



DataScientest BootCamp DataScience Sept24

Projet Radiographies Pulmonaires

Chris Hozé et Mickaël Melkowski

Mentor : Souhail Hadji

Contexte du projet

La COVID19

La COVID-19 est une pandémie mondiale qui a émergé en 2019 et affecte principalement les poumons.

Initialement, les outils principaux de diagnostic était la RT-PCR et la radiographie pulmonaire.

Difficulté de diagnostic

Les radiographies sont aisées à réaliser mais difficiles à expertiser.

Des universités d'Asie et de Moyen-Orient ont mis **en commun leur données de radiographies pour développer un outil de diagnostic automatisé.**

Objectif de l'étude

Développer un modèle permettant de classifier les radiographies pulmonaires en fonction de l'affection ou non du patient.

Faciliter le diagnostic.

Plan de la présentation

Exploration des données

- **Présentation des données**
 - Nombre total
 - Répartition par groupe
 - Images moyennes
- **Statistiques exploratoires**
 - Valeurs des pixels
 - Analyses en composantes principales (ACP)

Modèle de classification

- **Définition du modèle**
 - Modèle de base
 - Paramétrage
- **Modèle final**
 - Description
 - Performances

Généralisation du modèle de prédiction

- **Les données CXNET**
- **Génération de masques**
 - Segmentation
- **Prédiction de la positivité au COVID**
 - Classification

Exploration des données

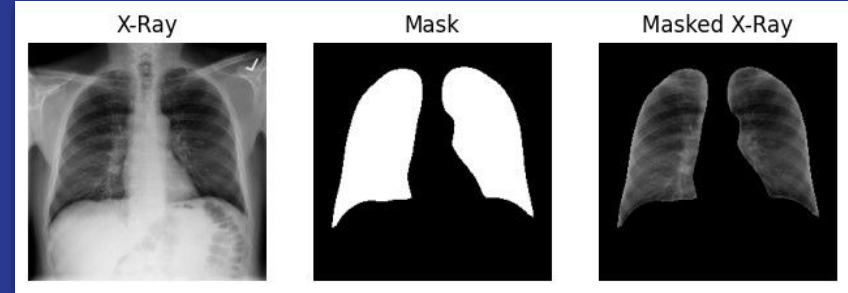
Les données

On dispose d'un jeu de données constitué de 21 165 images réparties en 4 catégories :

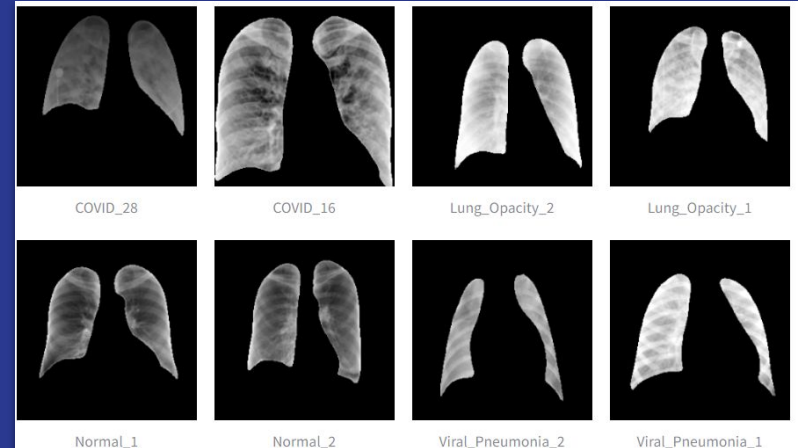
- COVID
- Opacité Pulmonaire
- Pneumonie Virale
- Normal

Chaque image est fournie avec son masque précalculé.

Le masque permet d'isoler l'information liée aux poumons



Les images sont au format 256 x 256 x 1

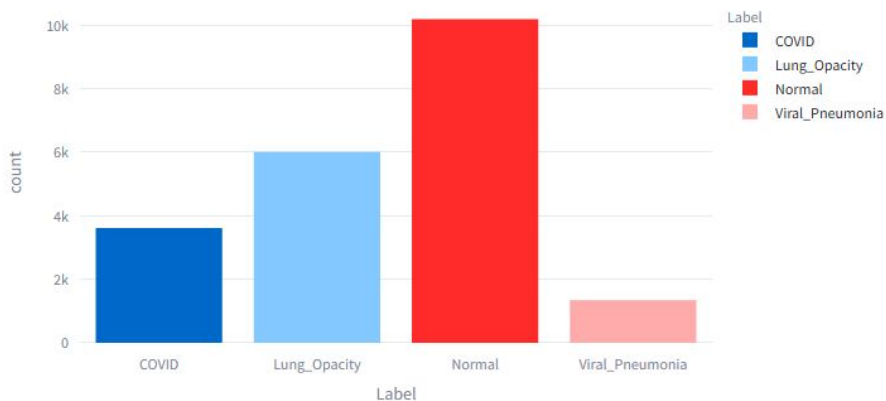


Nombre d'images par catégories et source

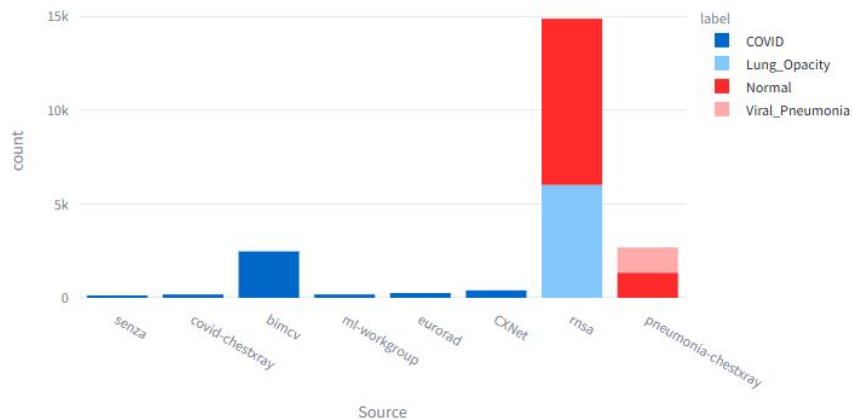
Le nombre d'images par source est très variable.
Les données COVID et Viral Pneumonia
sont minoritaires.

Les sources de données sont multiples
en particulier pour la catégorie COVID.

Nombre d'images par Label



Nombre d'images par Source



Analyse des images après masquage

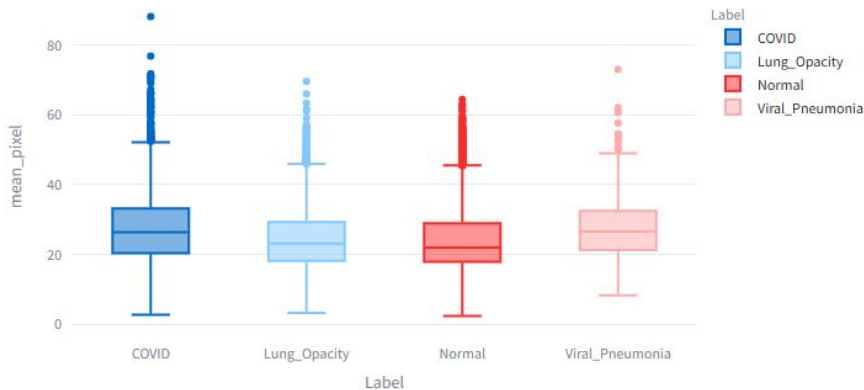
La valeur moyenne des pixels est plus élevée pour les catégories COVID et Viral Pneumonia.

Les données COVID semblent plus extrêmes.

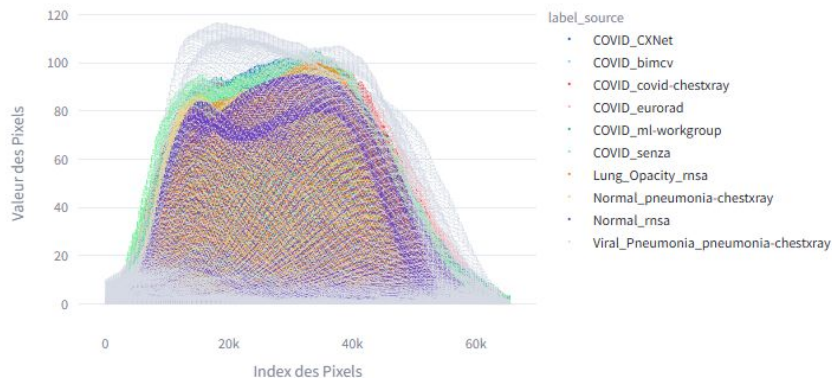
Selon les sources et les catégories, la valeur moyenne à un pixel donné varie fortement.

Ces variations pourraient s'expliquer par des différences de prise de vue des radiographies.

Distribution de la valeur moyenne des pixels par image et par Label



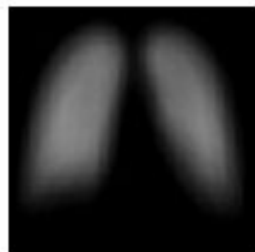
Valeurs des pixels pour Toutes les catégories



Images moyennes par catégorie et par source

Peu de différences sont visibles à l'oeil sur les poumons eux même.

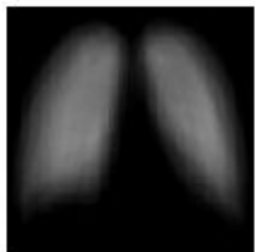
Pour la **pneumonie virale, le poumon gauche est moins visible** ce qui risque de biaiser la modélisation.



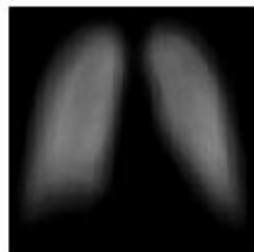
COVID - CXNET



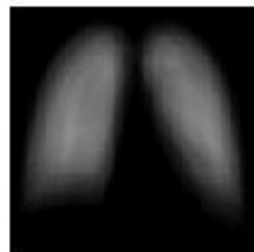
COVID - bimcv



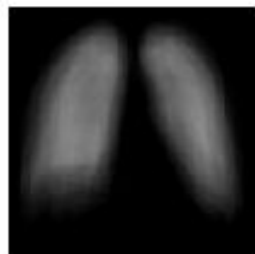
COVID - covid-chestxray



COVID - eurorad



COVID - mlworkgroup



COVID - senza



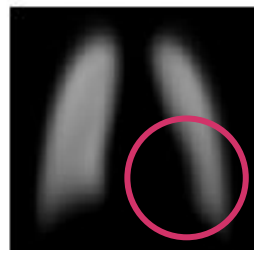
Lung Opacity - rnsa



Normal -
pneumonia- chestxray



Normal - rnsa

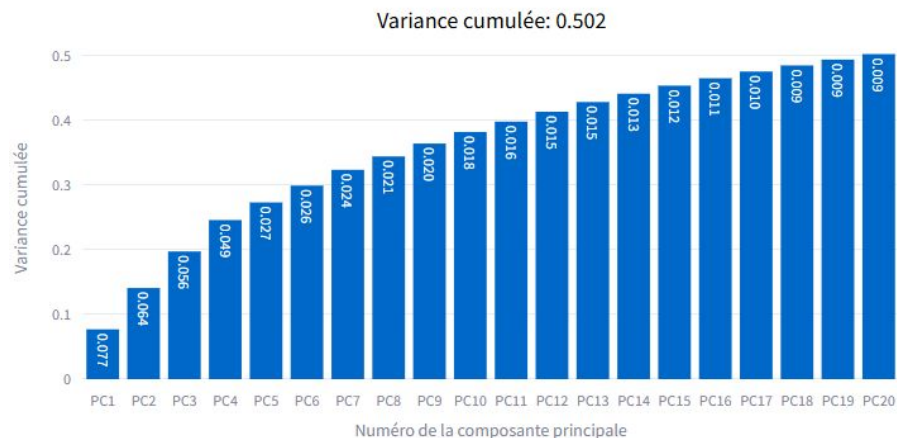


Viral Pneumonia
pneumonia- chestxray -

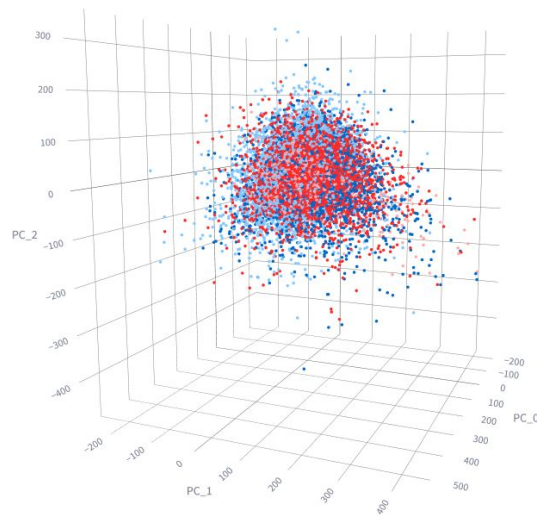
Analyse en composantes principales

Avec 20 composantes, on extrait 50% de la variance initiale du jeu de données.

Évolution de la variance expliquée par les composantes principales



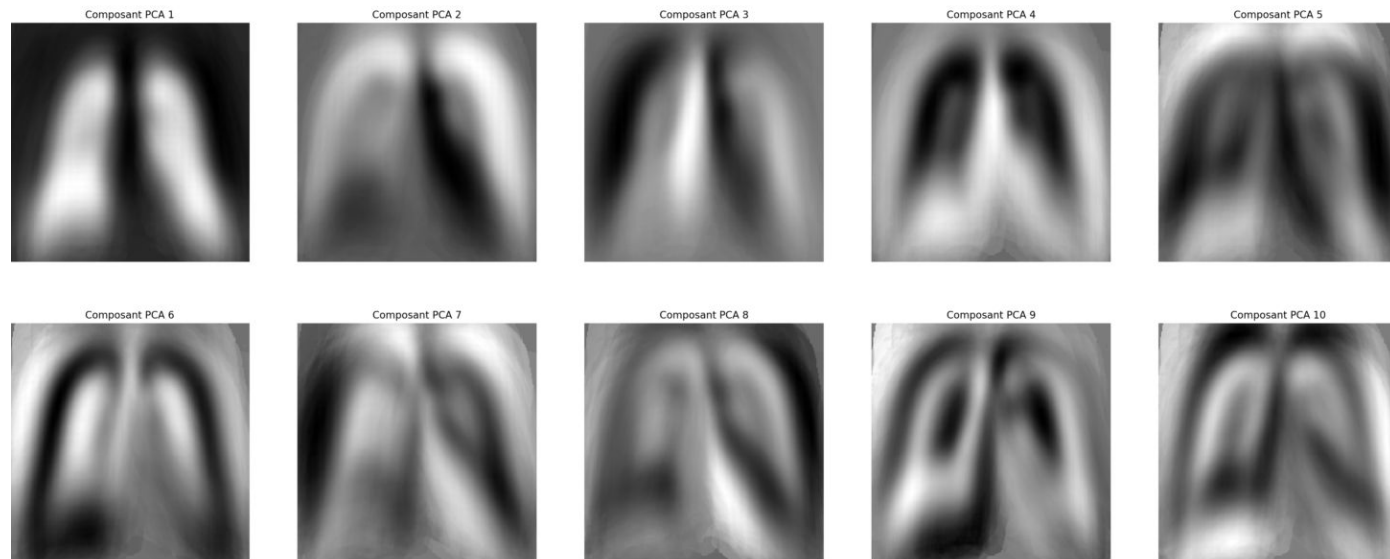
On ne peut pas distinguer de différences entre catégories sur les composantes de l'ACP.



Analyse en composantes principales

On remarque que les composantes mettent peu en évidence des différences fines de coloration.

Elles semblent principalement refléter des différences de prise de vue et de luminosité.



Les modèles de classification classiques ne seront donc pas performants sur ce jeu de données.

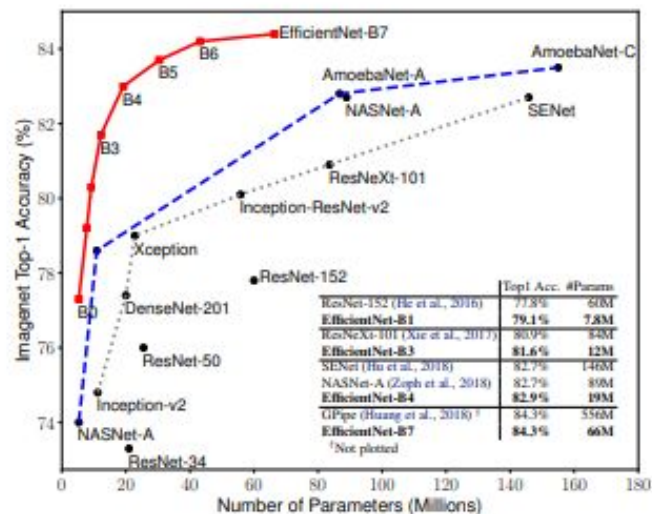
Il faudra se tourner vers des modèles de deep-learning et de computer vision.

Classification des images

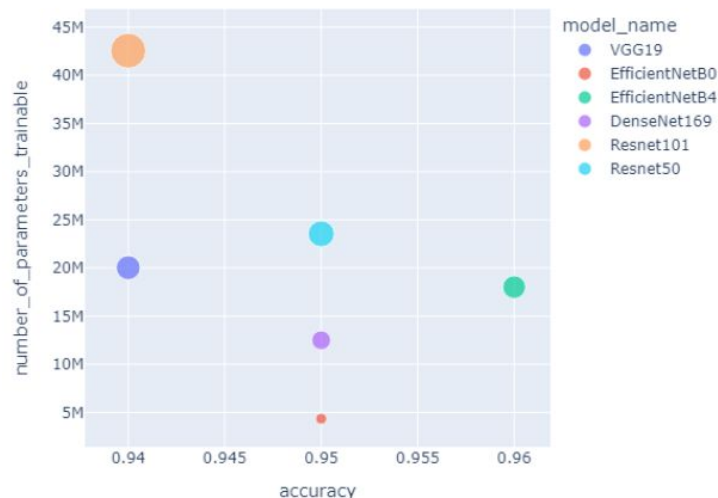
Choix du modèle de base

Précision et Nombre de paramètres
sur le jeu de données ImageNet

Tang et Le, 2019, <https://arxiv.org/abs/1905.11946>



Précision et Nombre de paramètres
sur notre jeu de données
(images non masquées)



On choisit de tester les modèles EfficientNetB0, EfficientNetB4 et ResNet50.

Paramétrage du modèle

Stratégies testées

Le jeu de données divisé en trois :

- 80% en training,
- 10% en test
- 10% en validation

On retient le modèle qui obtient la meilleure précision sur le jeu de données de test.

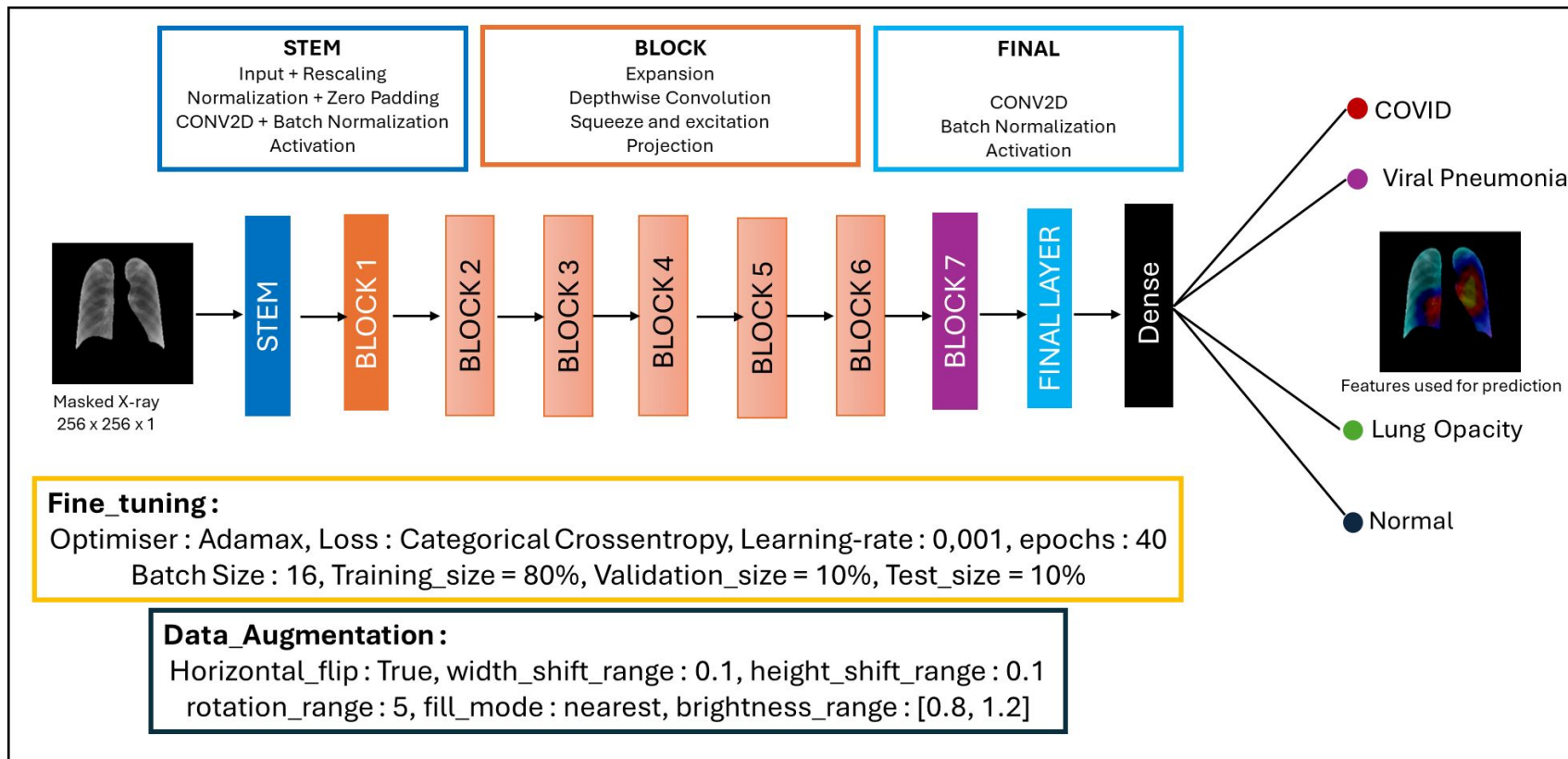
Modèle testé	EfficientNetB0	EfficientNetB4	ResNet50
Trainable layers	15 %	50%	100 %
Optimizer	SGD	Adam	Adamax
Learning-rate	0.01	0.001	0.0001
Decay	None	CosDecay	ExpDecay
Nb_epochs	10	20	40

Data Augmentation

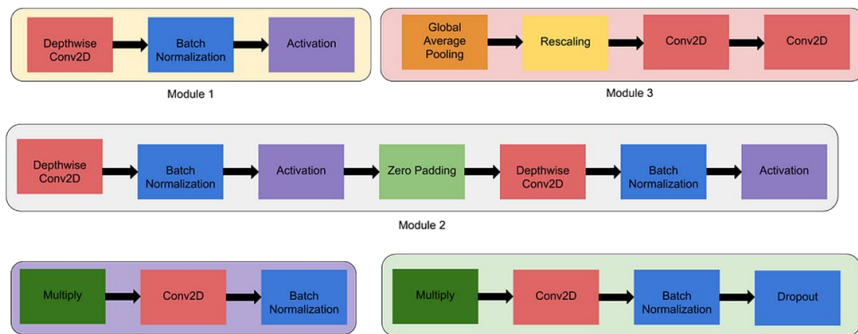
Augmenter la diversité du jeu de données en appliquant des transformations sur les images à chaque itération

Retournement horizontal	Horizontal_flip = True
Ajustement de luminosité	Brightness_range : [0.8, 1.2]
Rotation	rotation_range = 0.1
Décalage vertical	height_shift_range = 0.1
Décalage horizontal	width_shift_range = 0.1

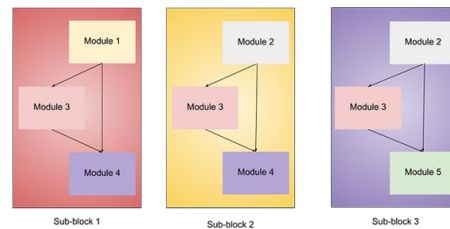
Modèle final : EfficientNetB4 avec fine-tuning



Détails de l'architecture Efficient Net B4

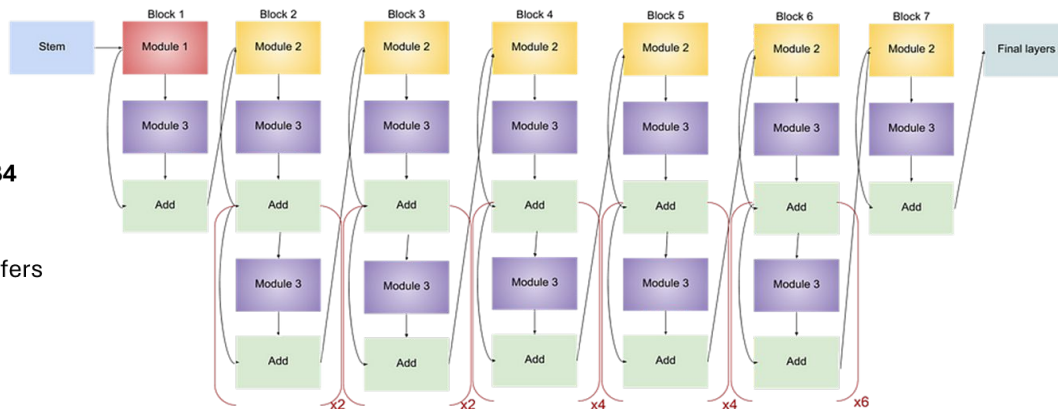


Sub-blocks



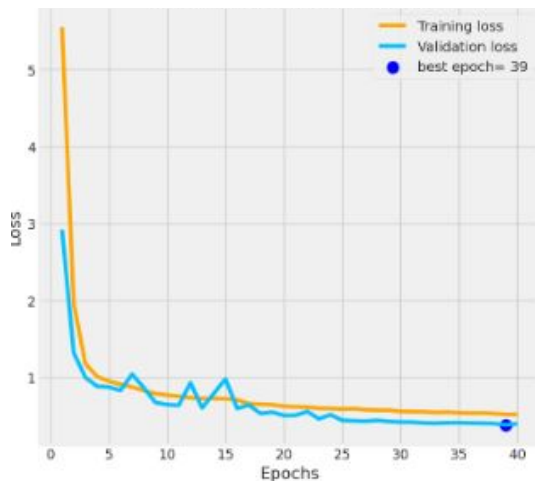
Efficient Net B4 architecture

Here, modules refers to sub-block

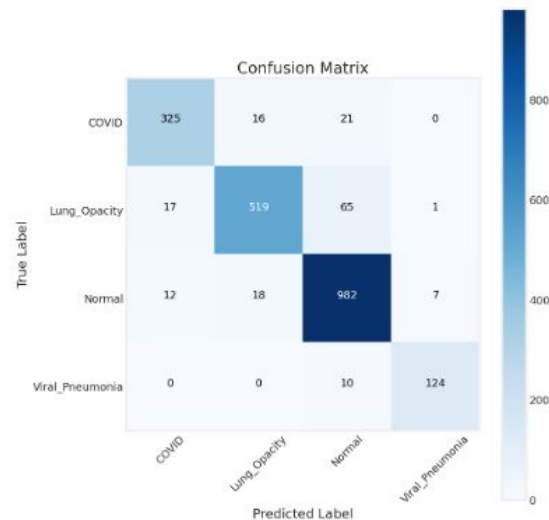


Performances du modèle

Evolution de la précision et de la fonction de perte
sur les jeux de données d'apprentissage et de validation



Matrice de confusion
sur le jeu de données de test



Le modèle permet d'atteindre une précision moyenne de 0.92 sur l'ensemble des catégories
Sur la catégorie COVID, la précision est de 0.93 et le recall de 0.90.

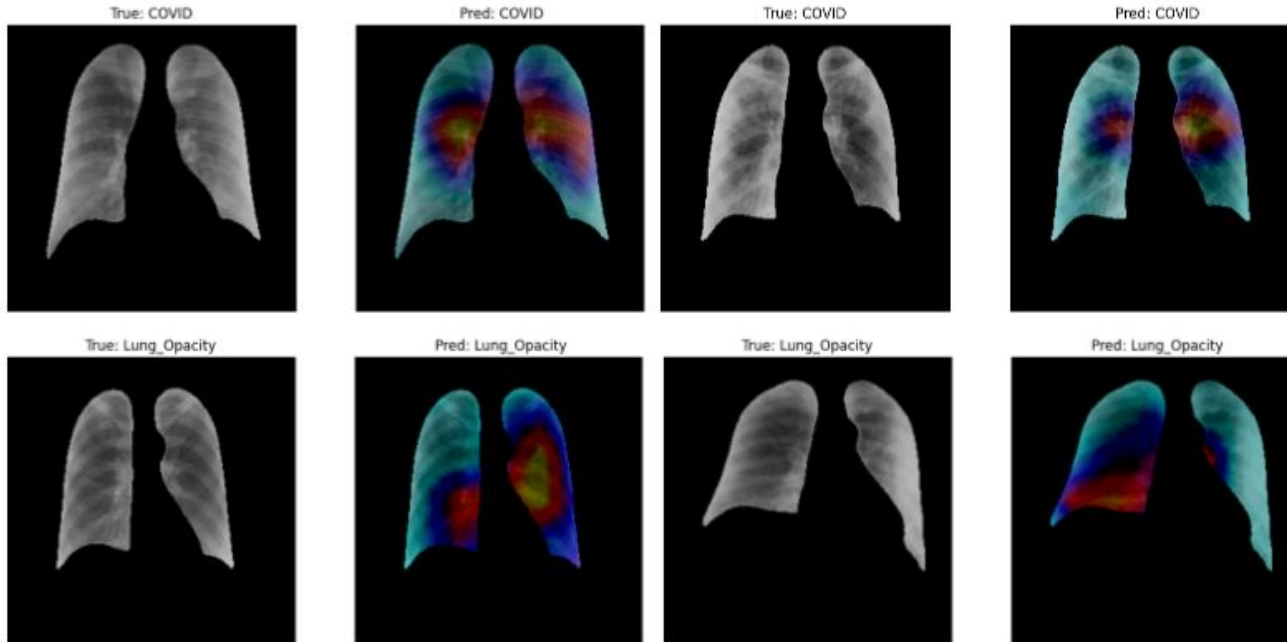
Performances du modèle

Rapport de classification

	Precision	Recall	F1-Score	Support
COVID	0.93	0.9000	0.91	362
Lung_Opacity	0.93	0.8400	0.88	602
Normal	0.9	0.9700	0.94	1019
Viral_Pneumonia	0.98	0.9400	0.96	134
accuracy		0.9200		2117
macro avg	0.94	0.9100	0.92	2117
weighted avg	0.92	0.9200	0.92	2117

Application sur le jeu de données de test :

Prédiction et visualisation des zones du poumon atteintes



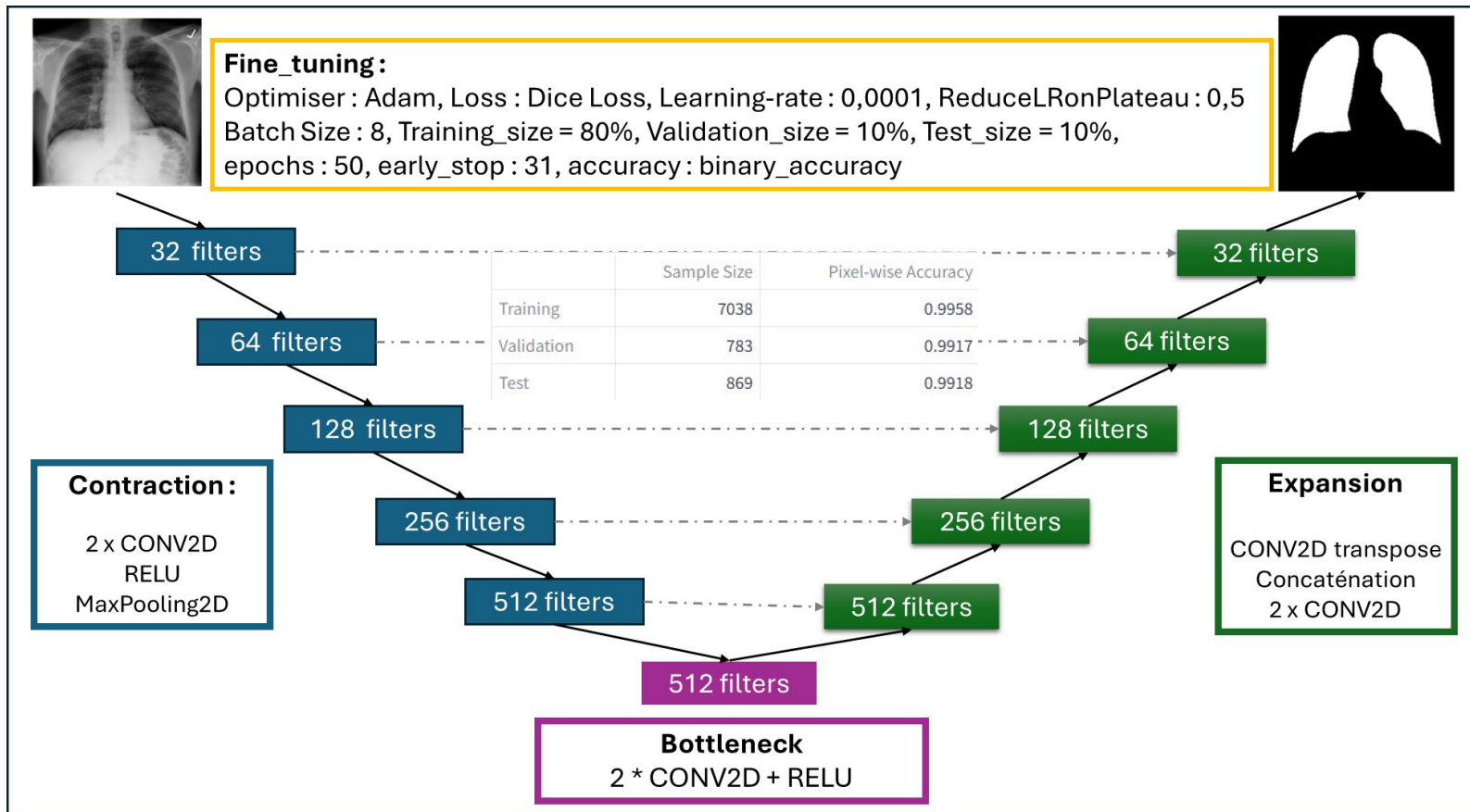
Généralisation du modèle

Généralisation du modèle

Nous cherchons à savoir quelle serait l'efficacité du modèle sur des données externes.

- Utilisation du dataset COVID-NET :
 - **+ de 80 000 images** de radiographies pulmonaires non masquées
 - Regroupées en deux catégories **“Positif” et “Négatif” pour le COVID.**
 - Sélection de 638 images non redondantes avec les données initiales
- Développement d'un **modèle de segmentation pour la génération de masques**
 - A partir des images et masques de notre jeu de données initial
 - **Architecture Unet** couramment utilisé en imagerie médicale

Modèle Unet pour la génération de masques

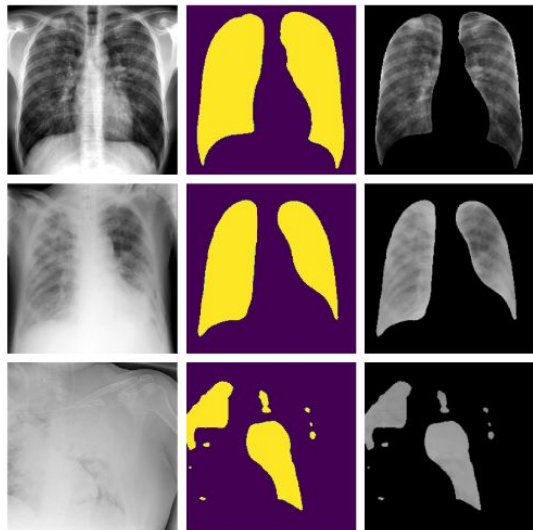


Applications des modèles sur CXNET

Segmentation pour la génération de masque

Une précision de 99,1% sur le dataset de validation

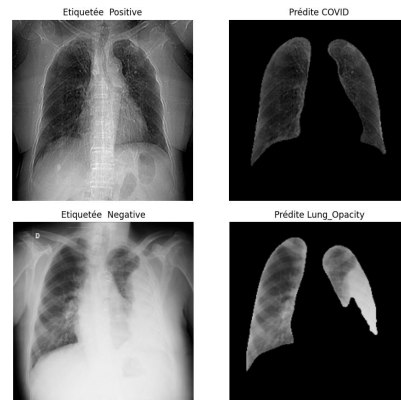
Une bonne efficacité, sauf cas particulier.



Classification

	COVID	Lung_Opacity	Normal	Viral_Pneumonia
Negatif	62,7 %	19,7 %	17,6 %	0,0 %
Positif	83,1 %	11,3 %	5,3 %	0,3 %

Il semble plus difficile de généraliser notre modèle de classification. **Une plus forte proportion de cas positif sont classés comme COVID.** Il faudrait s'assurer que les patients négatifs au COVID sont bien exempt d'affection.



Les résultats de la généralisation sur le jeu de données COVID-NET sont mitigés.

Pour conclure, il faudrait disposer de données plus formatées et classées comme dans le dataset initial.

Conclusion

21 165 images réparties en
COVID, Normal, Pneumonie, Opacité
Pas de classification évidente

Une précision moyenne de 92%
COVID : précision de 0.93, recall de 0.90
Conforme à l'état de l'art

Application Streamlit pour
la prédiction, le masquage
et la visualisation.



**Modèle de deep-learning
EfficientNetB4** avec
paramétrage fin et
augmentation de données

Génération de masques avec un
modèle de segmentation Unet.
Application aux données **COVID-NET**
(80 000 images, Positif / Négatif)

Retours d'expériences

Contraintes organisationnelles

Manque d'organisation et de synchronisation au démarrage du projet.

Choix de travailler en parallèle plutôt qu'en répartition des tâches sur le déploiement de modèles.

Contraintes techniques

Limitations en ressources informatiques liées aux tailles de nos GPU

Incompatibilités Windows vs WSL et versions de Tensorflow

Prise en main de Pytorch

Déploiement du Streamlit

Perspectives

Améliorer la précision en utilisant d'autres modèles en particulier Transformers

Prédire la sévérité du COVID à partir d'un jeu de données disposant d'annotations plus complexes.

Détecter plus finement les zones du poumon atteintes.

Place à la démo !

Application de classification de Radiographie Pulmonaire

Cette application permet la prédiction de l'état d'un patient à partir d'une radiographie pulmonaire pour les affections suivantes : Covid, pneumonie virale ou opacité pulmonaire.

La prédiction s'effectue sur des images brutes ou après isolation du poumon par masquage. Trois fonctionnalités sont disponibles :

- La prédiction des affections
- La génération de masque pour isoler les poumons
- La visualisation des zones les plus informatives pour la prédiction

Le fichier à importer doit être une image au format "png", "jpg".
Il est possible de prédire un ensemble de fichiers avec un dossier au format "zip".

Fichier ou dossier à prédire:

 Drag and drop file here
Limit 200MB per file • PNG, JPG, JPEG, ZIP, X-ZIP-COMPRESSED

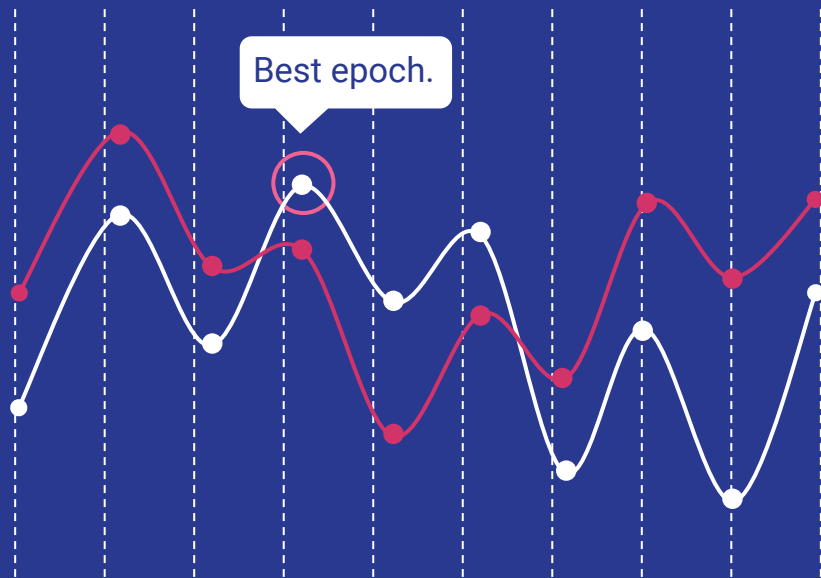
Browse files

Sinon choisir un fichier d'exemple

Aucun

Scripts, notebooks et modèles :
github.com/mmelkowski/radio_pulmo/

Application Streamlit :
radio-pulmo.streamlit.app/



Application de classification de Radiographie Pulmonaire

Cette application permet la prédiction de l'état d'un patient à partir d'une radiographie pulmonaire pour les affections suivantes : Covid, pneumonie virale ou opacité pulmonaire. [?]

La prédiction s'effectue sur des images brutes ou après isolation du poumon par masquage.

Trois fonctionnalités sont disponibles :

- La prédiction des affections
- La génération de masque pour isoler les poumons
- La visualisation des zones les plus informatives pour la prédiction

Le fichier à importer doit être une image au format "png", "jpg".

Il est possible de prédire un ensemble de fichiers avec un dossier au format "zip".

Fichier ou dossier à prédire:



Drag and drop file here

Limit 200MB per file • PNG, JPG, JPEG, ZIP, X-ZIP-COMPRESSED

Browse files

Sinon choisir un fichier d'exemple

masked_COVID-3.png

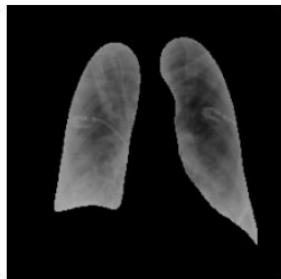


Image chargée après redimensionnement

Voulez vous prédire, masquer ou visualiser (Grad-CAM) l'image ?

Visualiser

Pour la visualisation, l'image doit être masquée.

Vous pouvez visualiser les pixels les plus importantes au début, au milieu et à la fin de prédiction.

Choix de la couche à visualiser (Première : stem_conv, Intermédiaire : block4f_expand_conv, Finale : top_conv):

top_conv

Démarrer la visualisation

✓ Visualisation en cours...

🕒 Chargement du model de prédiction...

🖌️ Création de l'overlay...

Voulez vous prédire, masquer ou visualiser (Grad-CAM) l'image ?

Prédire

L'image est-elle masquée ? (Les poumons sont isolés, on ne voit ni l'arrière-plan ni les autres organes) [?]

Oui



Démarrer la prédiction

✓ Prédiction en cours...

🕒 Chargement du model de prediction...

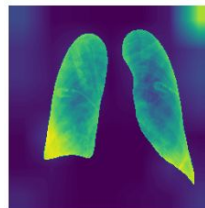
🤖 Prédiction...

👍 Interpretation...

Prédiction effectuée

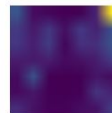
L'image est classée comme: COVID

Visualisation effectuée



Grad-CAM Applied

Download Grad-CAM



Heatmap generated

Download heatmap

Application de classification de Radiographie Pulmonaire

Cette application permet la prédiction de l'état d'un patient à partir d'une radiographie pulmonaire pour les affections suivantes : Covid, pneumonie virale ou opacité pulmonaire. [?]

La prédiction s'effectue sur des images brutes ou après isolation du poumon par masquage.

Trois fonctionnalités sont disponibles :

- La prédiction des affections
- La génération de masque pour isoler les poumons
- La visualisation des zones les plus informatives pour la prédiction

Le fichier à importer doit être une image au format "png", "jpg".

Il est possible de prédire un ensemble de fichiers avec un dossier au format "zip".

Fichier ou dossier à prédire:



Drag and drop file here

Limit 200MB per file • PNG, JPG, JPEG, ZIP, X-ZIP-COMPRESSED

Browse files

Sinon choisir un fichier d'exemple

raw_Normal-36.png



Image chargée après redimensionnement

Voulez vous prédire, masquer ou visualiser (Grad-CAM) l'image ?

Prédire


L'image est-elle masquée ? (Les poumons sont isolés, on ne voit ni l'arrière-plan ni les autres organes) [?]

Non




Démarrer la prédiction

✓ Prédiction en cours...

 Masquage à faire...

 Chargement du model de masquage...

 Calcul du Masque...

 Masquage...

 Chargement du model de prediction...

 Prédiction...

 Interpretation...

Prédiction effectuée

L'image est classée comme: Normal

Application de classification de Radiographie Pulmonaire

Cette application permet la prédiction de l'état d'un patient à partir d'une radiographie pulmonaire pour les affections suivantes : Covid, pneumonie virale ou opacité pulmonaire. ?

La prédiction s'effectue sur des images brutes ou après isolation du poumon par masquage.

Trois fonctionnalités sont disponibles :

- La prédiction des affections
- La génération de masque pour isoler les poumons
- La visualisation des zones les plus informatives pour la prédiction

Le fichier à importer doit être une image au format "png", "jpg".

Il est possible de prédire un ensemble de fichiers avec un dossier au format "zip".

Fichier ou dossier à prédire:



Drag and drop file here

Limit 200MB per file • PNG, JPG, JPEG, ZIP, X-ZIP-COMPRESSED

Browse files

Sinon choisir un fichier d'exemple

raw_Viral Pneumonia-9.png



Image chargée après redimensionnement

✓ Prédiction en cours...

🛑 Masquage à faire...

⌚ Chargement du model de masquage...

✏ Calcul du Masque...

✂ Masquage...

⌚ Chargement du model de prediction...

🤖 Prédiction...

👉 Interpretation...

Prédiction effectuée

L'image est classée comme: Viral Pneumonia

✓ Masquage en cours...

⌚ Chargement du model de masquage...

✏ Calcul du Masque...

✂ Masquage...

Masquage effectué




Image masquée



Masque

Application de classification de Radiographie Pulmonaire

Cette application permet la prédiction de l'état d'un patient à partir d'une radiographie pulmonaire pour les affections suivantes : Covid, pneumonie virale ou opacité pulmonaire. 

La prédiction s'effectue sur des images brutes ou après isolation du poumon par masquage.


Trois fonctionnalités sont disponibles :

- La prédiction des affections
- La génération de masque pour isoler les poumons
- La visualisation des zones les plus informatives pour la prédiction


Le fichier à importer doit être une image au format "png", "jpg".

Il est possible de prédire un ensemble de fichiers avec un dossier au format "zip".

Fichier ou dossier à prédire:

 Drag and drop file here
Limit 200MB per file • PNG, JPG, JPEG, ZIP, X-ZIP-COMPRESSED

Browse files

 Test_Covid.zip 1.9MB

×

Sinon choisir un fichier d'exemple


Aucun

▼

Voulez vous prédire, masquer ou visualiser (Grad-CAM) les images ?

Prédire

▼

Est-ce que les images sont masquées ? (Les poumons sont isolés, on ne voit pas l'arrière-plan et les autres organes) 


Non


▼

 Démarrer la prédiction

✓ Prédiction en cours...



 Chargement du modèle de prédiction...


 Chargement du modèle de segmentation...

 Prédiction fini.

Prédictions effectuées

Le tableau montre pour chaque fichier son résultat le plus probable et ensuite la probabilité par catégorie.

	filename	Prediction_results	pred
0	0a2c130c-c536-4651-836d-95d07e9a89cf.png	COVID	
1	0ab261f9-4eb5-42ab-a9a5-e918904d6356.png	Lung_Opacity	
2	0bc7f47a-ecea-4b66-811e-096004721a63.png	COVID	
3	0ed30a38-fd74-498f-aeaa-6feb7eb66be.png	COVID	
4	A034518-12-31-1900-NA-CHEST_AP_PORT-01141-2.000000-AP-84624-1-1.jpg	COVID	
5	A042028-01-18-1901-NA-CHEST_AP_VIEWONLY-74543-1.000000-AP-99716-1-1.jpg	COVID	
6	A055532-01-02-1901-NA-CHEST_AP_PORT-69637-1.000000-AP-93369-1-1.jpg	Lung_Opacity	
7	A099771-12-29-1900-NA-CHEST_AP_PORT-76117-1.000000-AP-94973-1-1.jpg	Lung_Opacity	

Download results as csv 

Application de classification de Radiographie Pulmonaire

Cette application permet la prédiction de l'état d'un patient à partir d'une radiographie pulmonaire pour les affections suivantes : Covid, pneumonie virale ou opacité pulmonaire. ?

La prédiction s'effectue sur des images brutes ou après isolation du poumon par masquage.

Trois fonctionnalités sont disponibles :

- La prédiction des affections
- La génération de masque pour isoler les poumons
- La visualisation des zones les plus informatives pour la prédiction

Le fichier à importer doit être une image au format "png", "jpg".

Il est possible de prédire un ensemble de fichiers avec un dossier au format "zip".

Fichier ou dossier à prédire:



Drag and drop file here

Limit 200MB per file • PNG, JPG, JPEG, ZIP, X-ZIP-COMPRESSED

Browse files



Test_Covid.zip 1.9MB



Sinon choisir un fichier d'exemple

Aucun



Voulez vous prédire, masquer ou visualiser (Grad-CAM) les images ?

Masquer



Démarrer le masquage

✓ Masquage en cours...

🔄 Chargement du model de masquage...

🟢 Masquage fini.

Masquage effectué

Download masked images as

zip 📁

Dossier initial



0a2c130c-c536-4
651-836d-95d07
e9a89cf



0ab261f9-4eb5-4
2ab-a9a5-e9189
04d6356



0bc7f47a-ecea-4
b66-811e-09600
4721a63



0ed30a38-fd74-4
98f-aeaa-6feb77e
b66be



A034518-12-31-1
900-NA-CHEST_
AP_PORT-01141-
2.000000-AP-84...



A042028-01-18-1
901-NA-CHEST_
AP_VIEWONLY-7
4543-1.000000-...



A055532-01-02-1
901-NA-CHEST_
AP_PORT-69637-
1.000000-AP-93...



A099771-12-29-1
900-NA-CHEST_
AP_PORT-76117-
1.000000-AP-94...

Dossier contenant la prédiction des masques



mask_0a2c130c-c5
36-4651-836d-95d
07e9a89cf.png



mask_0ab261f9-4e
b5-42ab-a9a5-e91
8904d6356.png



mask_0bc7f47a-ec
ea-4b66-811e-096
004721a63.png



mask_0ed30a38-fd
74-498f-aeaa-6feb
77eb66be.png



mask_A034518-12
-31-1900-NA-CHE
ST_AP_VIEWONLY
1-2.000000-AP-8...



mask_A042028-01
-18-1901-NA-CHE
ST_AP_VIEWONLY
7-74543-1.000000...



mask_A055532-01
-02-1901-NA-CHE
ST_AP_PORT-6963
7-1.000000-AP-9...



mask_A099771-12
-29-1900-NA-CHE
ST_AP_PORT-7611
7-1.000000-AP-9...

Dossier contenant les images masquées



masked_0a2c130c-c536
-4651-836d-95d07e9a8
9cf.png



masked_0ab261f9-4eb5
-42ab-a9a5-e918904d6
356.png



masked_0bc7f47a-ecea
-4b66-811e-096004721
a63.png



masked_0ed30a38-fd74
-498f-aeaa-6feb77eb66
be.png



masked_A034518-12-3
1-1900-NA-CHEST_AP_
PORT-01141-2.000000-
AP-84624-1-1.jpg



masked_A042028-01-1
8-1901-NA-CHEST_AP_
VIEWONLY-74543-1.00
0000-AP-99716-1-1.jpg



masked_A055532-01-0
2-1901-NA-CHEST_AP_
PORT-69637-1.000000-
AP-93369-1-1.jpg



masked_A099771-12-2
9-1900-NA-CHEST_AP_
PORT-76117-1.000000-
AP-94973-1-1.jpg