

Détection de la mutation PIK3CA dans le cancer du sein par OWKIN

La Team

Antoine Gauchot, Anisse Id-Benaddi, Loris Megy,
Adrien Moulinas, Pierre Samaha,

Contexte

Données

- 334 scans d'une lame histologique
- 1000 tuiles extraites de chaque scan
- 2048 features extraites de chaque tuile

Supervision faible

- Pas de label individuel pour chaque tuile
- Label global pour chaque scan

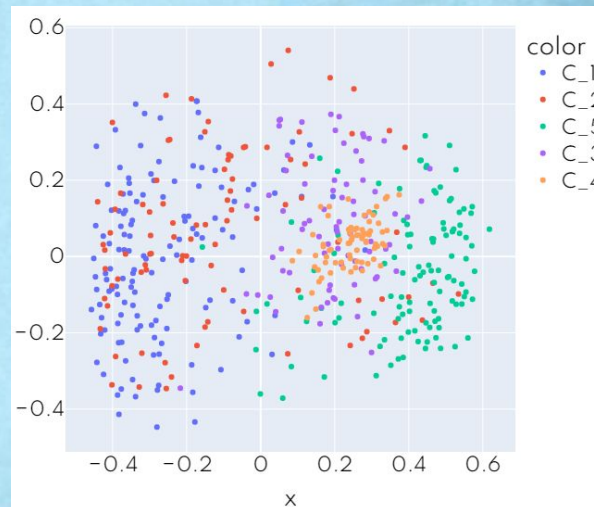
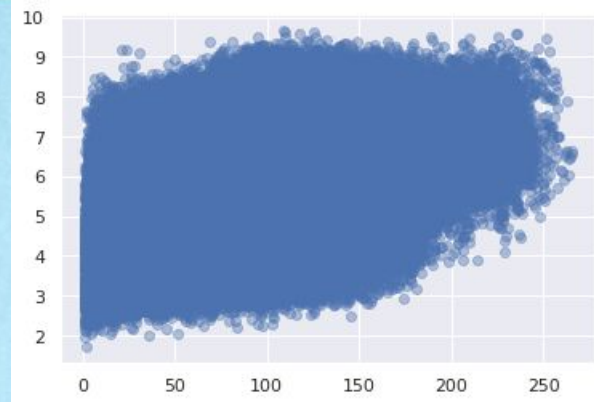
Scans de centres différents

Analyse des données

Vecteurs de features et métadonnées

Importance de l'origine des images
(Hôpital et positions spatiales)

Réduction des données ? (PCA)



Approches diverses

Une approche vocabulaire

- i/ Réduire la dimension avec une PCA
- ii/ Déterminer N clusters de tuiles sur l'ensemble des scans (344K vecteurs) (K-means, DBSCAN)
- iii/ Chaque scan est alors représenté par sa proportion de tuiles appartenant à chaque cluster (vecteur de taille N)
- iv/ régression logistique sur les vecteurs de proportion représentant chaque scan

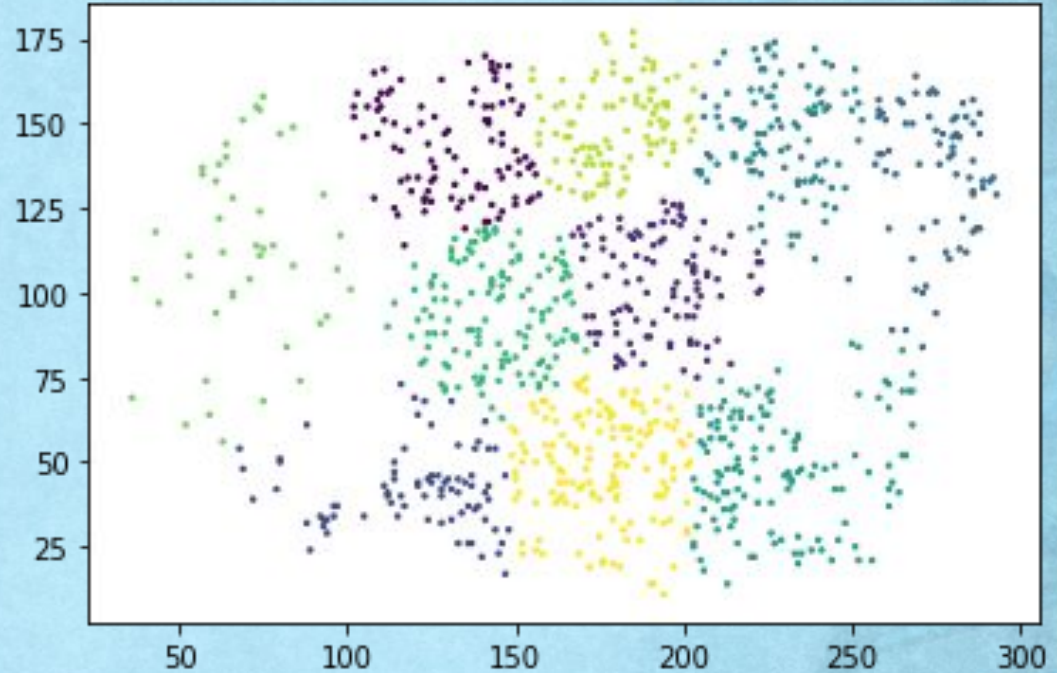
Clustering spatial

i/ Reprendre la Baseline avec un max.

ii/ Clustering spatial.

iii/ Pour chaque scan appliquer la Baseline.

iv/ Régression logistique sur les vecteurs des prédictions ordonnées.



Modèle

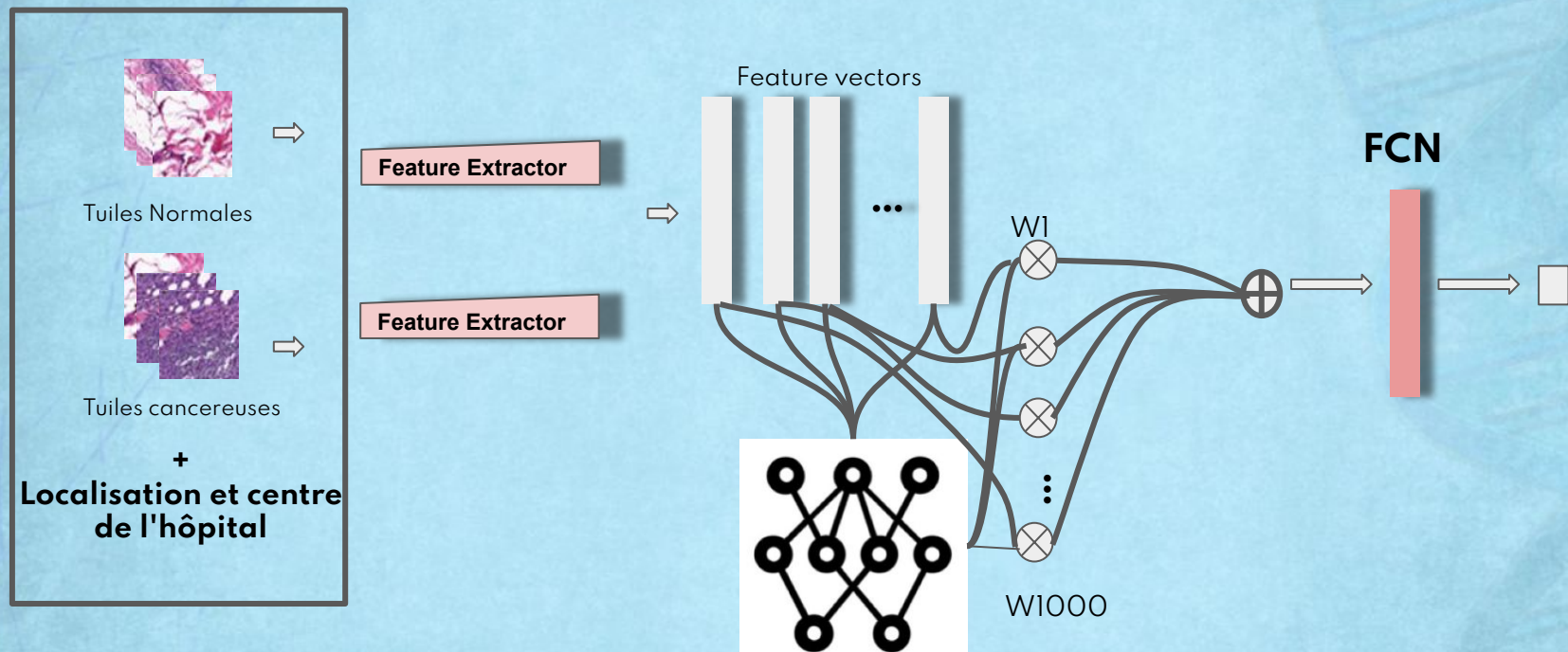
Classification avec Attention-based Deep Multiple Instance Learning (MIL)

- Entrée : les 1000 vecteurs de taille 2048+3
- **Couche d'attention** pour attribuer un score d'attention à chaque tuile
- **Multiplication des features** avec leur score d'attention
- **Concaténation** avec les métadonnées zoom et id du centre
- **Softmax** pour classifier

Le **mécanisme d'attention** permet de déterminer quelles tuiles dans chaque scan a le plus d'impact sur le label -> permet de résoudre en partie la supervision faible

Réseau invariant par permutation des tuiles

Modèle et Mécanisme d'attention



Résultats

Mieux que la baseline : AUC de **0.6466**

Mieux qu'hier, moins bien que demain

That's all Folks!