

Entretien technique

Contexte : dépistage du cancer du poumon basé sur l'Intelligence Artificielle.

Lien vers le répertoire zip contenant les données et instructions : <https://drive.google.com/file/d/1-1sSgOqVt1eqwuxHeQRT0uySFOeHe19D/view?usp=sharing>

Début : lundi 24/01.

Temps de travail estimé : 8 - 10h.

Deadline d'envoi : le jeudi 27/01 à midi.

Présentation pour le vendredi 28/01 à 14h (heure française) :

- 4 slides minimum – 10 maximum (objectif, méthodes, résultats, pistes)

Conseil : n'hésitez pas à réutiliser vos connaissances acquises dans d'autres domaines pour enrichir les solutions proposées (statistiques, visualisation de données, clustering...)

Data

Les nodules pulmonaires sont extraits de scanners thoraciques de la base LIDC

<https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI>.

Les scanners sont des volumes pseudo 3D dans la mesure où ils consistent en un empilement de coupes 2D. La figure ci-dessous illustre un volume de nodule extrait d'un scanner, avec 5 coupes. La coupe du milieu sera appelée coupe centrale.

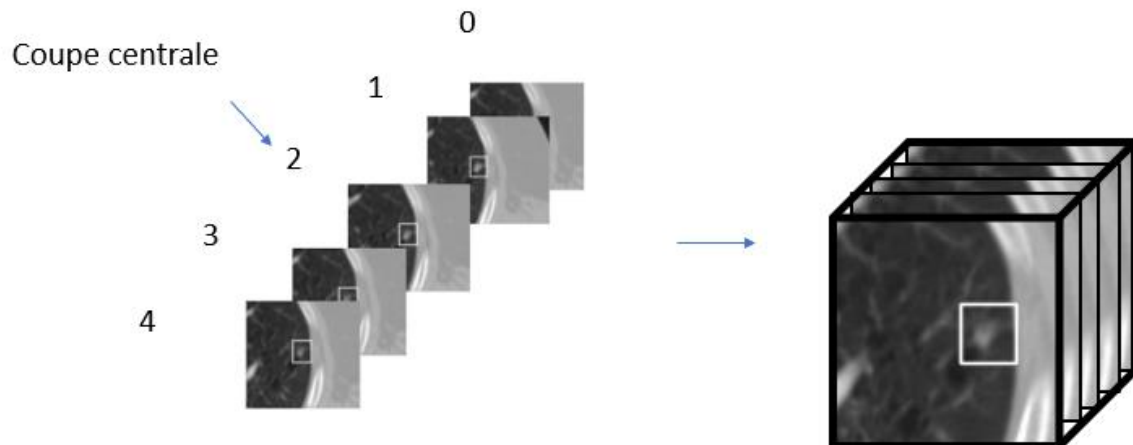


Figure 1: volume extrait d'un scanner correspondant à un nodule sur 5 coupes (à gauche). La figure de droite illustre le concept de pseudo-3D.

Les volumes des nodules sont au format `.npy` et les annotations au format `.csv`.

Une visualisation des nodules est possible grâce aux échantillons dans le répertoire `nodules_preview`. Il y a des annotations concernant

- la malignité (**malignant**) pour l'exercice Data Scientist décrit ci-dessous
- la texture (**density**) et spiculation (**spiculated**) pour l'exercice Data Manager décrit ci-dessous.

Instructions :

Code à réaliser sur Google Colab

```
[ ] from google.colab import drive
    drive.mount('/content/gdrive') # need authentication
```

```
!pip install tensorflow keras
!pip install scikit-learn
```

Data Manager

Objectif : attribuer un score selon la grille LungRads.

<https://pubs.rsna.org/doi/pdf/10.1148/rg.2015150079>

- Utiliser le code fourni pour entrainer l'algorithme à classer les nodules en fonction de leur texture (**density**).
- Segmenter la coupe centrale des volumes de nodules pour mesurer le diamètre et calculer une approximation du volume (rappel : $\frac{4}{3}\pi R^3$ où R est le rayon). Si possible, identifier le centre.
- Ecrire un algorithme permettant d'attribuer un score LungRads en fonction de la texture et du volume.

NB : on n'utilise pas les consignes liées à la croissance ou l'apparition d'un nouveau nodule (nous n'avons pas les données antérieures permettant d'évaluer la croissance d'un nodule)

- **Bonus** : dans cet exercice, vous aviez directement les volumes contenant des nodules. Pour détecter les nodules, il existe 2 approches :
 - o soit identifier des boîtes autour de régions d'intérêt (supposées correspondre à des nodules),
 - o soit segmenter l'ensemble du scan (0 : voxel ne contenant pas de nodule, 1 : voxel contenant un nodule) et utiliser un post-traitement pour prédire le centre des nodules à partir de la segmentation.

Pour la 2^e approche, indiquez quelle méthode statistique permettrait de prédire les centres des nodules. Vous pouvez vous inspirer d'articles état-de-l'art.