**Introduction**

L'algorithme d'apprentissage automatique développé permet d'analyser les images 1000 fois plus rapidement que les humains. Les grands acteurs du numérique ont basé de nombreux projets dans le domaine de la santé sur le *Machine Learning* autour notamment du séquençage génomique.

**Contexte**

La pneumonie est une infection qui enflamme les sacs aériens d'un ou des deux poumons. Les sacs aériens peuvent se remplir de liquide ou de pus (matière purulente), ce qui provoque une toux avec mucus ou plus, de la fièvre, des frissons et des difficultés respiratoires.

En 2017, 2,56 millions de personnes sont décédées des suites d'une pneumonie. Près d’un tiers de ces personnes étaient des enfants de moins de 5 ans. L'OMS a estimé que 45 000 de ces décès prématurés étaient dus à la pollution de l'air domestique. Avec une amélioration de l’efficacité des diagnostics, bon nombre de ces décès peuvent être réduits.

Avec l’évolution de Covid-19 les services hospitaliers sont débordés et cherchent un outil informatique qui permet d’aider les docteurs à détecter rapidement et en masse la pneumonie à partir des images X-ray données.

L'objectif de ce projet est de construire un modèle d'apprentissage automatique optimisé afin qu’il puisse aider les radiologues à détecter au plutôt la pneumonie à partir des rayons X du thorax, et pouvoir sauver bon nombre de vies.

**Matériels**

1. Outils à dispostion

Tout au long de ce projet nous avons décidé d’utiliser le langage Python et ses librairies Numpy, Pandas, sklearn et tensorflow.

Jupyter : environnement de programmation interactif basé sur le Web permettant de créer des documents Notebook (document JSON).

1. Le jeu de données

Le jeu de données est organisé en 3 dossiers (train, test, val) et contient des sous-dossiers pour chaque catégorie d'images (Pneumonie/Normal). Il y a 5 866 images de radiographie (JPEG) et 2 catégories (Pneumonie/Normal). Les images de radiographie du thorax (antérieures-postérieures) ont été sélectionnées à partir de cohortes rétrospectives de patients pédiatriques âgés d’un à cinq ans du Guangzhou Women and Children's Medical Center, Guangzhou. Toutes les radiographies du thorax ont été réalisées dans le cadre des soins cliniques de routine des patients. Pour l'analyse des images radiographiques thoraciques, toutes les radiographies thoraciques ont d'abord fait l'objet d'un contrôle de qualité en éliminant tous les scans de mauvaise qualité ou illisibles. Les diagnostics des images ont ensuite été évalués par deux médecins experts avant d'être autorisés à entraîner le système d'IA. Afin de tenir compte d'éventuelles erreurs de classement, l'ensemble d'évaluation a également été vérifié par un troisième expert.

**Méthodes**

1. Préparation des données

* Aplatir les données dans un tableau unidimensionnel (flatten) (règle le problème des différentes tailles d’images)
* Obtenir un histogramme de couleur en 3 dimensions de l’espace colorimétrique HSV
* Normaliser les données (valeurs des pixels ramenées entre 0 et 1)
* Convolution : Amélioration de la netteté : augmenter le poids des pixels en combinant les voisins
* Mise en commun : Réduire la taille de l’image en combinant les pixels : utilisation du pooling max (addition des valeurs )

1. Les différents modèles d’apprentissage

Algorithme KNN

Régression linéaire

SVM

Deep neural network (DNN)

1. Stratégie et choix du modèle

* Evaluation de la précision
* Evaluation de la perte

**Résultats (**et optimisation du modèle)