



Table des matières



04 Résultats 05 Références





Introduction

La peau est l'organe le plus volumineux dont nous soyons dotés. C'est elle qui fait barrière entre nos organes internes et l'environnement extérieur, nous protégeant ainsi des blessures et des rayonnements solaires. Cependant, il arrive qu'elle tombe malade, notamment lorsqu'elle n'est pas bien protégée. Près de 80 000 nouveaux cas de cancer de la peau sont diagnostiqués chaque année en France.

Le cancer de la peau est défini par la croissance anormale des cellules de la peau et qui peut avoir différentes causes, notamment l'exposition au soleil. Cette maladie peut toucher différentes parties du corps humain, principalement les zones du visage, du cuir chevelu, des mains et des jambes.



3 grands types du cancer de la peau



Le carcinome basocellulaire

Représente 70% des cas



Le carcinome spinocellulaire

Représente environ 15% des cas



Le mélanome malin

Le moins fréquent mais le plus dangereux





Diagnostique

Examen clinique

Régle **ABCDE** qui est une méthode très efficace pour detecter la nature des lésions

Interrogation du patient

Chercher les facteurs de risques environnementaux et constitutionnels



La régle ABCDE

Une lésion cutanée est suspecte lorsqu'elle répond aux critères suivants, réunis dans la règle ABCDE :

- **A**: asymétrie,
- B: bords irréguliers,
- **C** : couleur hétérogène
- D : diamètre supérieur à 6mm
- **E** : extension en taille récente





Pronontic de survie

Le mélanome représente 1,1 % de l'ensemble des décès par cancer se classant ainsi au 16e rang des cancers les plus mortels. Il est à lui seul responsable de 90 % des décès liés aux cancers cutanés.

Le propostic dépend de la **précocité** du diagnostique et du **stade** de la maladie.





Outils de réalisation









Métriques de comparaison

Modèles

Matrice de confusion

Accuracy

Recall

Précision

Score F1

TSM12

INCEPTION V3

ResNet152V2

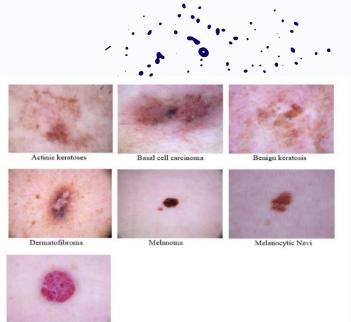
VGG16





Traitement des données

Selon la base de données, il existe :



Vascular Lesions

7 types de lésions cutanées

- Melanocytic Nevi [nv]
- Melanoma [mel]
- □ Benign Keratosis-like Lesions [bkl]
- □ Basal Cell Carcinoma [bcc]
- □ Actinic Keratoses [akiec]
- □ Vascular Lesions [vasc]
- □ Dermatofibroma [df]



Traitement des données

Selon la base de données, il existe :

4 manières de faire le diagnostique

- ☐ Histopathology
- □ Confocal
- □ Follow-up
- □ Consenus





Visualisation de données

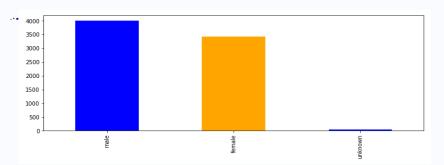


Figure qui montre la distribution des patients selon le sexe

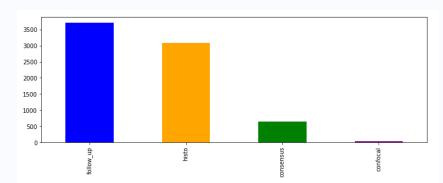


Figure qui montre la fréquence de chaque méthode de diagnostique

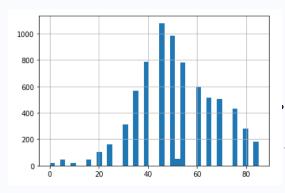


Figure qui montre le nombre de patients par âge



Visualisation de données

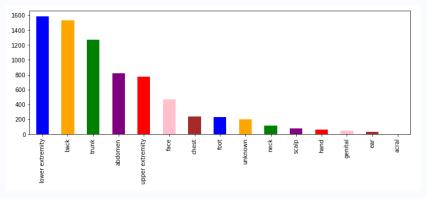


Figure qui montre les parties du corps humain concernée par ces lésions et la fréquence pour chaque partie

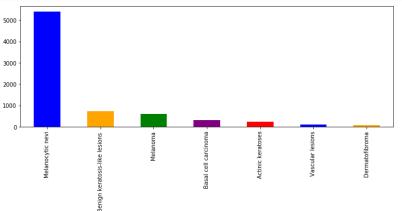


Figure qui montre la fréquence de chaque type de lésion



Visualisation de données

Conclusions par rapport à la visualisation de données :

- Les hommes sont plus touchés que les femmes par les lésions cutanées.
- ☐ La méthode de diagnostique follow-up est la plus utilisée suivie de l'histopathologie.
- ☐ La portion d'âge la plus touchée est entre 35 et 70.
- ☐ Le Melanocytic Nevi [nv] est le type de lésions le plus fréquent.





Prétraitement des données

- ✓ Création de deux dictionnaire, le premier contient les noms d'images, le deuxième contient les catégories de diagnostiques pour les lésions cutanées.
- ✓ Nettoyage de données : éliminer les duplications.
- ✓ Des **codes numériques** uniques ont été crées pour chaque catégorie de lésions, car les entiers sont plus faciles à gérer que les chaines de caractères.
- ✓ Pour assurer que l'apprentissage soit plus rapide les images ont été redimensionnées par un facteur de 1/4. Les dimensions des images sont devenues 150 * 112.5.
- Ensuite les images ont été stockées sous forme de liste d'images numériques.
- ✓ Convertir les étiquettes numériques en une matrice binaire (one hot matrix)





Division des données

- ✓ Nous divisons d'abord l'ensemble de données en 80 % de données de formation et 20 % de données de test. Nous prenons ensuite les 80 % de données d'entraînement et les divisons en 87,5 % de données d'entraînement et 12,5 % de données de validation, pour former nos 70 % d'entraînement - 10 % de validation - 20 % de test.
- ✓ Notez que la division sera appliquée à chaque classe individuellement pour s'assurer qu'il y a suffisamment d'échantillons de chaque classe dans chaque division pour une modélisation précise, ce qui signifie que chaque classe sera également divisée 70:10:20.

Méthodologie



La normalisation

Selon l'équation suivante :

$$x_n = \frac{x - u}{\sigma}$$



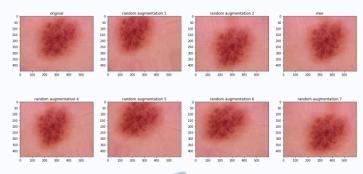
Méthodologie



Augmentation de données

pour éviter l'overfitting, et pour obtenir des résultats plus exactes et plus robustes, nous avons ajouté des données artificielles par :

- ☐ Une rotation aléatoire de 20%
- ☐ Une translation aléatoire horizontale de 20%
- ☐ Une translation aléatoire verticale de 20%
- ☐ Un shear aléatoire de 10%
- ☐ Un zoom aléatoire de 10%







L'apprentissage des modeles



Les hyperparamètres :

- Optimiseur (Adam) : est l'optimiseur le plus utilisé et le plus simple à implémenter. Très efficace et peut manipuler de larges données.
- Fonction de perte (Categorical Cross-Entropy) : est utilisé lorsqu'une seule catégorie est appliquée pour chaque point de données ce qui était très optimale pour notre projet.
- **Epochs**: De multiples essais avec les valeurs: 20, 50, 100, 150 et 200 ont été effectués pour trouver le meilleur résultat. 50 était la valeur la plus optimale.
- **Batch size** : e multiples essais avec les valeurs : 5, 10, 20 et 40 ont été effectués pour trouver le meilleur résultat.10 était la valeur la plus optimale.
- Learning rate (taux d'apprentissage): Décroissant qui commence à 0.001 pour les 2 premiers modèles et 0.0001 pour les autres.

Hyperparameter	Value
Optimizer	Adam
Loss Function	Categorical Cross-Entropy
Epochs	50
Batch Size	10
Learning Rate	0.001 - 0.00001 (Reduces on Plateau)



L'apprentissage des modeles



Les poids:

Des poids ont été ajoutés pour rendre les modèles plus sensibles à la catégorie des lésions cutanées des naevus mélanocytaires [nv] en raison de la déséquilibre des classes, puisqu'il représentait environ les deux tiers des données. Les pondérations étaient les facteurs par lesquels la valeur de la perte était multiplié en interne pour être utilisé dans l'algorithme de rétropropagation. L'équation suivante montre comment cela a fonctionné.

weighed loss class[i] = loss[i] _ class weights[i]

Skin Lesion Category	Weight
Actinic keratoses [akiec]	1.0
Basal cell carcinoma [bcc]	1.0
Benign keratosis-like lesions [bkl]	1.0
Melanoma [mel]	1.0
Melanocytic nevi [nv]	3.0
Vascular lesions [vasc]	1.0
Dermatofibroma [df]	1.0

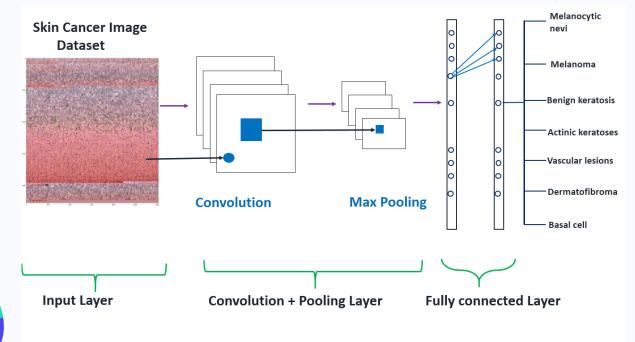


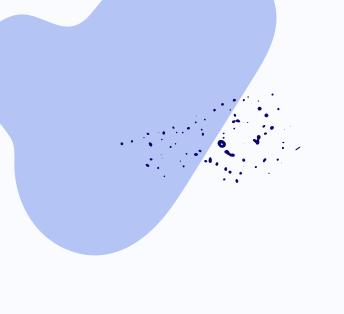


L'apprentissage des modeles



Exemple de l'architecture du modèle TSM12(12 couches) :

















Résultats



		TSM12	InceptionV3	Resnet152V2	VGG16
Exactitude(%)	train	80.04	77.07	99.43	94.56
	val	77.67	76.25	78.51	76.83
Loss	train	0.55	0.64	0.02	0.17
	val	0.78	0.76	1.35	1.41
RMSE	train	0.11	0.11	0.005	0.034
	val	0.12	0.12	0.11	0.11
MSE	train	0.038	0.044	0.001	0.01
	val	0.044	0.049	0.05	0.055

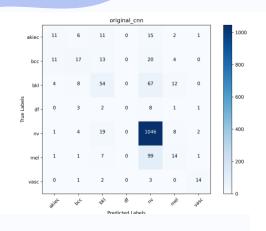
Métriques	TSM12	InceptionV3	ResNet152V2	VGG16
Exactitude(%)	77	75	78	77
Précision	72	68	76	74
Recall	77	75	78	77
Score F1	74	67	77	75

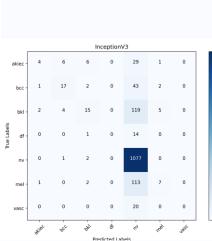
- Selon les résultats, il y a eu un overfitting pour les modèles VGG16 et ResNet152V2(high bias).
- Un low bias pour TSM12 et InceptionV3 (ni overfitting ni underfitting).
- On a remarqué que le modèle TSM12 était le plus rapide par contre ResNet152V2 était le plus lent, cela est du au fait que ResNet152V2 a plus de couches que TSM12(152 couches).

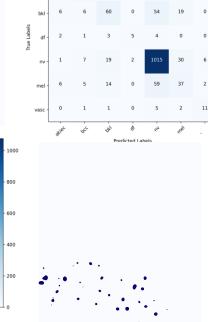


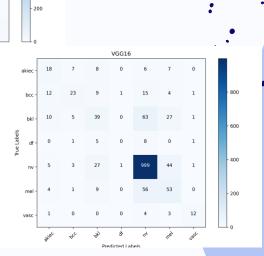
Résultats







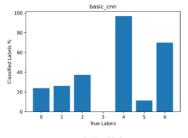


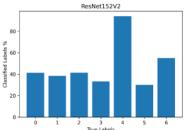


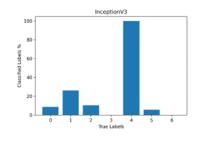


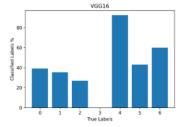
Résultats







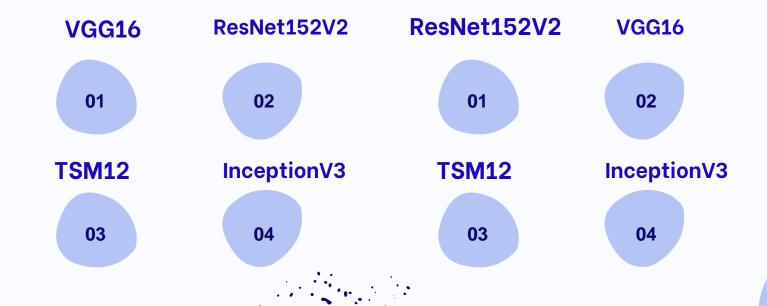




- Les 4 modèles ont une exactitude supérieure à 85% pour la classe Melanocytic Nevi [nv](Classe avec nombre maximum d'images).
- Bien au contraire, les 4 modèles ont des valeurs d'exactitude minimes pour la classe
 Dermatofibroma [df] (Classe avec nombre minimum d'images).
- On voit que le modèle InceptionV3 a eu des résultats inférieurs aux autres (le moins performant).
- On peut conclure que plus il plus d'images, plus les modèles sont plus performants.



Les résultats de l'article Mes résultats (approche naïve en ignorant l'overfitting)

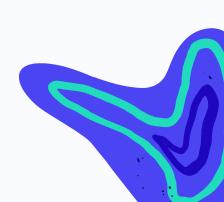


Conclusion

- Ce projet avait pour but de comparer les différentes architectures cnn dans la prédiction des lésions cutanées sur des images de lésions cutanées.
- La base de données a été prétraité par : le nettoyage de données-la division des données-la normalisation- l'augmentation et le redimensionnement de données.
- Nous avons construit une nouvelle architecture cnn appelé TSM12. Nous avons utilisé d'autres modèles existants : InceptionV3, ResNet152V2 et VGG16.
- Des hyperparamètres et des métriques ont été utilisés pour comparer entre les modèles.
- Selon l'article le modèle VGG16 était le plus performant, mais en considérants nos résultats et en ignorant l'overfitting, le modèle ResNet152V2 était plus performant que VGG16.
- Et enfin nous pouvons dire que plus les données sont plus larges pour une classe, plus l'apprentissage est meilleur.
- Il est conseillé d'augmenter les données pour la classe des lésions malignes pour une meilleur prédiction de ces tumeurs.

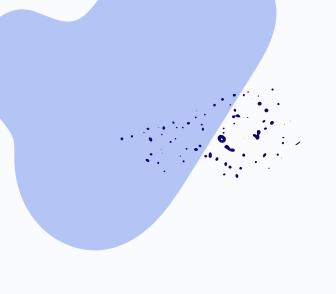








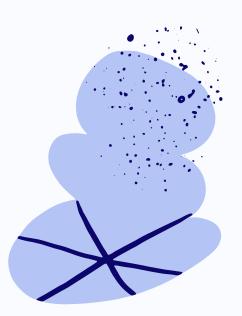
















References

- https://www.e-cancer.fr/Comprendre-prevenirdepister/Se-faire-depister/Depistage-des-cancersde-la-peau/Les-cancers-de-la-peau
- https://www.cancer-environnement.fr/340-Melanome-cutane.ce.aspx
- https://www.kaggle.com/kmader/skin-cancermnist-ham10000
- Skin Lesion Classification using Convolutional Neural Networks

Tareq Tayeh, Member, IEEE, Sulaiman Aburakhia, Member, IEEE, Moataz ElKhishen, Member, IEEE