

# Prédiction de biomarqueur sur des lames histologiques de tissus cancéreux

Presenté par :  
Mohamed GHAMRI

Encadré par :  
M. Camille KURTZ  
M. Nicolas LOMENIE

# Plan de présentation

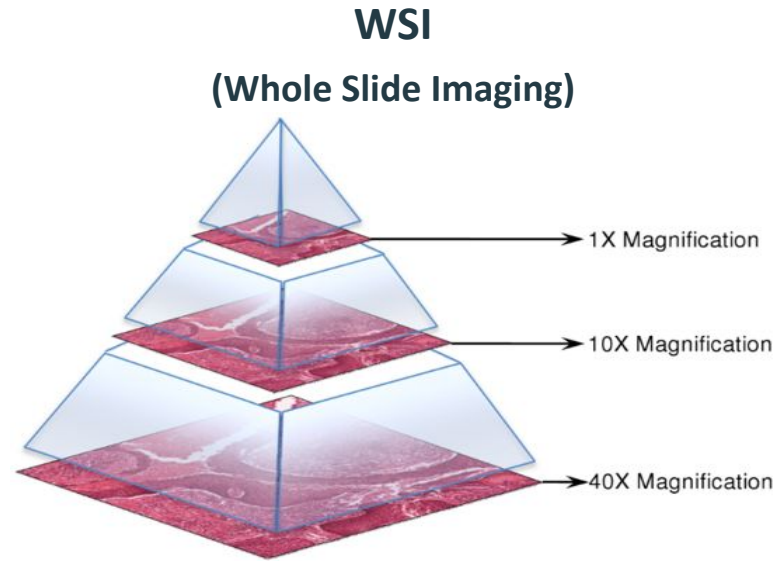
1. Introduction
2. Ensemble de jeu
3. Méthode
4. Résultats
5. Conclusion
6. Limite et Perspective
7. Références

# Introduction

## Cancer du sein

Le cancer du sein est l'une des affections les plus fréquemment diagnostiquées chez les femmes à l'échelle mondiale.

Les médecins utilisent généralement des lames histologiques pour détecter les tumeurs.



# Ensemble de jeux

	WSI	Entraînement	Test
Camelyon16	<b>400 WSI</b>	<b>270 WSI</b>	<b>130 WSI</b>
Classification des patches	<b>5 WSI (171883 patches 8,9% tumeurs)</b>	<b>70% , 10%</b>	<b>20% (34376 patches )</b>
Classification des slides	<b>58 WSI (29 tumeurs, 29 normal)</b>	<b>70%</b>	<b>30% (17 WSI)</b>

# Méthode

# Architecture générale

## Prétraitement

Générer les masques  
(Grand Truth)

Segmentation et  
suppression de fond

Générer les paires de  
patches

## Classification des Patches

l'entraînement de  
modèle (FCN)

Tester le modèle (FCN)

Générer les heatmaps

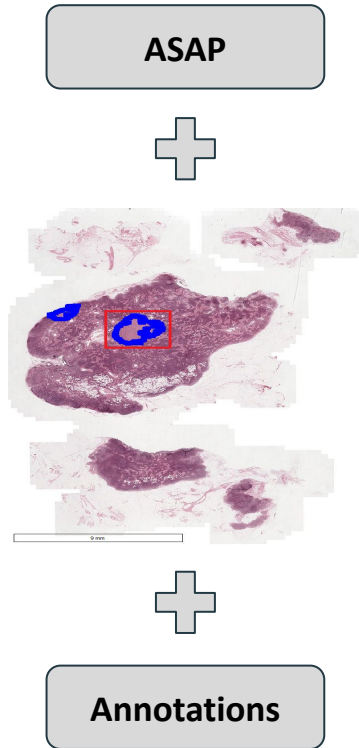
## Classification des slides

Extraire les  
caractéristique pour le  
classificateur

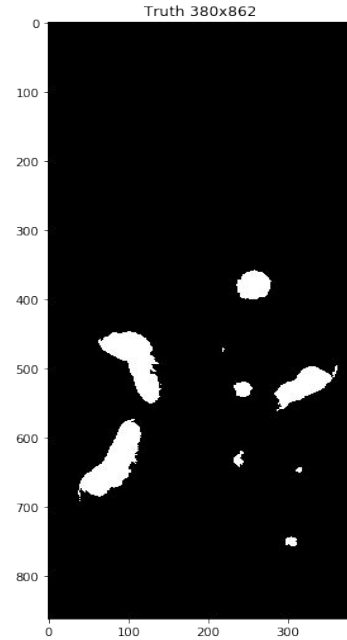
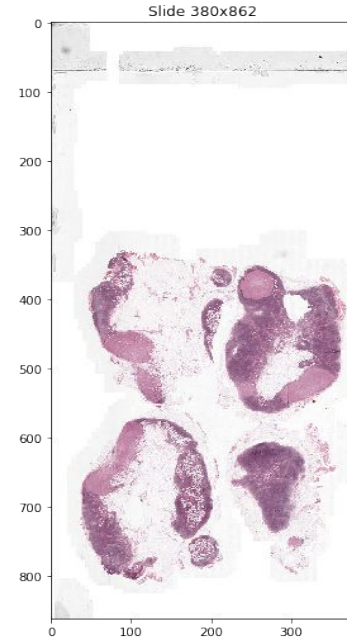
l'entraînement de  
Random Forest

Teste le classificateur  
Random forest

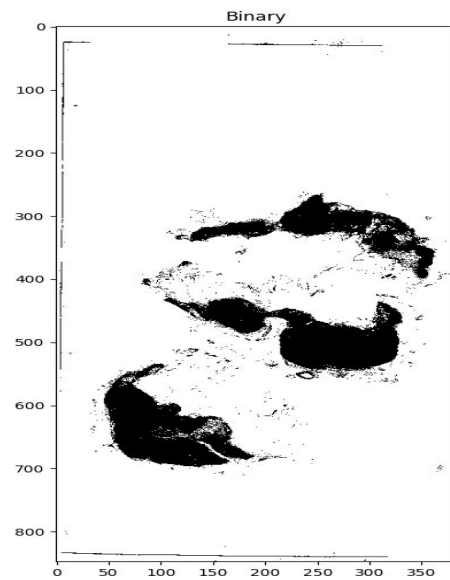
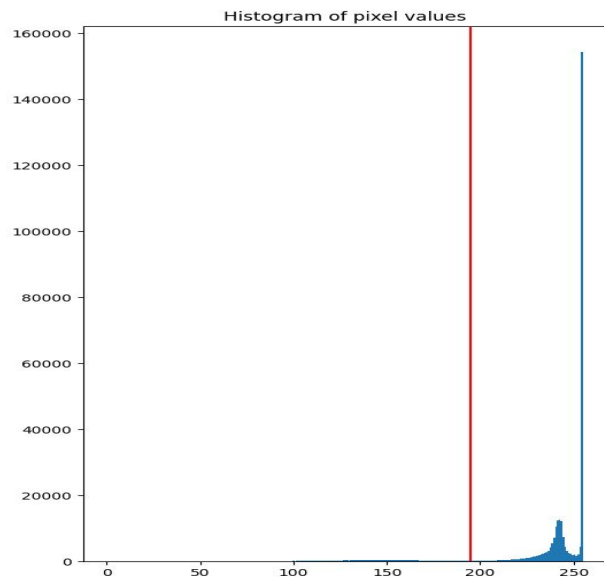
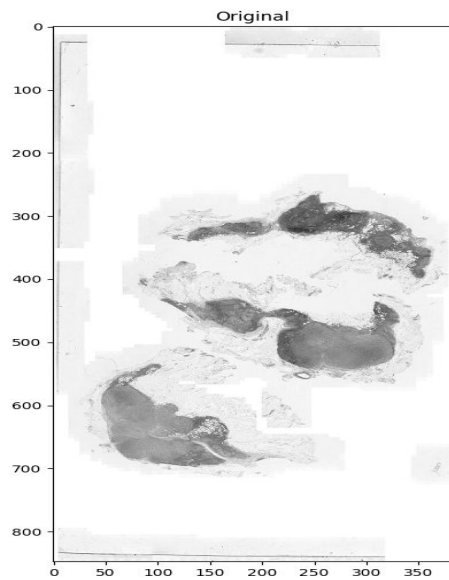
# Prétraitement (1)



Slide & Truth Thumbnails (downsampled 256x)

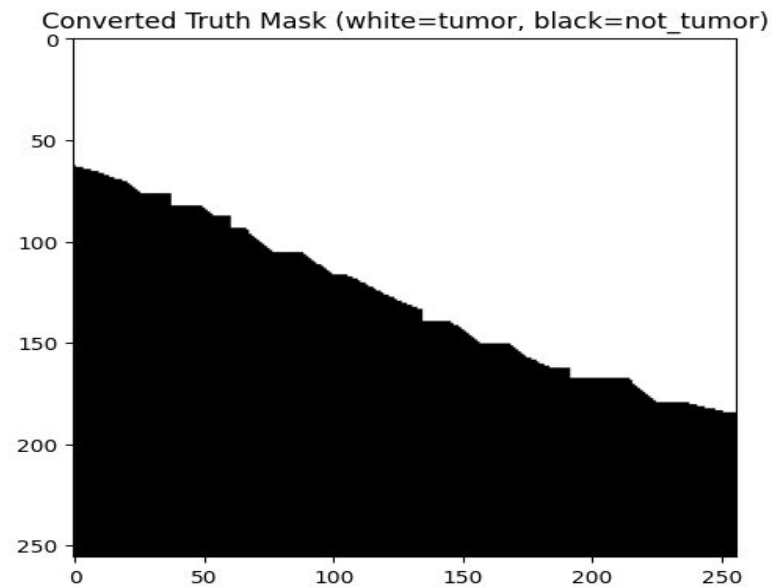
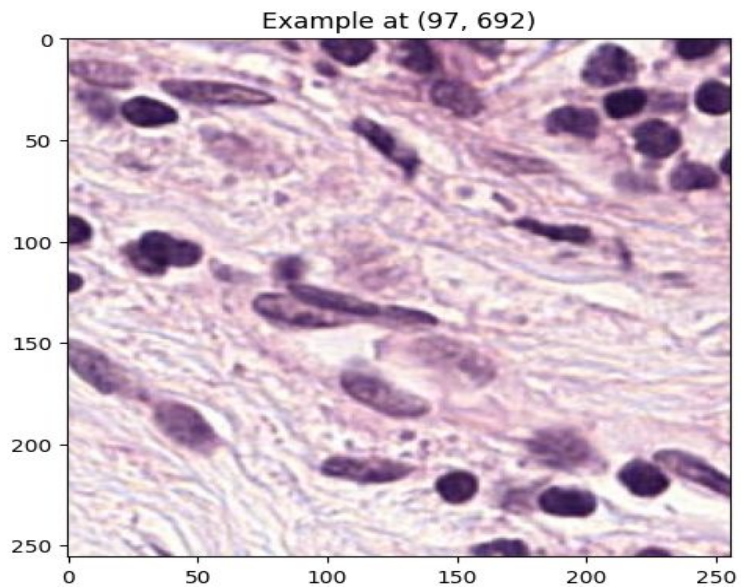


## Prétraitement (2)





## Prétraitement (3)



# Classification des patches

Layer (type)	Output Shape	Param #
lambda (Lambda)	(None, 256, 256, 3)	0
conv2d (Conv2D)	(None, 128, 128, 100)	7600
max_pooling2d (MaxPooling2D)	(None, 64, 64, 100)	0
conv2d_1 (Conv2D)	(None, 32, 32, 200)	500200
max_pooling2d_1 (MaxPooling2D)	(None, 16, 16, 200)	0
conv2d_2 (Conv2D)	(None, 16, 16, 300)	540300
conv2d_3 (Conv2D)	(None, 16, 16, 300)	810300
dropout (Dropout)	(None, 16, 16, 300)	0
conv2d_4 (Conv2D)	(None, 16, 16, 2)	602
conv2d_transpose (Conv2DTranspose)	(None, 256, 256, 2)	3846

## Hyperparamètres

Taille de batch : 32

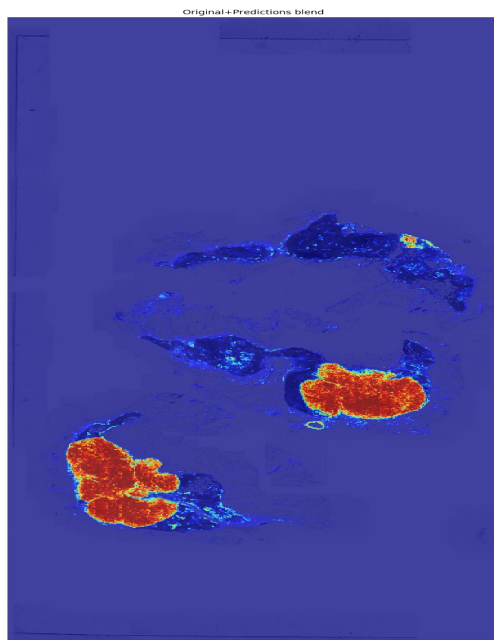
Fonction d'activation: ELU > RELU (vanishing)

Technique de régularisation : Dropout

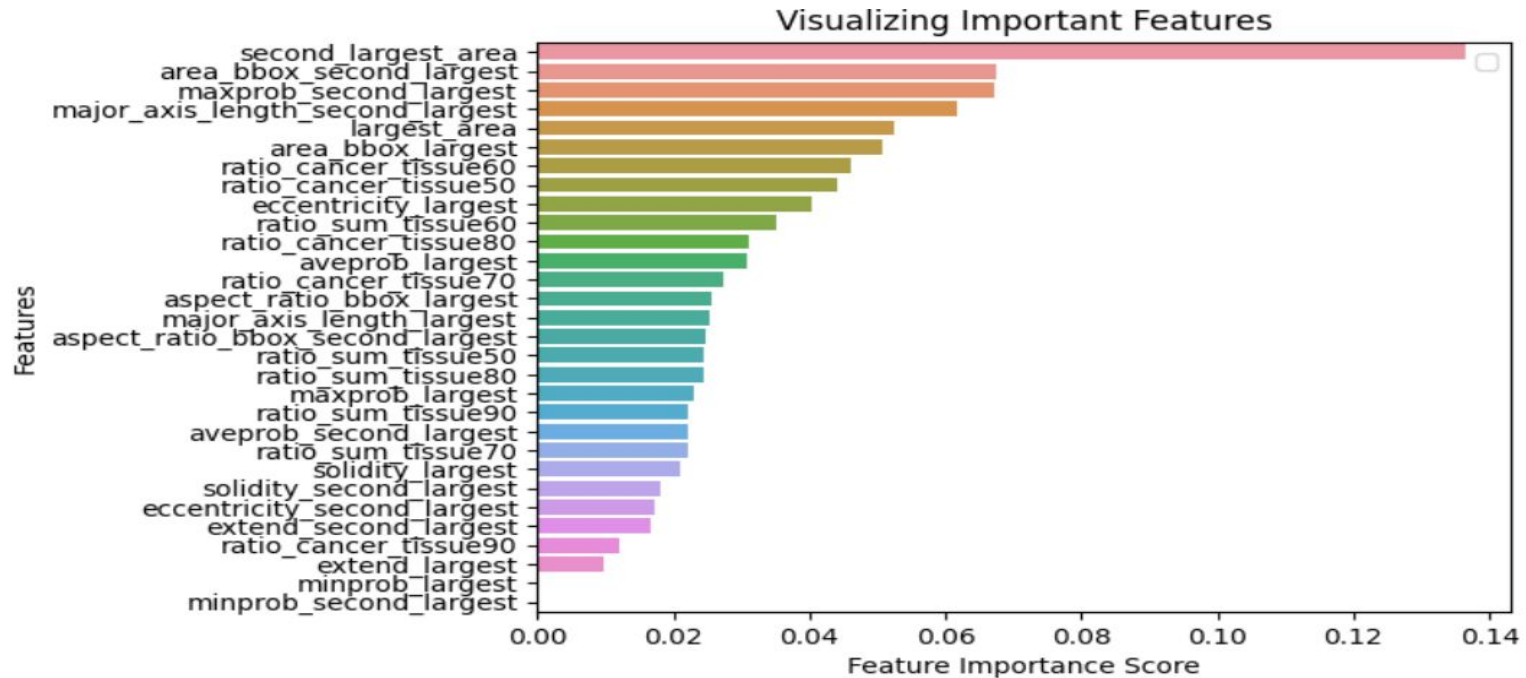
Fonction de Perte : categorical Cross-Entropy

Optimiseur : ADAM

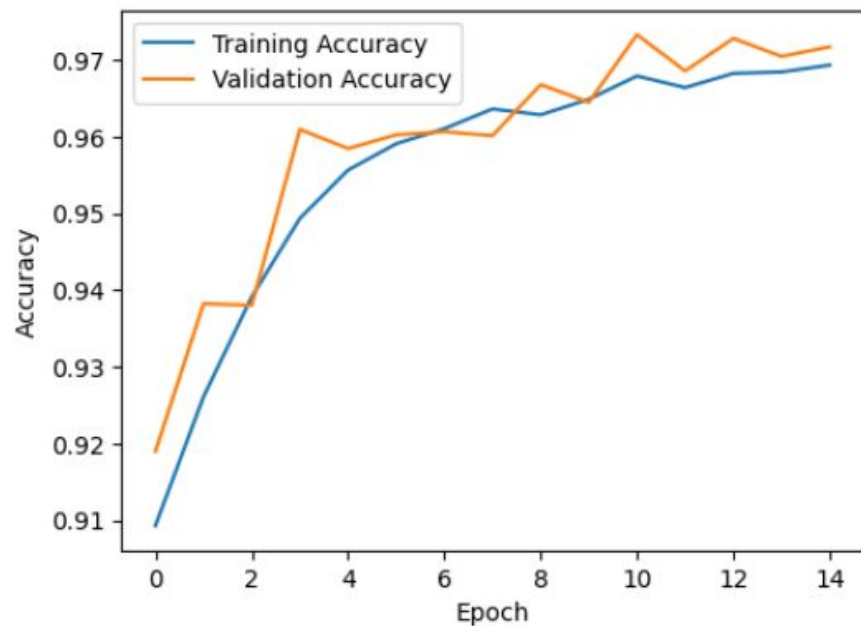
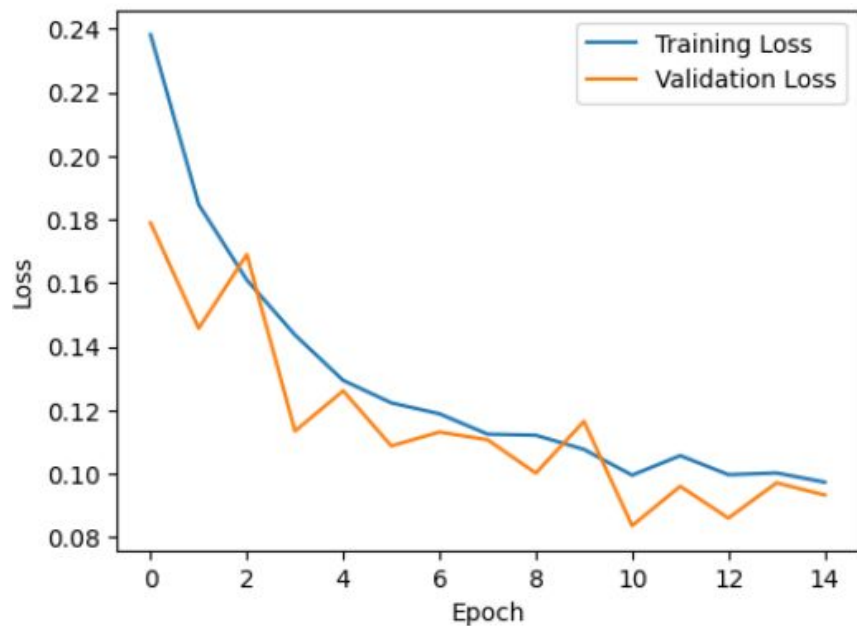
# Générer les Heatmaps



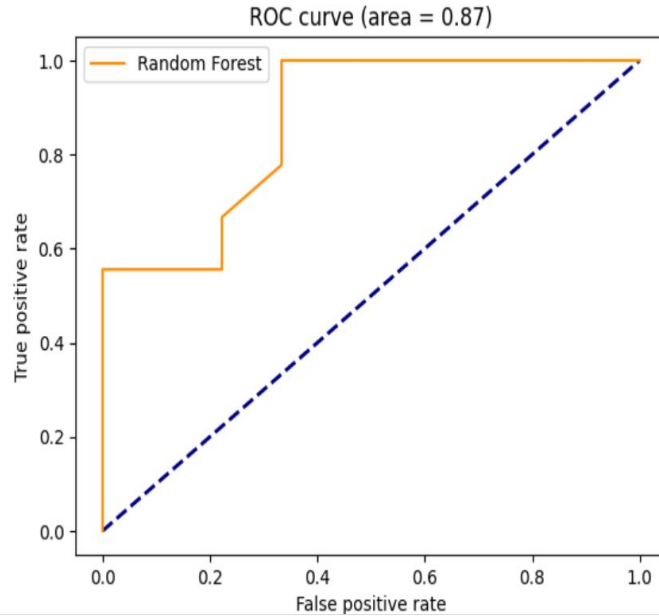
# Classification des slides



# Résultats (1)



## Résultats (2)



	Accura cy	Recal	Précisio n	F1 -score	AUC
Classificati on des patches	<b>0,97</b>	<b>0,80</b>	<b>0,84</b>	<b>0,82</b>	<b>/</b>
classificati on des slides	<b>0.83</b>	<b>0.85</b>	<b>0.85</b>	<b>0.85</b>	<b>0.87</b>

# Conclusion

La méthode démontre qu'elle est acceptable pour la détection du cancer du sein et qu'elle peut contribuer à la prise de décision.

Les résultats de cette méthode montrent qu'elle est comparable à d'autres méthodes différentes, voire même aux transformeurs.

# Limites

La réalisation de cette méthode demande beaucoup de ressources matérielles et de temps.

# Perspectives

Modifier le modèle de classification des lames en utilisant un modèle plus performant (deep learning), augmenter la taille de l'ensemble de données.

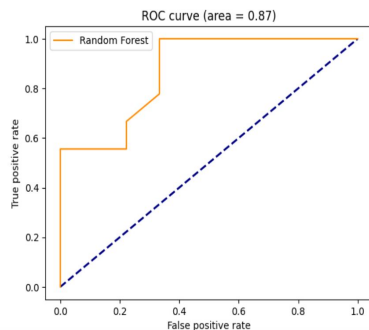


Merci pour votre attention

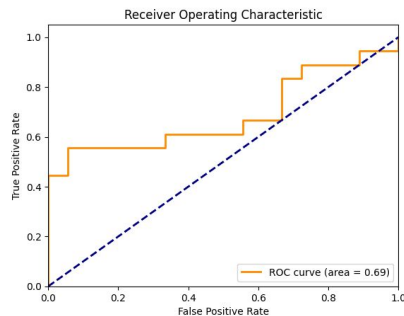
# Références

- Deep Learning for Identifying Metastatic Breast Cancer (<https://arxiv.org/pdf/1606.05718.pdf>)
- Babak Ehteshami Bejnordi et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer.
- <https://camelyon16.grand-challenge.org/>

# Comparaison des deux méthodes



Random forest - AUROC = 0.87



CLAM - AUROC = 0.69

(peut atteindre 0.9 en fonction des données)

STEP 1

Étape commune : extraction de features (avec ou sans apprentissage)

STEP 2

Étape différente : agrégation (DMIL ou random forest)

Les deux méthodes sont envisageables, le choix dépend du nombre de données et de labels disponibles :

- Si il y'a beaucoup de données mais peu de labels : CLAM est une bonne option.
- Si le nombre de données est moyen mais qu'il y'a assez de labels : il vaut mieux opter pour l'agrégation par random forest.