

N° candidat 47392, TIPE 2022

Détection de tumeurs grâce à l'intelligence artificielle

Grégoire PELLETIER

Thème : Santé prévention

Sommaire

- ➊ Introduction
 - Utilité
 - Principe de fonctionnement
- ➋ Traitement d'images
 - Morphologie
 - Reconnaissance d'objet
- ➌ Modélisation 2D
 - Propagation
 - Modélisation 2D
- ➍ Intelligence artificielle
 - Influence de la taille du dataset
- ➎ Modélisation 3D

Sommaire

- 1 Introduction
 - Utilité
 - Principe de fonctionnement
- 2 Traitement d'images
 - Morphologie
 - Reconnaissance d'objet
- 3 Modélisation 2D
 - Propagation
 - Modélisation 2D
- 4 Intelligence artificielle
 - Influence de la taille du dataset
- 5 Modélisation 3D

Utilité

- La majorité des tumeurs, pour le cancer du sein, sont détectées par mammographie. Parfois, des lésions malignes passent inaperçues et d'autres lésions suspectes s'avèrent bénignes
- L'IA détecte des tumeurs malignes avec un meilleur taux de réussite qu'un radiologue

Utilité

- Une détection plus rapide et efficace permet un meilleur traitement, prévient l'avancée du cancer et améliore les chances de guérison du malade
- L'IA est donc très utile pour les radiologues, et permet la détection précoce de cancers

Principe de fonctionnement

- Problème de classification entre groupe "sain" et "malade"
- On utilise la méthode des plus proches voisins

Principe de fonctionnement

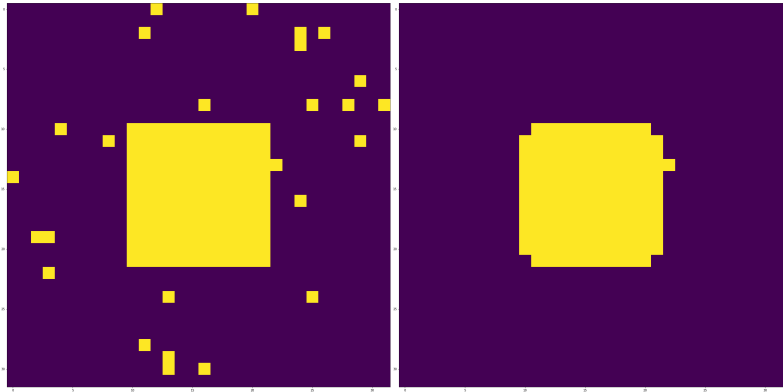
- On place les points de la base d'entraînement (groupe connu) sur un graphique, le modèle attribue un groupe aux points de la base de test (groupe inconnu) selon ses k plus proches voisins sur le graphique
- On vérifie les taux de réussite des différents modèles, on conserve le meilleur
- On valide le meilleur modèle sur la base de validation (élimine "l'overfitting")

Sommaire

- 1 Introduction
 - Utilité
 - Principe de fonctionnement
- 2 Traitement d'images
 - Morphologie
 - Reconnaissance d'objet
- 3 Modélisation 2D
 - Propagation
 - Modélisation 2D
- 4 Intelligence artificielle
 - Influence de la taille du dataset
- 5 Modélisation 3D

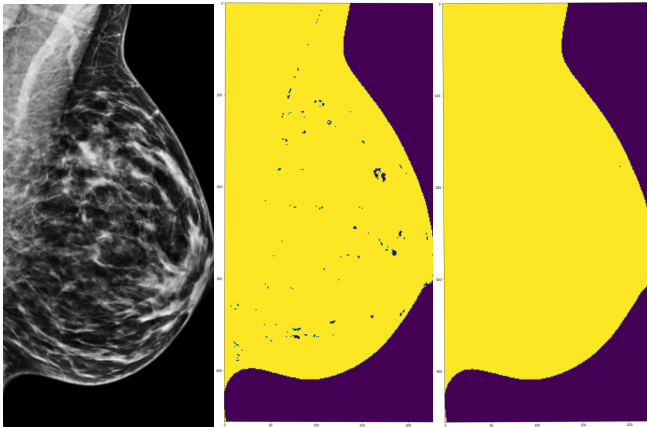
Morphologie

- dilation et érosion

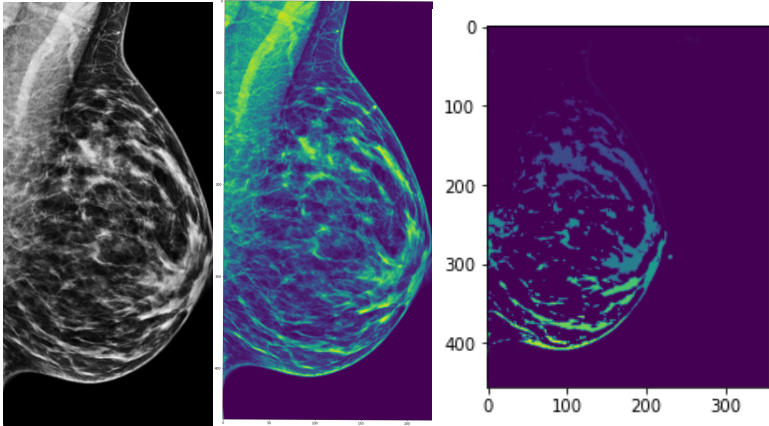


Morphologie

- Masque

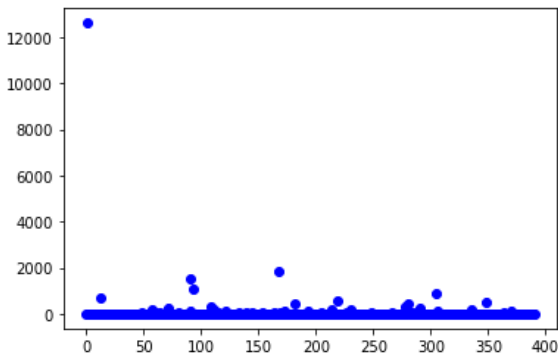


Reconnaissance d'objet



Reconnaissance d'objet

- Histogramme des tailles (en pixels) des groupes



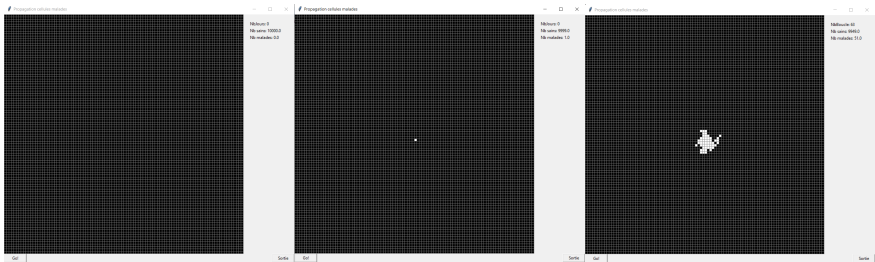
Sommaire

- 1 Introduction
 - Utilité
 - Principe de fonctionnement
- 2 Traitement d'images
 - Morphologie
 - Reconnaissance d'objet
- 3 Modélisation 2D**
 - Propagation
 - Modélisation 2D
- 4 Intelligence artificielle
 - Influence de la taille du dataset
- 5 Modélisation 3D

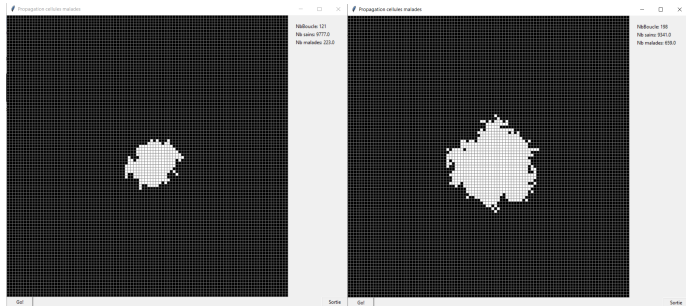
Propagation à partir d'une cellule "malade"

- On va étudier la propagation de la tumeur à partir d'une première cellule malade, la probabilité de se propager aux voisins sera aléatoire (et inférieure à 0.1)
- On visualise la propagation en regardant l'état des cellules au fil des boucles

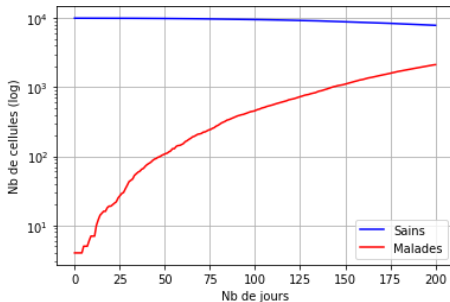
Propagation à partir d'une cellule "malade"



Propagation à partir d'une cellule "malade"

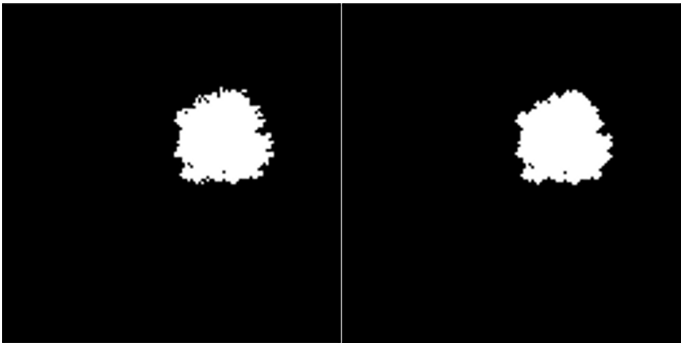


Evolution du nombre de cellules malades au court du temps



Modélisation 2D

- On automatise le processus et on utilise les procédés de traitement d'images précédents



Modélisation 2D

- On récupère deux données : le nombre total de cellules malades (grâce au traitement d'images), et le maximum des distances des cellules malades à la plus proche cellule saine
- On considère qu'il y a présence de tumeur maligne à partir d'un certain seuil de cellules malades, ou une certaine distance maximale, c'est à dire si la tumeur est suffisamment étendue

Modélisation 2D

- Pour cela on crée le tableau des distances à la cellule saine la plus proche

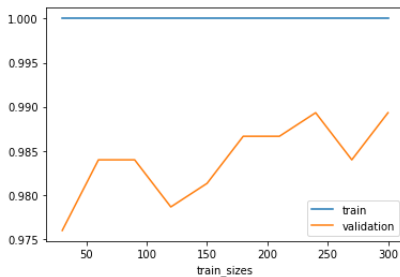
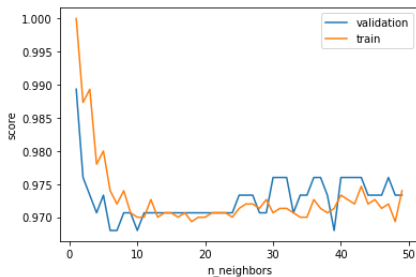
	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	
39	0	0	0	0	0	0	0	0	1	1	1	2	3	3	3	3	3	3	4	5	5	5	5	5	5	4	4	3	2	2	1	1	0	0	0	0		
40	0	0	0	0	1	1	0	1	1	2	2	2	3	4	4	4	4	4	4	5	6	6	6	6	5	5	4	3	3	3	2	2	1	0	0	0	0	
41	0	0	0	1	1	1	1	1	2	2	3	3	3	4	5	5	5	5	5	6	7	7	6	6	5	4	4	4	3	2	1	1	0	0	0	0		
42	0	0	0	1	2	2	2	2	2	3	3	4	4	4	5	6	6	6	6	6	7	7	7	7	6	5	5	5	4	3	2	2	1	0	0	0		
43	0	0	1	1	2	2	3	3	3	3	4	4	5	5	5	6	7	7	7	7	7	8	7	6	6	6	5	4	3	3	2	1	1	0	0	0		
44	0	0	0	1	1	2	2	3	4	4	4	5	5	6	6	6	7	8	8	8	8	8	7	7	7	7	6	5	4	3	3	2	2	1	1	0	0	
45	0	0	0	0	1	1	2	3	4	5	5	5	6	6	7	7	7	8	9	9	9	9	8	8	8	7	6	5	4	3	2	2	1	1	0	0	0	
46	0	0	0	0	0	1	2	3	4	5	6	6	6	7	7	8	8	8	9	10	10	9	9	9	8	7	6	5	4	3	2	1	1	0	0	0	0	
47	0	0	0	0	1	1	2	3	4	5	6	7	7	7	8	8	8	9	9	10	10	10	10	9	8	7	6	5	4	3	2	1	0	0	0	0	0	
48	0	0	0	0	1	2	2	3	4	5	6	7	8	8	8	9	9	10	10	10	11	11	11	10	9	8	7	6	5	4	3	2	1	1	1	0	0	1
49	0	0	0	0	1	2	3	3	4	5	6	7	8	9	9	9	10	10	11	11	11	11	10	9	8	7	6	5	4	3	2	2	2	1	1	1	1	
50	0	0	0	0	1	2	3	4	4	5	6	7	8	9	9	9	10	10	11	12	11	10	9	8	7	6	5	4	3	3	2	2	2	2	2	1	1	
51	0	0	0	1	1	2	3	4	5	5	6	7	8	8	8	9	9	10	11	11	11	10	9	8	7	6	5	4	4	4	3	3	3	3	2	1	1	
52	0	0	0	0	1	2	3	4	5	5	6	7	7	7	7	8	8	9	10	10	10	10	10	9	8	7	6	5	5	5	4	4	3	2	2	1	1	
53	0	0	0	1	1	2	3	4	4	5	6	6	6	6	7	7	8	9	9	9	9	9	9	8	7	6	6	6	5	5	5	4	3	2	2	1	1	
54	0	0	0	1	2	2	3	3	4	5	5	5	5	6	6	7	8	8	8	8	8	8	8	8	7	7	7	7	6	6	5	5	4	3	2	1	0	0
55	0	0	1	1	2	2	2	3	4	4	4	4	5	5	6	7	7	7	7	7	7	7	7	7	7	7	7	6	6	5	5	4	4	3	2	1	1	1
56	0	0	0	1	1	1	2	3	3	3	3	4	4	5	6	6	6	6	6	6	6	6	6	6	6	6	6	5	5	4	4	3	3	2	2	1	1	
57	0	0	0	0	0	1	2	2	2	2	3	3	4	5	6	5	5	5	5	5	5	5	5	5	5	5	6	6	5	4	4	3	3	2	2	1	1	
58	0	0	0	0	0	1	1	1	1	2	2	3	4	5	6	5	4	4	4	4	4	4	4	4	4	5	6	6	5	4	3	3	2	2	1	1	0	0
59	0	0	0	0	0	0	1	0	1	1	2	3	4	5	5	5	4	3	3	3	3	3	3	3	4	5	5	5	4	3	2	2	1	1	0	0	0	
60	0	0	0	0	0	0	1	1	0	1	2	3	4	4	4	4	4	3	2	2	2	2	2	3	4	5	4	4	4	4	3	2	1	1	0	0	0	0
61	0	0	0	0	0	1	1	1	1	1	2	3	3	3	3	3	3	3	2	1	1	1	2	3	4	4	4	3	3	3	2	1	0	0	0	0	0	0

Sommaire

- 1 Introduction
 - Utilité
 - Principe de fonctionnement
- 2 Traitement d'images
 - Morphologie
 - Reconnaissance d'objet
- 3 Modélisation 2D
 - Propagation
 - Modélisation 2D
- 4 Intelligence artificielle**
 - Influence de la taille du dataset**
- 5 Modélisation 3D

dataset composé de 500 grilles

- Pour 500 grilles, les résultats sont bons, mais quelques incohérences sont présentes, et le modèle peut encore être amélioré



dataset composé de 500 grilles

Code

```
k = np.arange(1, 50)
train_score, val_score = validation_curve(model,
                                          X_train, Y_train, 'n_neighbors', k, cv=5 )
plt.plot(k, val_score.mean(axis=1), label='validation')
plt.plot(k, train_score.mean(axis=1), label='train')
plt.ylabel('score')
plt.xlabel('n_neighbors')
plt.legend()
```

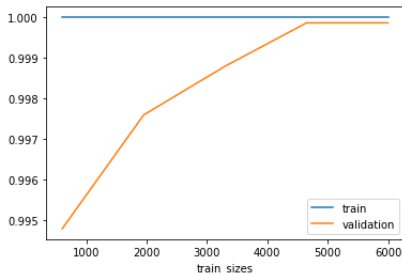
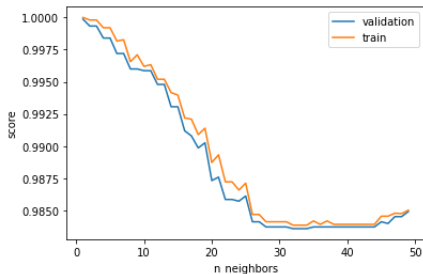

dataset composé de 500 grilles

Code

```
N, train_score, val_score = learning_curve(model,  
      X_train, Y_train, train_sizes = np.linspace(0.1,1.0,5), cv=5)  
print(N)  
plt.plot(N, train_score.mean(axis=1), label='train')  
plt.plot(N, val_score.mean(axis=1), label='validation')  
plt.xlabel('train_sizes')  
plt.legend()
```

dataset composé de 10000 grilles

- Pour 10000 grilles, les résultats sont meilleurs, on peut même remarquer que nous disposons de "trop de données"



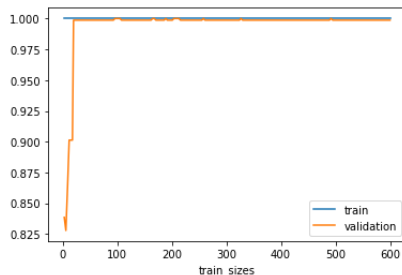
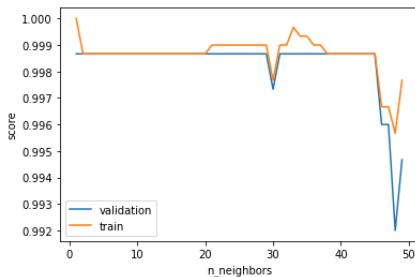
Sommaire

- 1 Introduction
 - Utilité
 - Principe de fonctionnement
- 2 Traitement d'images
 - Morphologie
 - Reconnaissance d'objet
- 3 Modélisation 2D
 - Propagation
 - Modélisation 2D
- 4 Intelligence artificielle
 - Influence de la taille du dataset
- 5 Modélisation 3D

- On adapte les programmes à la trois dimensions. On peut récupérer la taille totale de la tumeur et la distance maximale comme précédemment
- Les procédés de traitement d'images sont aussi adaptables à la 3D
- Par simplicité et efficacité, on réduit les dimensions des grilles et la propagation démarre désormais toujours du milieu
- De gauche à droite : bas, milieu, haut



- Les modèles d'intelligence artificielle fonctionnent alors exactement de la même façon qu'en deux dimensions
- On remarque que le modèle, dont la taille est moindre, est moins sujet à la variation des hyperparamètres
- Très peu de données sont nécessaires, on atteint rapidement un "palier de performance"



Bibliographie

- 1 IA et dépistage du cancer du sein : sensibilité accrue et temps de diagnostic raccourci, Bruno Benque avec RSNA, Thema-radiologie
- 2 Une IA de Google surpasse les radiologues pour détecter le cancer du sein, Coralie Lemke, Sciences et Avenir
- 3 FORMATION PYTHON MACHINE LEARNING, Machine Learnia, Youtube
- 4 Detection of Cancer in Lung with K-NN Classification Using Genetic Algorithm, P. Bhuvaneswari et A. Brintha Therese, Science direct
- 5 Jeu de la vie épidémique en Python, Hugues Meunier