**PREDICTION CANCER DU SEIN**

**I – Identifier et définir le problème**

Le problème est : Pouvons-nous détecter des cellules cancéreuses sur une coupe d’échantillon de tumeur afin de localiser précisément les zones cancéreuses et ainsi évaluer la gravité du cancer ?

Pour répondre à cette question, le modèle va prendre comme entrée une image provenant de la base de données, et en sortie il dira s’il y a cancer ou non. Il s’agit donc d’une classification.

**II – Comprendre les données**

* Charger la base de données dans un tableau

Pour pouvoir utiliser la base de données que nous avons téléchargé, il faut la charger dans un tableau. Ce tableau sera composé d’une colonne avec l’image (c’est à dire un vecteur à 3 dimensions car les images font 50x50 pixels et sont en couleur) et d’une colonne label, qui pour chaque image, sera 1 si cancéreux, 0 sinon.

Pour cela, nous parcourons récursivement les dossiers et les sous dossiers car la base de données téléchargée est triée par patient (il y a 279 patients, et pour chacun deux dossiers avec des photos cancéreuses et non cancéreuses). Il y a 277 524 photos dont 198 738 IDC négative et 78 786 IDC postive.

**III – Choisir une méthode**

**IV – choisir un modèle et comprendre les grandes lignes**

(Être capable en 2-3 phrases de faire comprendre à vos collègues) le fonctionnement du modèle et le mécanisme d'apprentissage sous-jacent

**V – Appliquer le modèle et commenter les résultats**