	api	api	api	api	Sentinel hub plateforme	api	
<b>Commentaires</b>	Catalogue existe	Catalogue d'images existe			Configurer le compte		Image Landsat

Prend du temps pour la 1ère configuration

# Téléchargement des données depuis Google Cloud Platform

## Les catalogues disponibles :

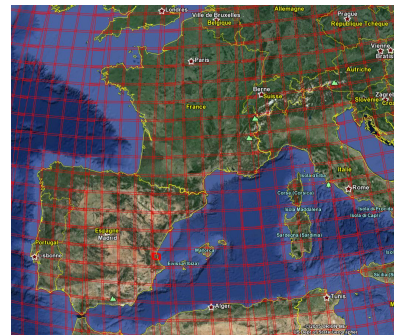
- Landsat (américain) : <https://cloud.google.com/storage/docs/public-datasets/landsat?hl=fr>
- Sentinel 2 : <https://cloud.google.com/storage/docs/public-datasets/sentinel-2?hl=fr>

les données utilisées par Copernicus notamment

## Utilisation : il existe 2 méthodes pour extraire une zone d'intérêt :

- via les tuiles : chaque tuile a un "id" → il faut connaître l'id des tuiles de la zone d'intérêt
- via le contour de la zone d'intérêt qu'on définit dans l'api

## Résultat du téléchargement :



# Présentation des données

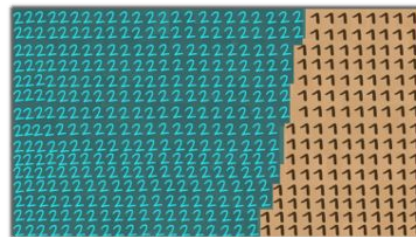
Les données géographiques peuvent être représentées, numériquement, sous 2 formes différentes :

- le mode vecteur
- le mode raster

Qu'est-ce qu'un raster ? Une image composée de pixels.



Exemple de données continues en mode raster



Exemple de données discrètes en mode raster

Qu'est-ce qu'un vecteur ? Un objet géométrique à deux composantes (graphique et attributive).



Les trois géométries de base



## C. Présentation de l'objectif de la formation

# | Présentation de l'objectif à atteindre

## Domaine d'application sélectionnée :

→ La gestion des ressources agricoles et la répartition des parcelles agricoles

## Objectifs techniques :

→ Étudier un **indice de végétation** afin de suivre l'évolution de la végétation de parcelles agricoles sur une zone spécifique sélectionnée

→ Démonstration de la prédiction de **la classe de parcelles agricoles** à l'aide d'un indice de végétalisation obtenu à partir d'images satellitaires.

# Présentation d'un indice de végétation

**Objectif pratique :** Suivre l'évolution de la végétation de parcelles sur une zone spécifique à l'aide d'un indicateur scientifique


## Définitions de l'indice de végétation NDVI (Normalized Difference Vegetative Index) :

- ❖ c'est un indice de végétation par différence normalisé
- ❖ il permet de déterminer la santé de la végétation en mesurant la teneur en chlorophylle des plantes
- ❖ il permet d'évaluer la vigueur et la quantité de végétation d'une zone géographique


## Comment calculer l'indice NDVI ?

Indice de végétation :

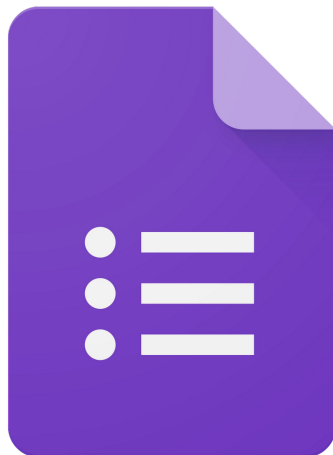
$$NDVI = (NIR - RED) / (NIR + RED)$$



réflectance des canaux proche  
infrarouge (NIR)



réflectance des canaux rouge  
(RED)



[Lien QCM Télédetection](#)



3

## Applications IA

## D. L'IA et les images satellites

# Présentation d'un indice de végétation

**Objectif pratique :** Suivre l'évolution de la végétation de parcelles sur une zone spécifique à l'aide d'un indicateur scientifique


## Définitions de l'indice de végétation NDVI (Normalized Difference Vegetative Index) :

- ❖ c'est un indice de végétation par différence normalisé
- ❖ il permet de déterminer la santé de la végétation en mesurant la teneur en chlorophylle des plantes
- ❖ il permet d'évaluer la vigueur et la quantité de végétation d'une zone géographique


## Comment calculer l'indice NDVI ?

Indice de végétation :

$$NDVI = (NIR - RED) / (NIR + RED)$$

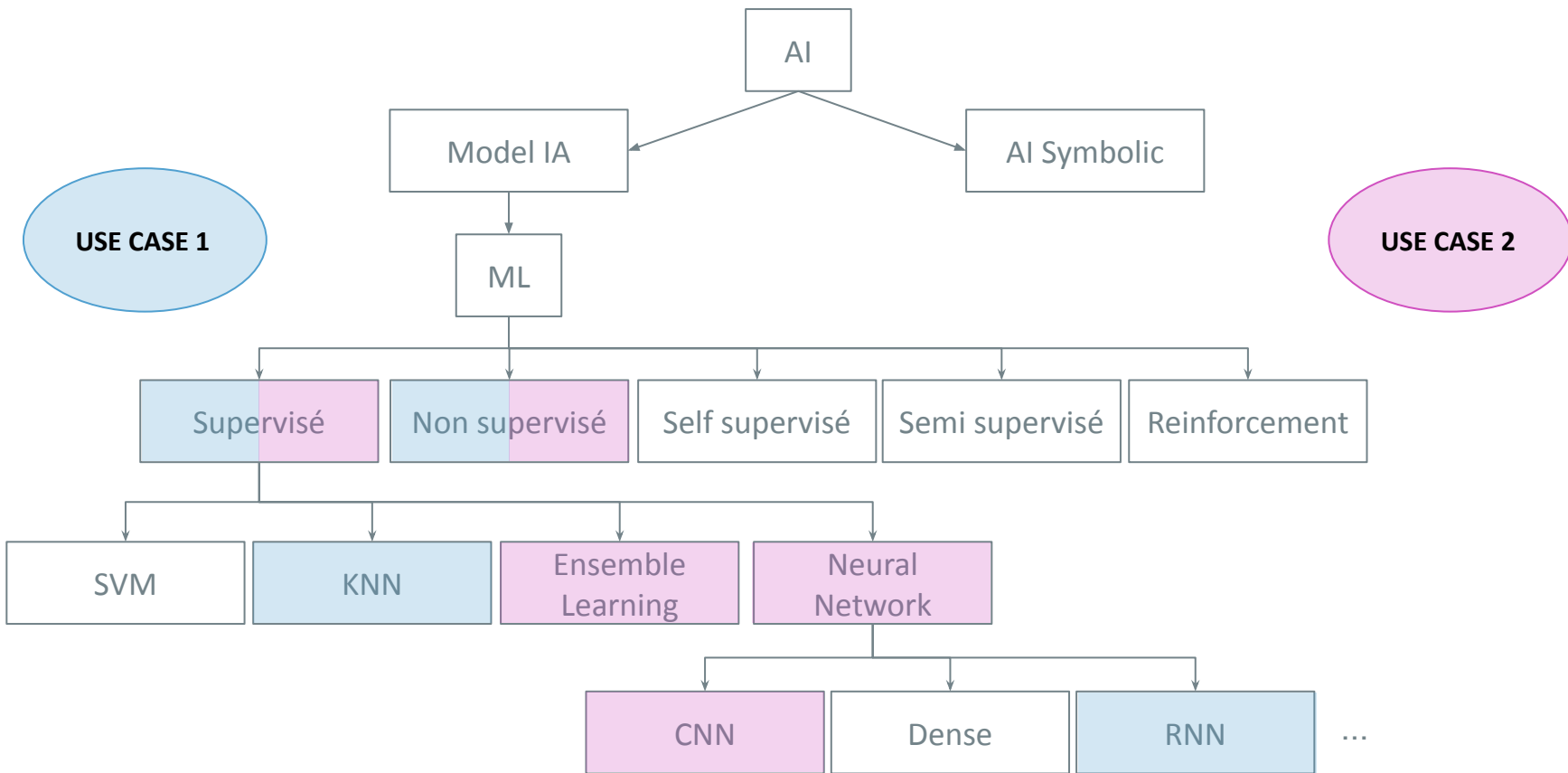


réflectance des canaux proche  
infrarouge (NIR)



réflectance des canaux rouge  
(RED)

# Artificial Intelligence, Machine Learning et Deep Learning





# USE CASE n° 1 : dataset sous forme **vectorel**

**Objectif pratique : Suivre l'évolution de la végétation de parcelles sur une zone spécifique**

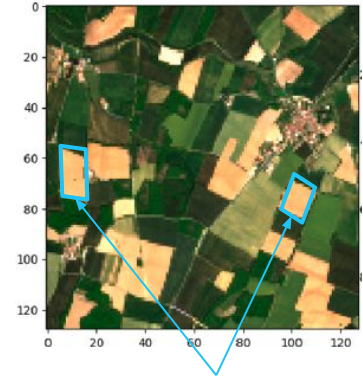
**La préparation des données :**

1. Préparer dataset : depuis une image satellite (.tif) et un contour de parcelle en format .shapefile
2. Extraction de features via connaissance métiers ➡ Calculer l'indice de végétation

$$\text{Indice de végétation :}$$

$$\text{NDVI} = (\text{NIR} - \text{RED}) / (\text{NIR} + \text{RED})$$

3. Sauvegarder les calculs des NDVI
4. Calculer des statistiques par polygone de parcelle agricole
5. Créer une série temporelle des statistiques par polygone

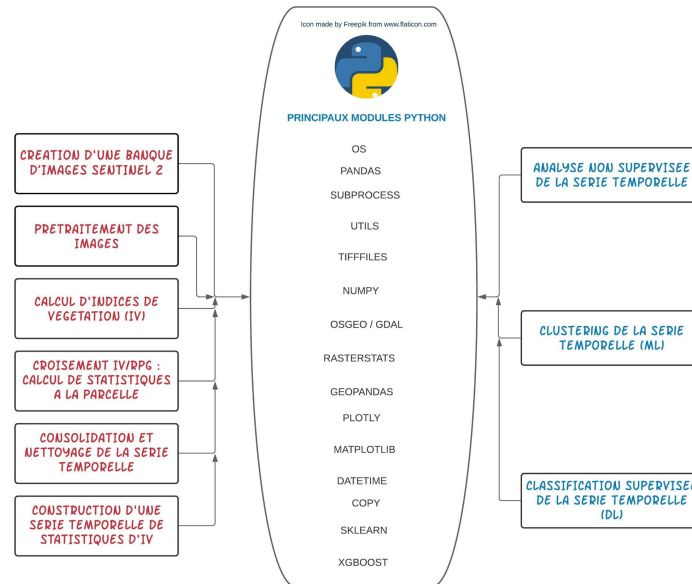


exemple : contours de parcelles spécifiques à étudier

# 1 - Collecte des données

**Objectif :** Se familiariser avec la manipulation de données structurées

**Objectif pratique :** Suivre l'évolution de la végétation de parcelles sur une zone spécifique



# | 1 - Collecte des données (TP)

USE CASE 1

**Application :**

**[Lien notebook google colab "1 - Prepare data"](#)**

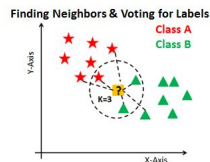
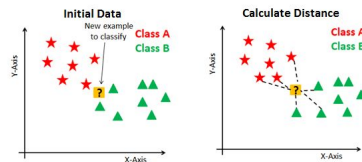
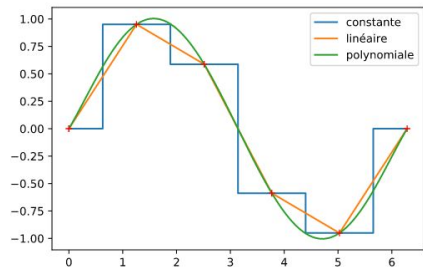
## 2 - Data preprocessing : traiter les valeurs nulles



### CONSOLIDATION ET NETTOYAGE DE LA SERIE TEMPORELLE

#### Quelques méthodes :

- Imputation :
  - par une constante,
  - par la moyenne,
  - par la médiane,
  - par la donnée la plus représentée, ...
- Interpolation : à partir d'un nombre fini de points reconstruire une fonction
  - linéaire,
  - polynomiale,
  - spline (une fonction polynomiale par morceaux),
  - cubic (cas particulier de polynôme d'ordre 3), ...
- K-nearest neighbors reconstruction



## | 2 - Data preprocessing : traiter les valeurs nulles (TP)

USE CASE 1

**Application :**

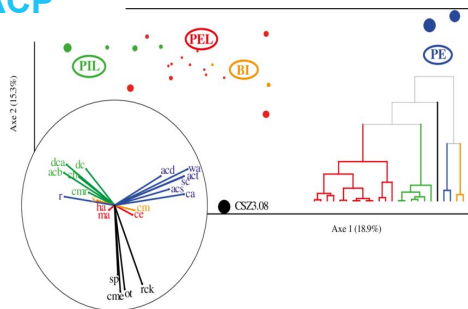
**[Lien notebook google colab "2 - Inspect temporal stats"](#)**

# 3 - Analyse non supervisé de la série temporelle

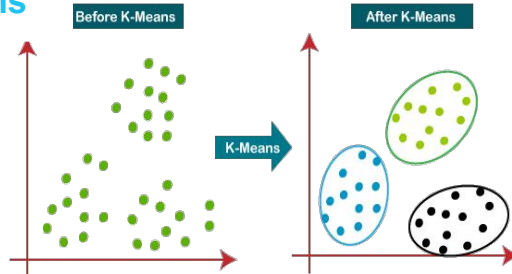
## Les algorithmes étudiés :

- ACP (Analyse des Composantes Principales) → réduction de dimensions, extraction de features
- KMeans (K-moyennes) → clustering
- Variational Auto Encodeur → réduction de dimensions (réseaux de neurones récurrents)

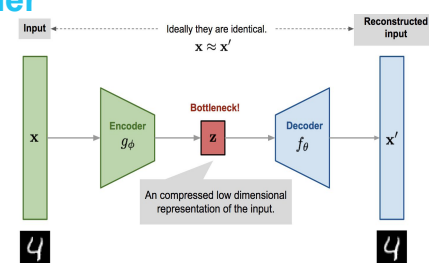
### ACP



### KMeans



### AutoEncoder



## 3 - Analyse non supervisé de la série temporelle (TP)

**Application :**

**[Lien notebook google colab "3 - Unsupervised learning"](#)**

## | 4 – Analyse supervisée de la série temporelle (ML)

### Qu'est-ce que l'apprentissage automatique ?

- "A computer program is said to learn from **experience  $E$**  with respect to some class of **tasks  $T$**  and **performance measure  $P$**  if its performance at tasks in  $T$ , as measured by  $P$ , **improves with experience  $E$ .**" - Tom M. Mitchell
- "Machine learning algorithms build a model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, where it is difficult or **unfeasible to develop conventional algorithms to perform the needed tasks**. The study of **mathematical optimization** delivers methods, theory and application domains to the field of machine learning." - Wikipedia



## 4 – Machine learning (introduction)



USE CASE 1

### Comment les machines apprennent-elles ?

- Quand on dit qu'une machine apprend, on sous-entend qu'elle essaie de résoudre un **problème d'optimisation**. Oui, c'est le cas :)
- Résoudre un problème d'optimisation = trouver le meilleur modèle possible.
- La machine va essayer de minimiser une **fonction de coût\*** en utilisant un outil : l'algorithme de **descente de gradient**.
- La descente de gradient est un algorithme d'optimisation qui permet de trouver **le minimum de toute fonction convexe** en y convergeant progressivement.

\* **Mean Squared Error par exemple**

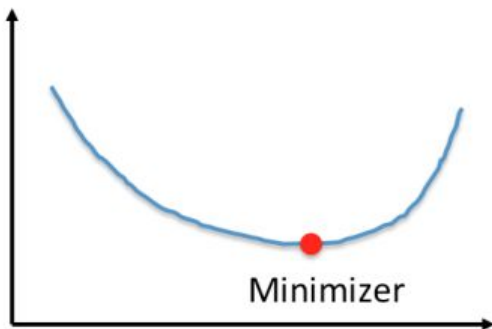
## 4 - Machine learning (fonction convexe)



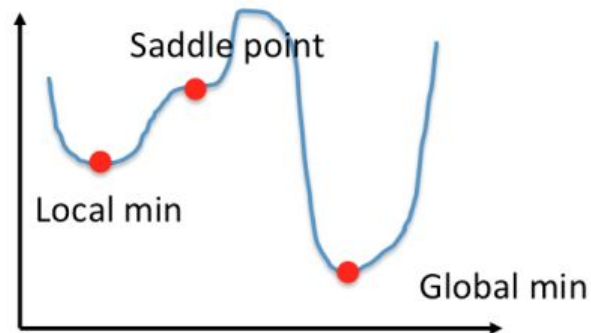
USE CASE 1

**Descente de gradient** : comment les algorithmes apprennent ?

**Convex**



**Non-Convex**



## 4 - Machine learning (gradient descent)



USE CASE 1

La meilleure analogie pour l'algorithme de descente du gradient est celle de la randonnée en montagne.



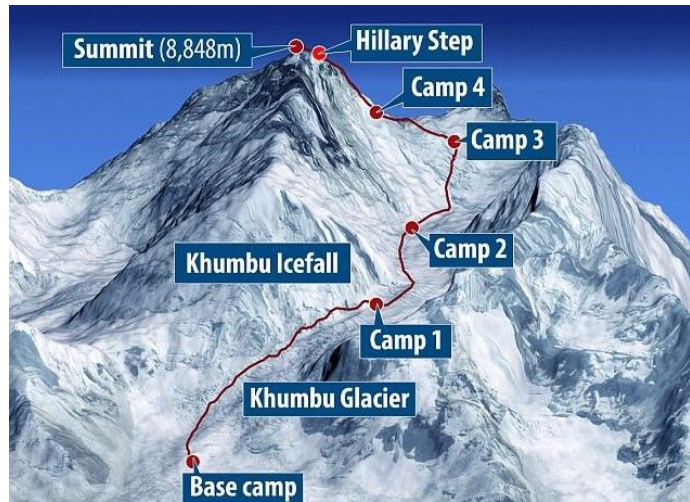
## 4 - Machine learning (fonction convexe)



### USE CASE 1

La meilleure analogie pour l'algorithme de descente du gradient est celle de la randonnée en montagne.

- Vous êtes au sommet de l'Everest et vous devez descendre.
- Vous commencez par chercher un chemin qui vous mène à un point plus bas. Vous le suivez.
- Vous atteignez ensuite ce point plus bas, mais vous n'êtes toujours pas en bas.
- Tu fais exactement la même chose, tu cherches un chemin vers un point plus bas et tu le suis.
- Tu continues à faire ça jusqu'à ce que tu atteignes le fond.
- Cela prend un certain temps, mais vous finirez par atteindre le fond.



## 4 – Machine learning (gradient descent)



USE CASE 1

### Un exemple de descente de gradient (théorie vs pratique) :

#### *Exemple théorique :*

Hypothesis:  $h_{\theta}(x) = \theta_0 + \theta_1 x$

Parameters:  $\theta_0, \theta_1$

Cost Function:  $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$

Goal:  $\underset{\theta_0, \theta_1}{\text{minimize}} J(\theta_0, \theta_1)$

#### *Exemple pratique :*

- Hypothesis: for example here a linear regression with an equation of type  $y = ax + b$
- Parameters: **a** and **b**
- Cost function: **Mean Squared Error**
- Goal: **minimize the cost function**

*Minimiser la fonction de coût signifie trouver les paramètres  $a, b, c$ , etc. qui donnent les plus petites erreurs entre le modèle et les points  $y$  de l'ensemble de données réelles.*

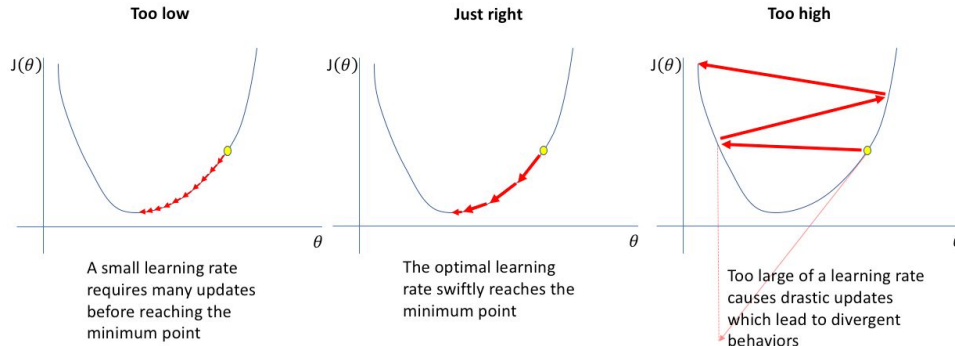
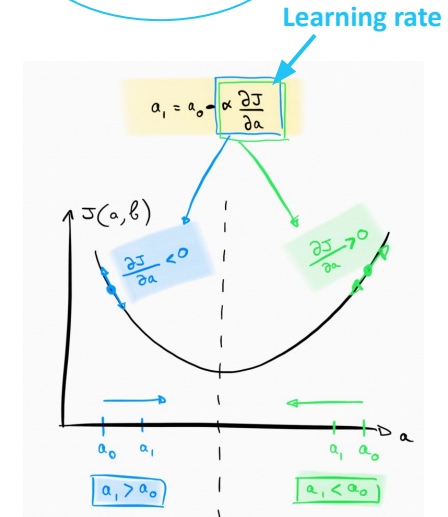
## 4 - Machine learning (gradient descent)



### USE CASE 1

- Nous choisissons un point de départ aléatoire  $a_0$  sur la courbe de la fonction de coût.
- Nous calculons le gradient (=dérivées partielles pour chaque paramètre (a, b))
- Nous calculons la perte
- Nous gardons le nouveau point  $a_1$  seulement si la perte a diminué.
- Nous continuons jusqu'à atteindre les minima globaux

### Ajuster le paramètre du taux d'apprentissage :



## 4 – Analyse supervisée de la série temporelle (ML)

### Les algorithmes étudiés :

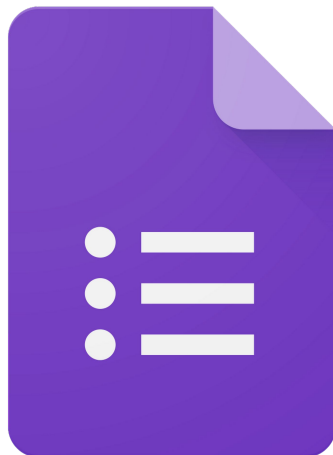
- **XGBoost (extrême gradient boosting) :**
  - Bagging (méthode ensembliste parallèle) → combiner les résultats d'un ensemble de modèles afin de fournir une meilleure prédiction
  - Boosting (méthode ensembliste séquentielle) → apprentissage en prenant en compte les erreurs de l'itération antérieure à l'aide de pondération
- **LSTM (Long Short Term Memory) :**
  - Input data : données triées temporellement
  - Domaine d'application :
    - Problématique : savoir quelles décisions ont été prises dans le passé afin de prendre une décision optimale à l'instant  $t$ .
    - En pratique : classification de séries temporelles, avec gestion des valeurs nulles

## | 4 - Analyse supervisé de la série temporelle (TP)

**Application :**

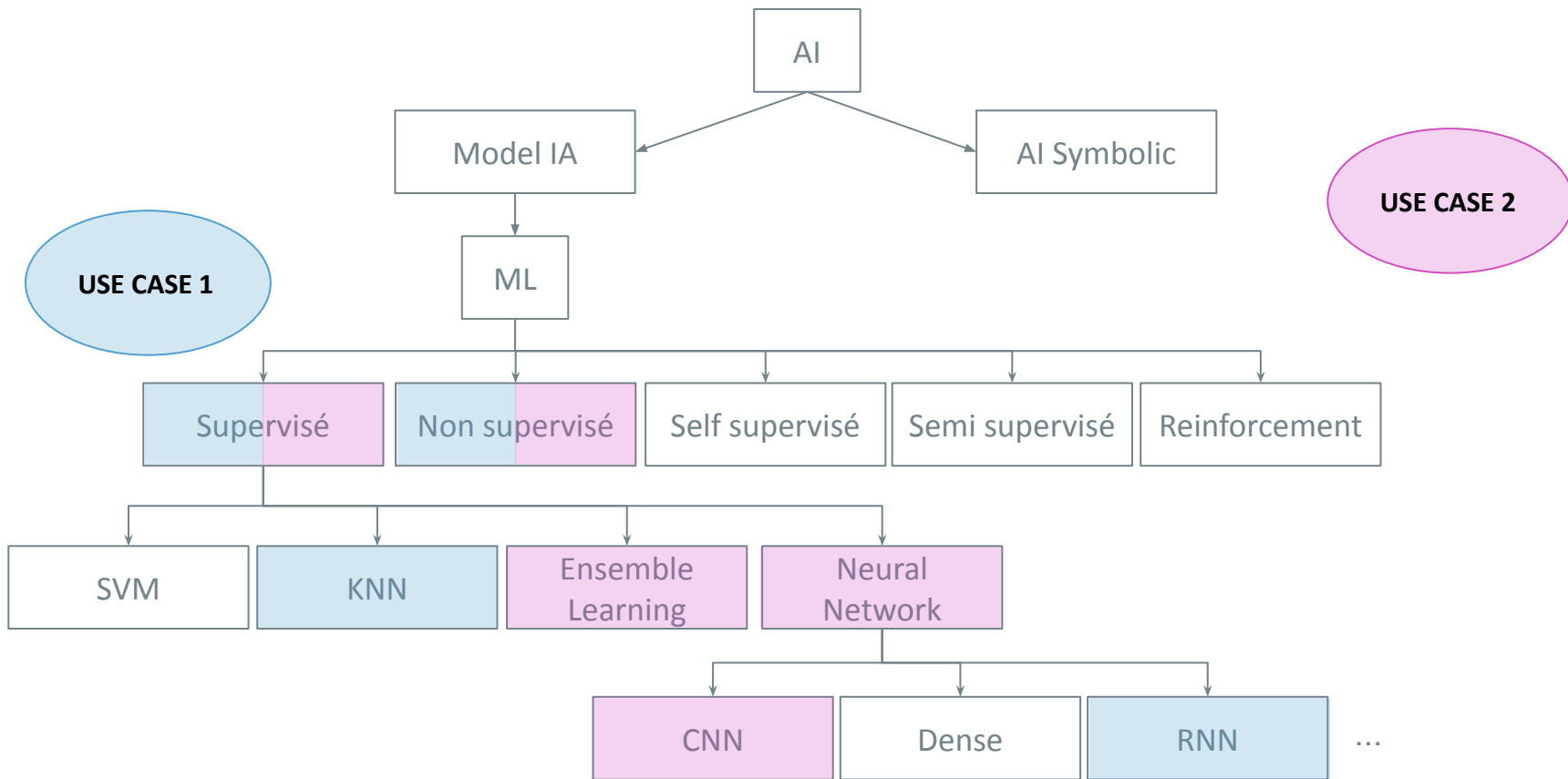
**[Lien notebook google colab "4 - Supervised learning"](#)**





[Lien QCM Application IA](#)

# Artificial Intelligence, Machine Learning et Deep Learning



# USE CASE n°2 : dataset sous forme raster

**Objectif** : se familiariser avec les ensembles de données non structurées (données d'image).

**Objectif pratique** : traitement d'image, classification de pixels.

*Inspection, visualisation et prétraitement des données :*

1. **Data inspection** : données manquantes, caractéristiques principales, la distribution des observations

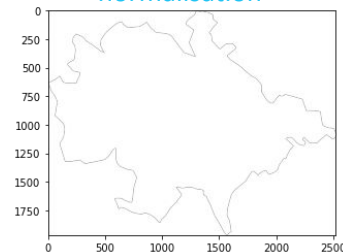
2. **Data preprocessing** :

a. **Normalisation des données**

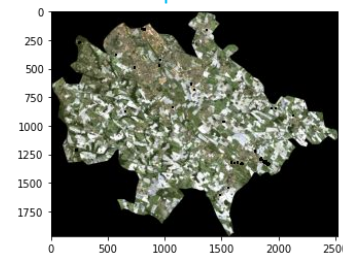
- Standardisation : centrer et réduire les données  $\rightarrow \frac{X - \mu}{\sigma}$
- Min Max scaling : transformation linéaire sur l'intervalle  $[0,1] \rightarrow \frac{X - X_{\min}}{X_{\max} - X_{\min}}$

=> ramener l'ensemble des données à des échelles communes

Avant  
normalisation



Après



# | 1 – Data preprocessing : traiter les valeurs nulles

**Objectif** : se familiariser avec les ensembles de données non structurées (données d'image).

**Objectif pratique** : traitement d'image, classification de pixels.

*Inspection, visualisation et prétraitement des données :*

## 2. Data preprocessing :

### b. Traitement des valeurs manquantes

- i. “Inpainting” méthode : reconstruction des pixels manquants / détériorés en utilisant le contexte des pixels voisins
- ii. Imputation / Interpolation : en ayant préalablement transformé l'image en un tableau 2D
- iii. KNN : pas assez de données ici pour avoir confiance en la méthode

# 1 – Data preprocessing : traiter les valeurs nulles



**METHODE INPAINTING  
DE LA LIBRAIRIE OPENCV**

Plusieurs algorithmes ont été conçus et OpenCV en fournit deux :

- Inpainting TELEA :
  - Auteur : article "An Image Inpainting Technique Based on the Fast Marching Method ", par Alexandru Telea en 2004
  - Méthode : la marche rapide depuis la limite d'une région vers l'intérieur, à l'aide de voisinage, on remplace le pixel manquant par la somme pondérée normalisée de tous les pixels connus dans le voisinage
- Inpainting NS
  - Auteur : "Navier-Stokes, Fluid Dynamics, and Image and Video Inpainting ", de Bertalmio, Marcelo, Andrea L. Bertozzi, et Guillermo Sapiro en 2001
  - Méthode : basée sur la dynamique des fluides et des équations différentielles partielles. Il parcourt d'abord les bords des régions connues vers les régions inconnues. Il poursuit les isophotes (lignes joignant des points de même intensité, tout comme les contours joignent des points de même élévation) tout en faisant correspondre les vecteurs de gradient à la limite de la région à re-peindre. Puis, la couleur est remplie pour réduire la variance minimale dans cette zone

# | 1 – Data preprocessing (TP)

USE CASE 2

**Application :**

**[Lien notebook google colab "5 - Image inspection"](#)**

## 2 - Segmentation non supervisée d'images

### Qu'est-ce que la segmentation d'image ?

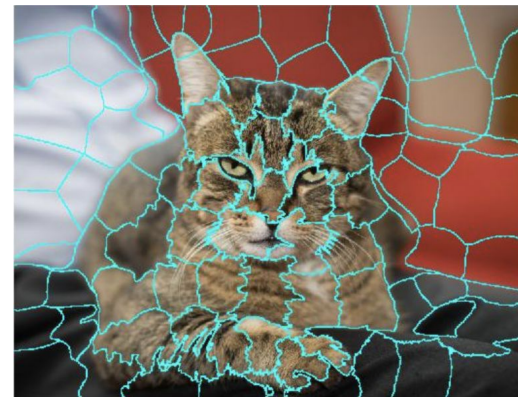
- Diviser ou partitionner une image
- Regrouper des pixels / des régions selon les propriétés des régions

### Exemple étudié : Segmentation d'images en superpixels

#### => Découpage en régions et superpixel

L'idée générale est de rechercher des groupes homogènes de pixels par apport à plusieurs critères :

- à la connection spatiale des pixels,
- la signature spectrale et
- la texture (consiste à calculer une série de mesures dans le but de définir une texture perçue sur une image, par exemple la moyenne du niveau de gris)



## | 2 - Segmentation non supervisée d'images (use cases)

### Quelles sont les applications de la segmentation d'images ?

- **Imagerie médicale** : permet de découper les images en zones pour compter des cellules, détecter des changements ou des anomalies.
- **Analyse d'images satellitaires** : permet de détecter les routes, les bâtiments ou les champs.
- **Voitures autonomes** : intégrées aux caméras embarquées dans les voitures autonomes. De manière générale, cette technique peut être utilisée à chaque fois que le tracé d'une bounding box (un rectangle) autour de l'objet ne suffit pas, et que l'on a besoin de plus de précision.



## | 2 - Segmentation non supervisée d'images (TP)

**Objectif n°1 (pre-processing)** : Regrouper des pixels afin de créer des groupes de pixels de même nature.

**Objectif n°2 (clustering)** : Appliquer des modèles non supervisés afin de détecter/identifier des parcelles agricoles des différentes cultures.

### Travail :

1. **Pre-processing** : application de l'algorithme de segmentation en superpixels Felzenszwalb.

Documentation :

<https://github.com/RaphaelGervillie/Superpixel>

<https://scikit-image.org/docs/stable/api/skimimage.segmentation.html#skimimage.segmentation.felzenszwalb>

Exemple : l'algorithme trouvera 5 superpixels (ie. 5 groupes homogènes de pixels différents),  
et pour chaque pixel la valeur du groupe correspondant (de 1 à 5) sera sauvegardée.

2. **Clustering** : l'algorithme de clustering Mini-batch K-means est une version de K-means, avec l'utilisation de petits lots de données aléatoires de taille fixe, utilisé à chaque itération pour mettre à jour les clusters.

Documentation : <https://thecleverprogrammer.com/2021/09/10/mini-batch-k-means-clustering-in-machine-learning/>

## | 2 - Segmentation non supervisée d'images (TP)

USE CASE 2

**Application :**

**[Lien notebook google colab "6 - Unsupervised superpixel segmentation"](#)**

# 3 - Classification supervisée des pixels

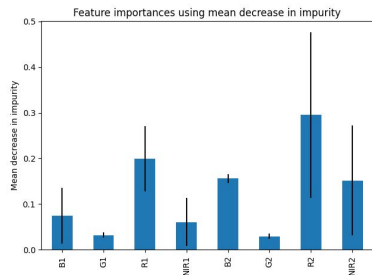
## Exercice :

**Construire un modèle (supervisé) de classification  
avec pour objectif de classer chaque pixel selon une nature de sol**

## Consignes :

- **Les algorithmes utilisées :**
  - Random Forest Classifier : avec ajustement des hyperparamètres selon une des méthodes suivantes :
    - Random Search : de manière aléatoire le modèle va tester un nombre de combinaison de paramètres
    - Grid Search : on définit des intervalles de définition des paramètres → plus coûteux

- **Analyse des variables influentes**



fonction “features importance”, sklearn.

## | 3 - Classification supervisée des pixels (TP)

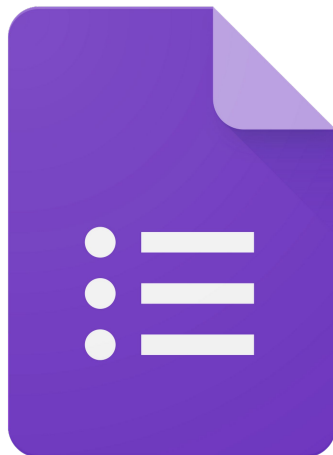
USE CASE 2

**Application :**

**[Lien notebook google colab "7 - Supervised pixel classification"](#)**

A lightbulb hangs from a cord on the right side of the frame. The background is a solid blue color with several out-of-focus, circular light spots (bokeh) scattered across it. The text "MILESKER - MERCI" is centered in the middle of the image.

MILESKER  
-  
MERCI



[Lien QCM Retour expérience](#)

*“Garapen ekonomikoa xedea baino gehiago baliabide bat dela uste du HUPIk”*

*“Le Développement Économique est un Moyen et pas une Finalité”*

HUPI SAS  
Technopole Izarbel  
2 Terrasse Claude Shannon  
64210 Bidart

HUPI IBERICA SLU  
Paseo Miramon N°170  
20009 Donostia  
/ San Sebastian

*contact@hupi.fr*

# Régression supervisée (DL)

## USE CASE 2

### Présentation d'un exemple réel :

Le LAI (Leaf Area Index) est un paramètre biophysique qui mesure la surface totale des feuilles par unité de surface au sol et qui est directement corrélé avec la quantité de lumière interceptée par la plante. Ce paramètre a de nombreuses utilisations telles que la prédiction de la production primaire photosynthétique, le suivi de la croissance des cultures et l'estimation du rendement. De plus, le LAI est requis par de nombreux modèles mondiaux du climat, de la productivité des écosystèmes et de l'écologie.

Les méthodes existantes sont trop complexes et prennent beaucoup de temps pour l'inférence (modèles d'inversion physique). Il n'existe pas de modèles opérationnels pour estimer le LAI des produits Sentinel-2 comme pour les produits MODIS LAI.

Avec DL, nous pouvons obtenir des valeurs de LAI similaires à celles obtenues par SNAP (Sentinel Application Platform) mais en 17 secondes environ contre 15 minutes pour SNAP.

**La famille de réseaux de neurones utilisés est : CNN 2D, car permet d'utiliser l'information spectrale, temporelle et spatiale. Utilisé aussi pour la détection de contour, la liaison de similarité/dissimilarité dans des images. Ici l'objectif était d'extraire de l'information spatiale, qui a été atteint grâce au réseaux convolutif et la robustesse de l'algorithme.**