

# Rapport de modélisation du projet: **Radiographie Pulmonaires**

Bootcamp DS Sept24 : Chris HOZE, Antoine LIONET, Mickael MELKOWSKI



# Table des matières:

<b>Table des matières:</b>	<b>0</b>
<b>Premiers modèles de classification</b>	<b>2</b>
Forêts aléatoires	2
Premier modèle de deep-learning : LeNet via Tensorflow	3
Accuracy et Loss	3
Matrice de confusion	3
Second modèle de deep-learning : Dense121 via Pytorch	4
Accuracy et Loss	4
Matrice de confusion	4
<b>Choix du modèle de base</b>	<b>5</b>
Modèle Resnet50 (177 couches)	5
Accuracy et loss	5
Matrice de confusion	5
Modèle EfficientNetB2 (342 couches)	6
Conclusion	6
<b>Augmentation de données, choix du learning rate</b>	<b>7</b>
<b>Fine tuning : Dégel des dernières couches</b>	<b>8</b>
Conclusion	9
<b>Ré-estimation de l'ensemble des paramètres</b>	<b>9</b>
Modèle VGG19 (16 couches)	9
Modèle EfficientNetB0 (237 couches)	10
Modèle EfficientNetB4 (381 couches)	11
Modèle DenseNet169 (x couches)	12
Modèle Resnet101 (x couches)	13
Conclusion	15

# Premiers modèles de classification

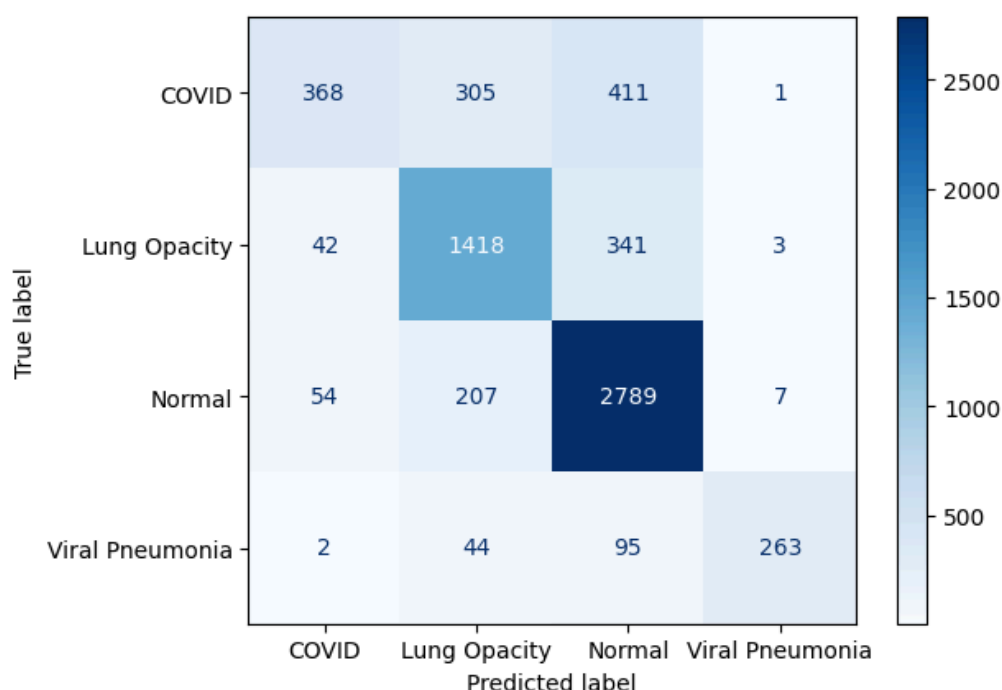
## Forêts aléatoires

En première approche, on applique une classification Random Forest sur les données transformées par l'ACP. L'ACP a été réalisée sur jeu de données complet après réduction du dataset en image de taille (100,100) soit 1000 variables correspondant aux pixels plus une correspondant à la source des données. Les données ont été séparées entre apprentissage (80%) et validation (20%) avant la réalisation de l'ACP. Le jeu de données transformé pour garder 95% de la variance contient 782 variables.

Une optimisation des paramètres est effectuée par validation croisée, On teste des valeurs entre 100 et 700 pour le nombre d'estimateurs, on cherche à optimiser le f1\_macro car le jeu de données est déséquilibré. Le meilleur modèle est celui avec un n\_estimators à 200.

Les résultats sont donnés dans le tableau de classification et la matrice de confusion suivante. On observe que malgré une précision moyennement élevée (0.76), le recall pour la classe COVID est particulièrement bas (0.34). Le modèle ne semble donc pas adapté pour cette analyse, ce qui est cohérent avec les résultats d'exploration.

	precision	recall	f1-score	support
COVID	0.79	0.34	0.47	1085
Lung_Opacity	0.72	0.79	0.75	1804
Normal	0.77	0.91	0.83	3057
Viral Pneumonia	0.96	0.65	0.78	404
accuracy			0.76	6350
macro avg	0.81	0.67	0.71	6350
weighted avg	0.77	0.76	0.74	6350

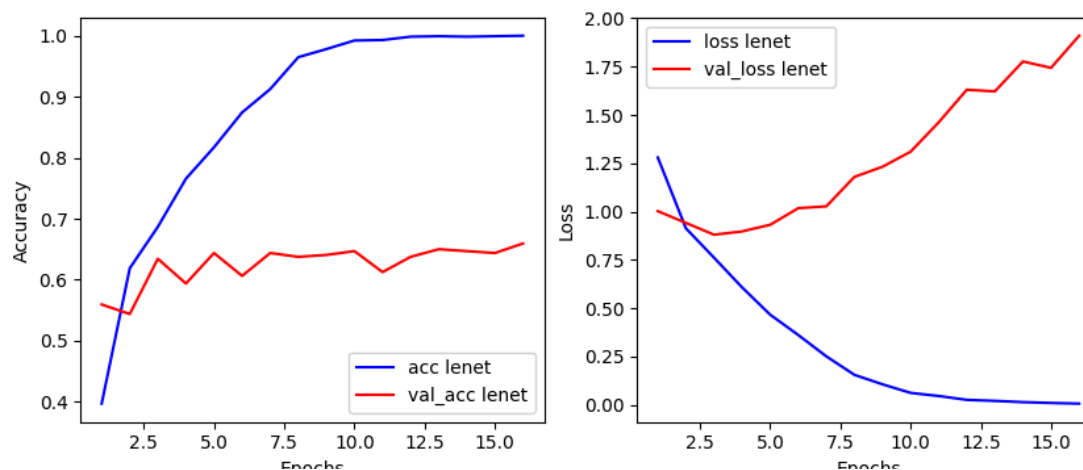


## Premier modèle de deep-learning : LeNet via Tensorflow

Des premiers modèles de deep-learning ont été réalisés avec un jeu de données réduit composés de 2000 images (500 par type) redimensionnées en 256\*256 pixels. On teste tout d'abord un modèle de LeNet avec Keras et TensorFlow.

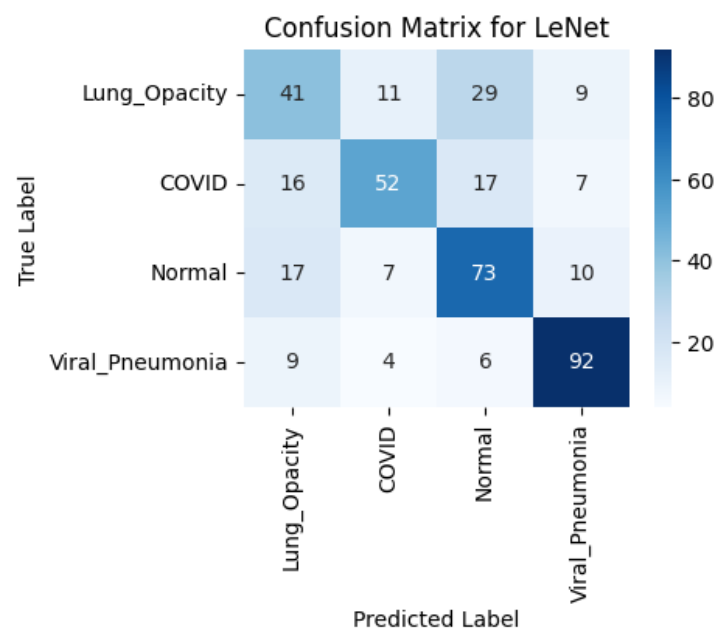
Le test a été réalisé avec 16 epochs et une taille de batch de 32

### Accuracy et Loss



Il y a ici un overfitting sur les données, l'accuracy sur le jeu de validation ne s'améliorant que peu et la loss augmentant aussi.

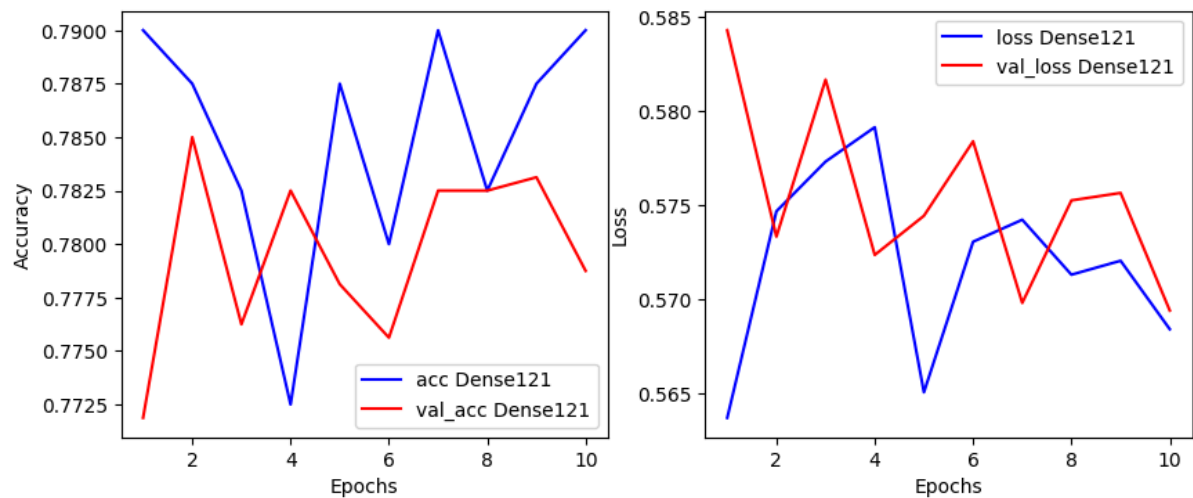
### Matrice de confusion



## Second modèle de deep-learning : Dense121 via Pytorch

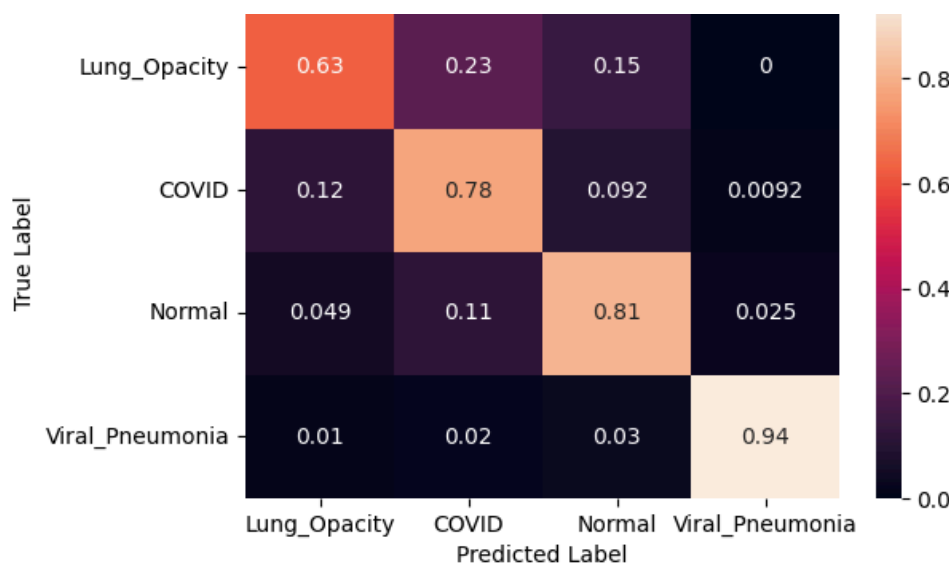
Le jeu de données et le paramétrage est identique à celui du modèle précédent.

### Accuracy et Loss



La précision de ce modèle semble légèrement meilleure même si l'on est encore loin d'un modèle abouti.

### Matrice de confusion



Ces premiers modèles présentent des résultats encourageants sur la pertinence de l'utilisation du deep learning avec en moyenne environ 80% de classification correctes pour Dense-121.

L'amélioration ou le choix de modèles plus adaptés pourra nous permettre d'assurer de meilleurs résultats pour les prochaines phases du projet.

## Choix du modèle de base

Plusieurs réseaux de neurones de convolution (CNN) ont été testés. Les modèles définis par nos soins n'ayant pas des performances suffisantes nous nous sommes tournées vers des modèles plus complexes disponibles directement via keras ou d'autres sources en ligne.

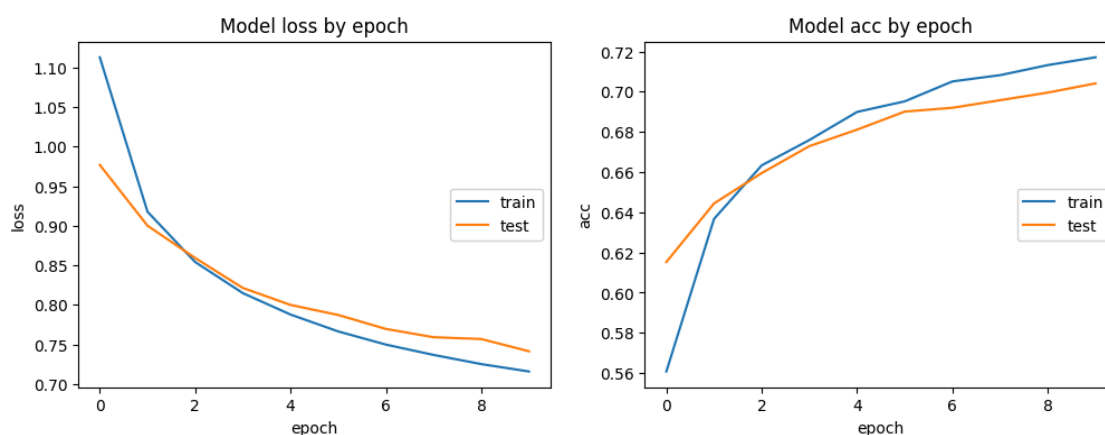
Les modèles choisis étaient pré-entraînés sur le jeu de données imagenet. Le choix du modèle de base a été effectué en utilisant les poids associés à leur entraînement via image\_net. Les masques ont été appliqués en amont et les images ont été pré-traitées avec les fonctions ad hoc fournies avec le modèle.

Plusieurs modèles ont d'abord été testés, le modèle Resnet, EfficientB2,. On choisit le meilleur modèle à partir des résultats du modèle brut avec un learning rate de 0.001 et 10 epochs et une taille de batch de 64.

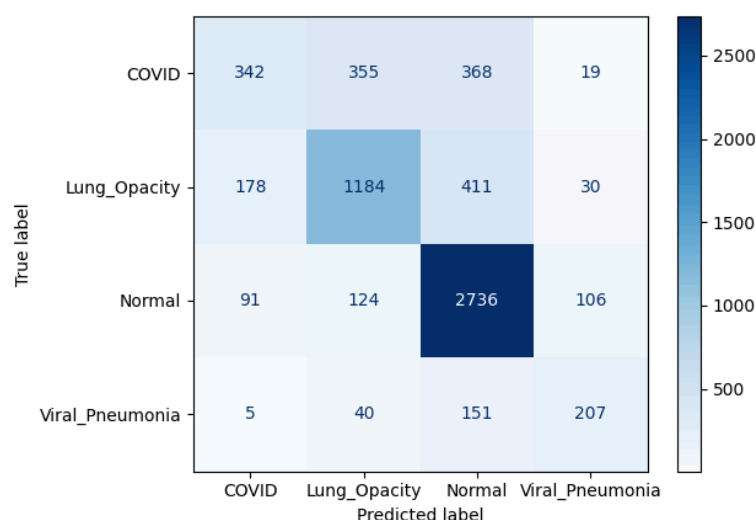
70% des données sont utilisées en apprentissage et 30% en validation.

### Modèle Resnet50 (177 couches)

#### Accuracy et loss

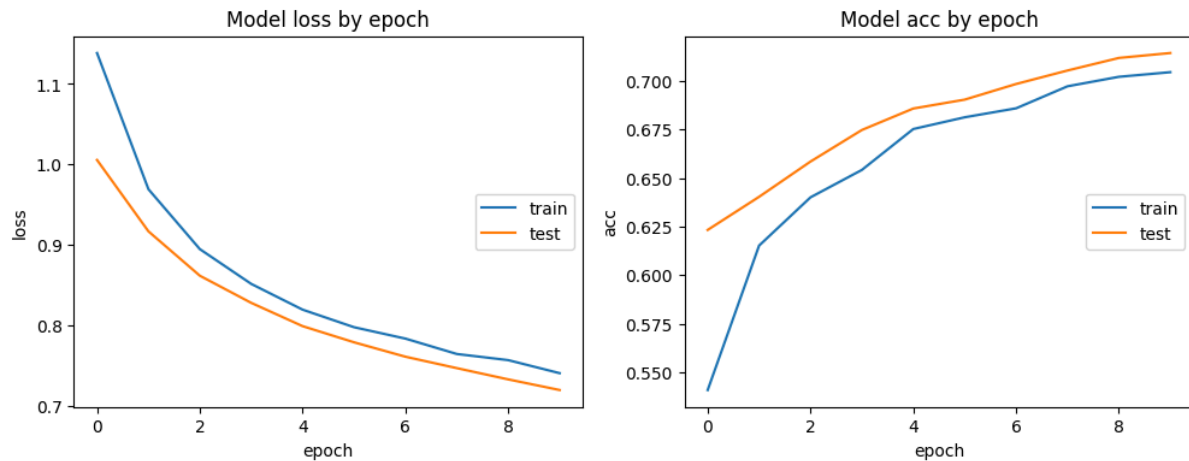


#### Matrice de confusion

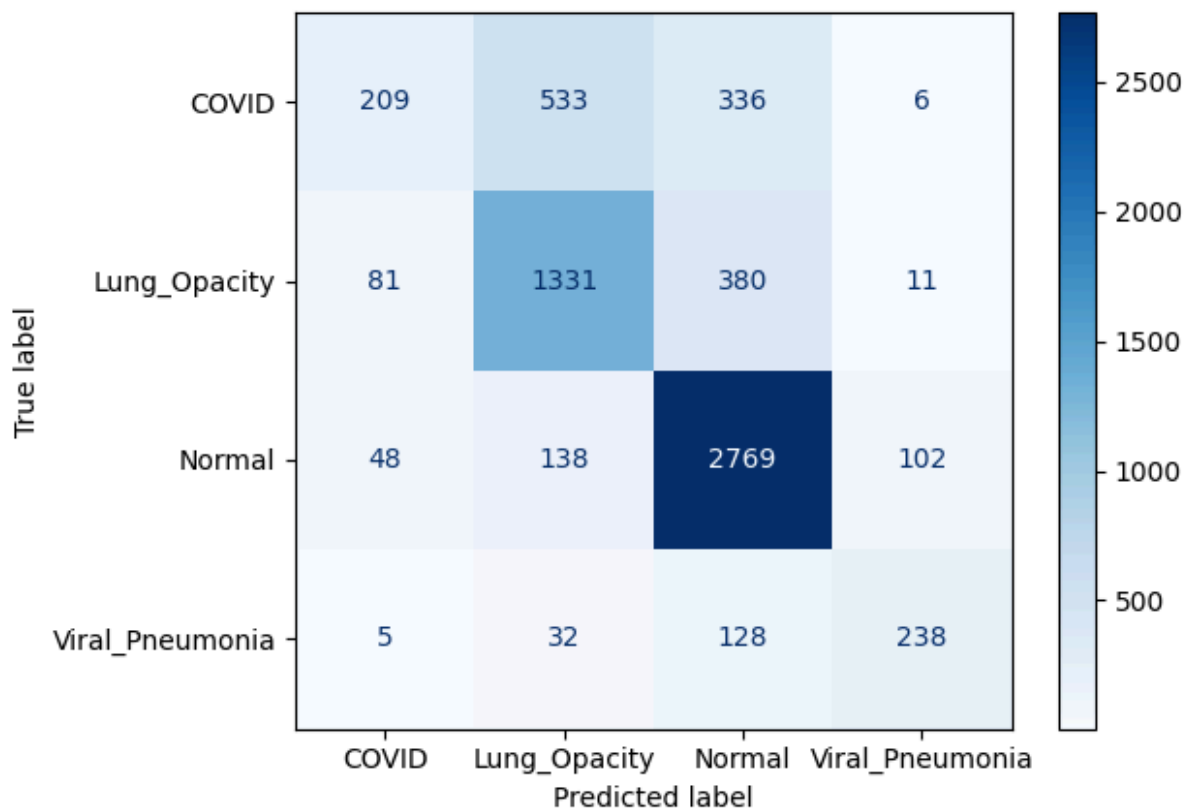


La précision du modèle est de 0.70, mais on observe que les radiographies COVID ne sont bien classées que dans un tiers des cas (recall = 0.33)

## Modèle EfficientNetB2 (342 couches)



La précision obtenue est de 0.72 avec un recall de 0.19 sur la classe COVID.



Les radiographies de la classe COVID sont souvent mal classées avec comme confusion principale l'opacité pulmonaire suivie de la classe Normal.

## Conclusion

Au vu de ces premières analyses, on choisit le modèle ResNet50.

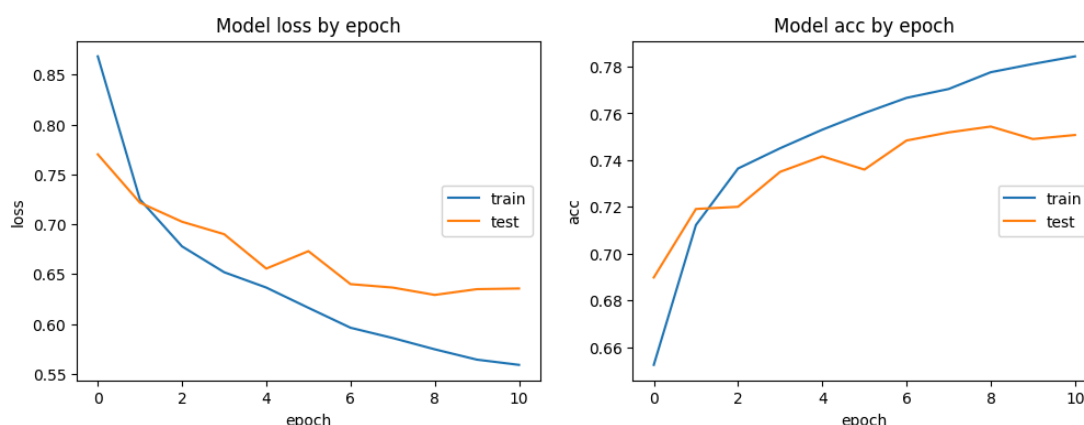
## Augmentation de données, choix du learning rate.

Les données ont été augmentées en utilisant les fonctions de zoom, d'étirement, de décalage vertical et horizontal ainsi que la modification de la luminosité. Cette augmentation a pour objectif d'améliorer la précision du modèle en fiabilisant l'analyse et facilitant sa généralisation.

Trois premières valeurs de learning-rate ont été testées : 0.01, 0.001 et 0.0001. On remarque que la précision augmente plus rapidement avec un learning rate de 0.01 sans pour autant empêcher la convergence de la loss.

On choisit donc de démarrer l'entraînement du modèle avec un learning-rate de 0.01 et de le diminuer ensuite au fur et à mesure des epochs. On intègre également un stop pour arrêter l'entraînement si la loss dans le jeu de données de validation varie de moins de 0.005.

Avec ces modifications, l'accuracy est de 0.75 et le recall sur la catégorie COVID de 0.49.



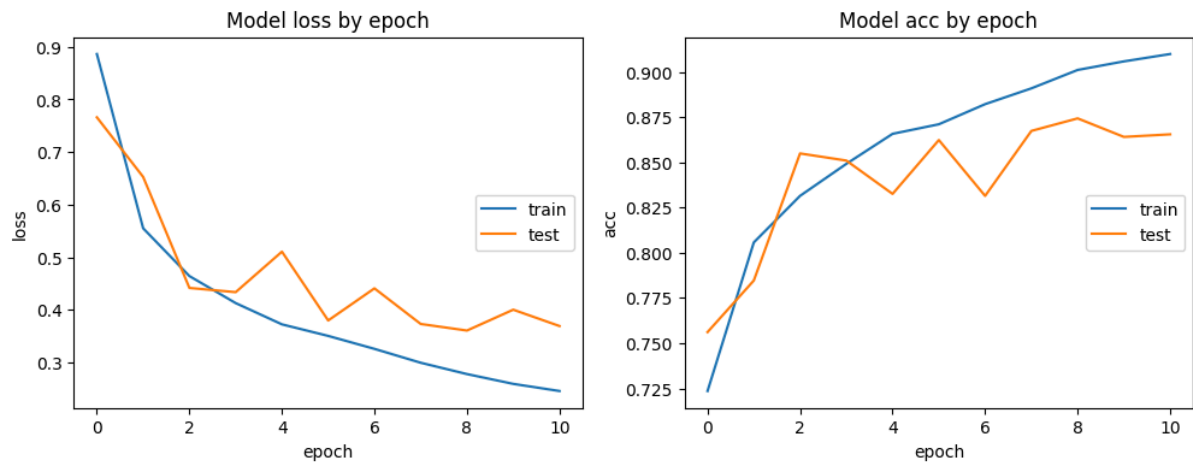
## Fine tuning : Dégel des dernières couches

Trois options ont été testées, l'une en dégelant les 7 dernières couches, la seconde en dégelant les 15 dernières couches et la dernière en dégelant les 30 dernières couches. On présente ici les résultats de la 3ème option qui dégèle le plus de couches et donne la meilleure précision (0.86 de val\_accuracy, vs 0.83 vs 0.77).

## Performance du modèle

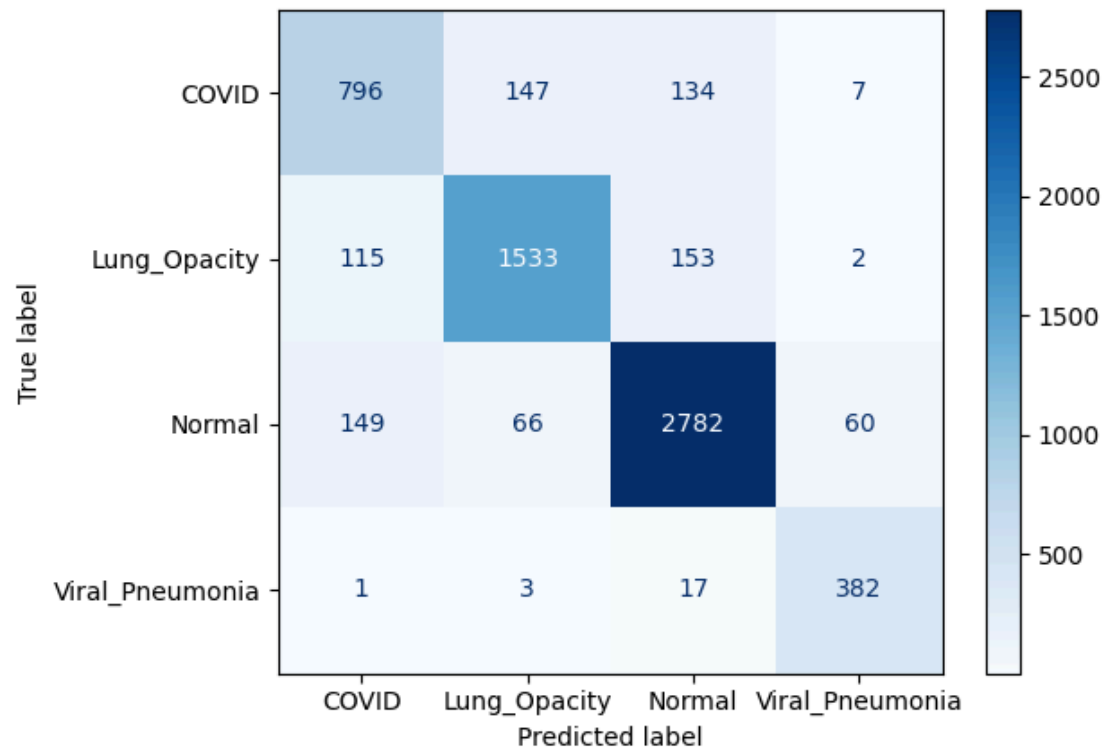
La loss et la précision sont affichées pour le jeu d'entraînement et de validation sont affichés dans le graphique ci-dessous.





On remarque un léger overfitting puisqu'à partir de l'epoch 7 la précision continue d'augmenter sur le jeu d'entraînement sans impact visible sur le jeu de validation.

La matrice de confusion sur le jeu de données test est la suivante,

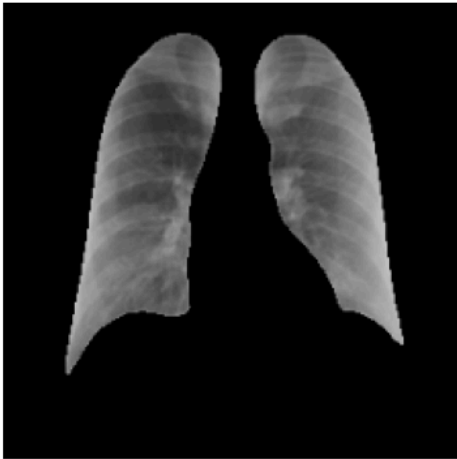


On remarque qu'il y a encore une légère confusion entre les labels COVID et Lung Opacity mais elle est atténuée par rapport au modèle précédent. La précision pour la catégorie d'intérêt COVID est de 0.75 avec un recall de 0.73. L'accuracy macro est de 0.85.

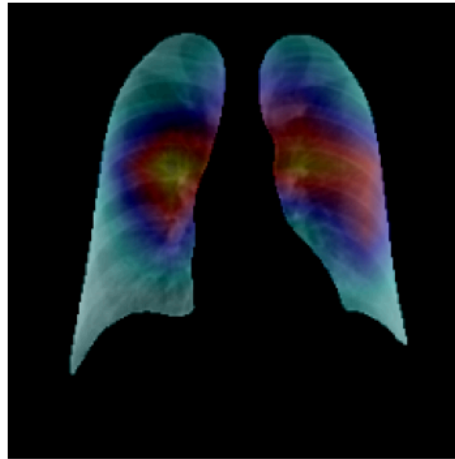
## Gradcam pour visualiser les régions utiles à la prédiction

Les images ci-dessous permettent de superposer les régions qui ont été utilisées pour la prédiction et l'image initiale, y sont également indiqués les labels réels et prédits.

True: COVID



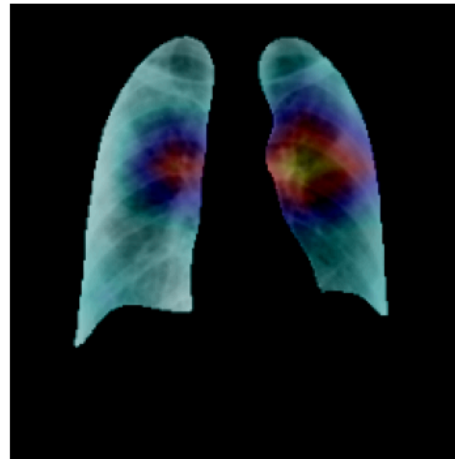
Pred: COVID



True: COVID



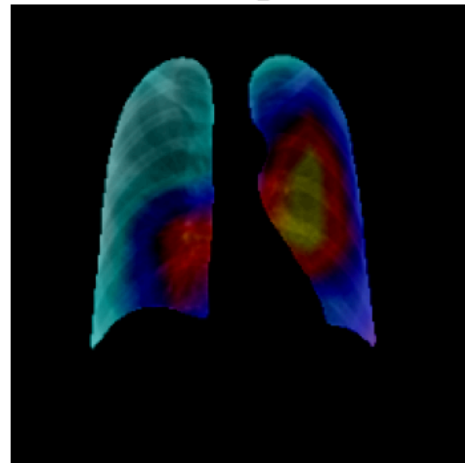
Pred: COVID



True: Lung\_Opacity



Pred: Lung\_Opacity



## Conclusion

Malgré une précision élevée du modèle, il persiste une confusion entre l'opacité pulmonaire et les cas de COVID et l'on observe une précision plus faible pour ces deux catégories. Il semble nécessaire pour améliorer encore la fiabilité du modèle d'utiliser des modèles plus complexes.

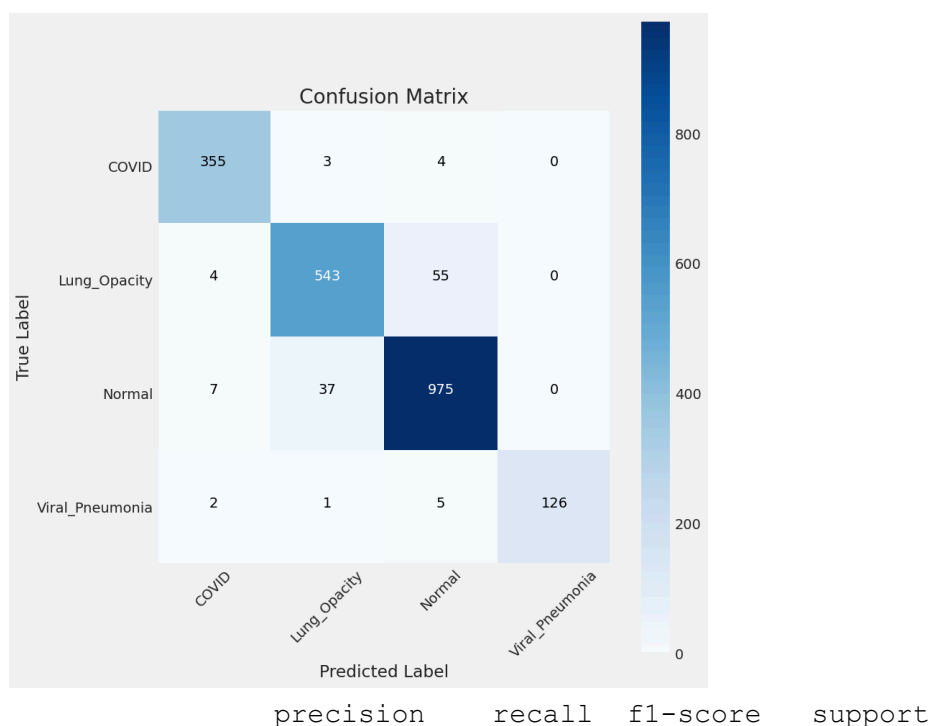
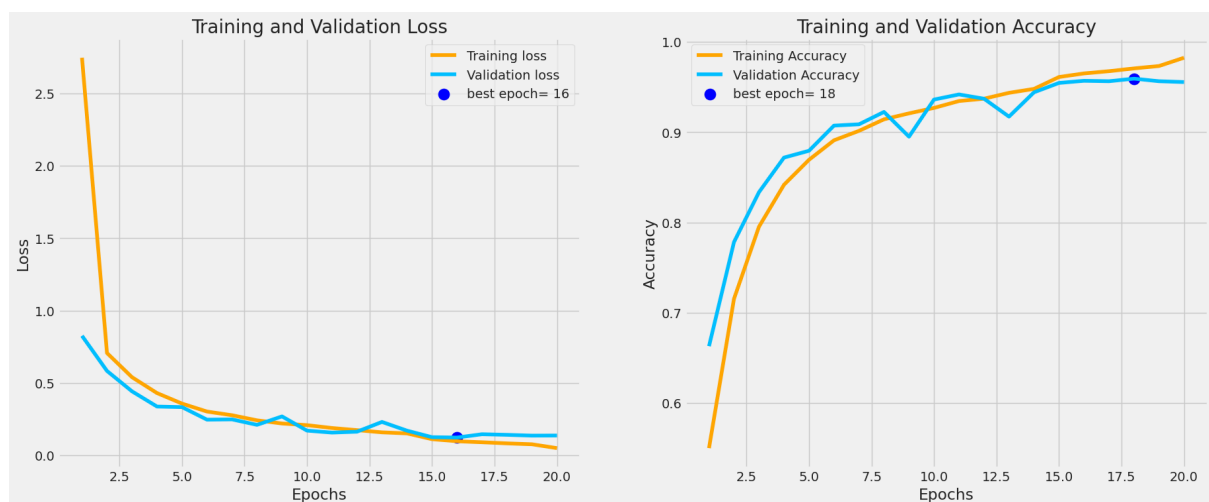
# Ré-estimation de l'ensemble des paramètres

Modèle Micka (training sur 0.8, test sur 0.1 et prédiction sur 0.1)

Ici l'apprentissage a été fait sur 80% du jeu de données, 10% des données ont été utilisées pour estimer les paramètres (test) et 10% pour la prédiction finale (validation). La taille des lots a été fixée à 32. On utilise ici les images brutes sans application du masque.

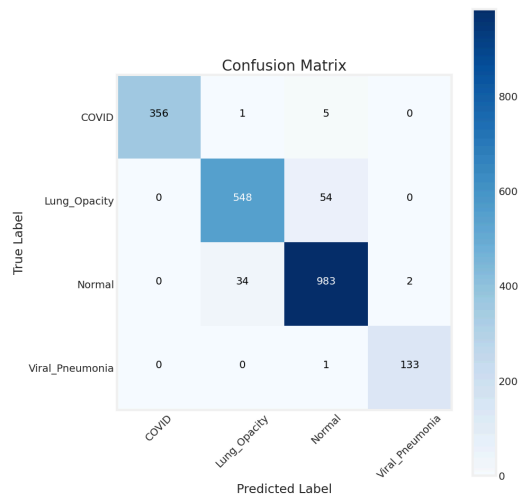
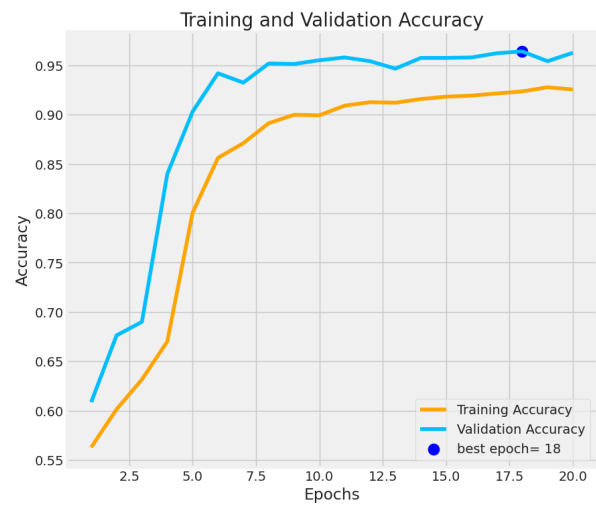
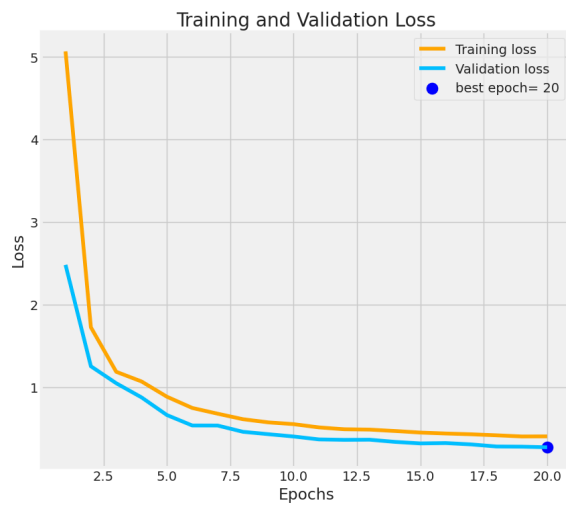
Les modèles testés sont VGG19, EfficientNetB4, EfficientNetB0 et DenseNet169 et Resnet101.

## Modèle VGG19 (19 couches)



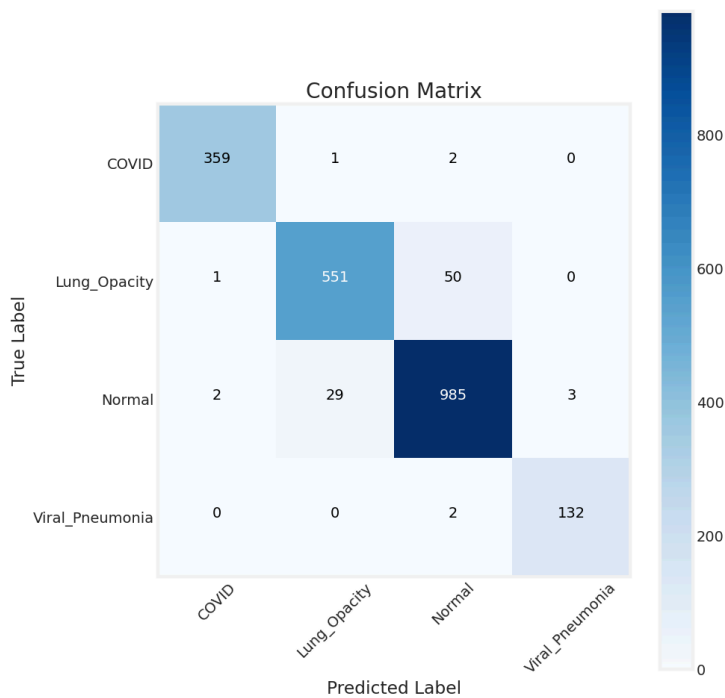
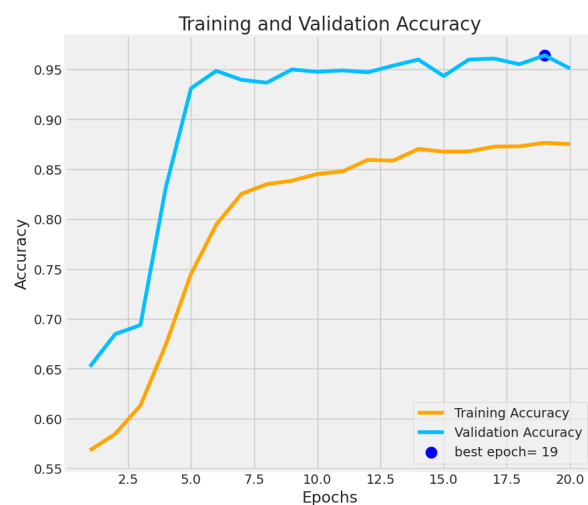
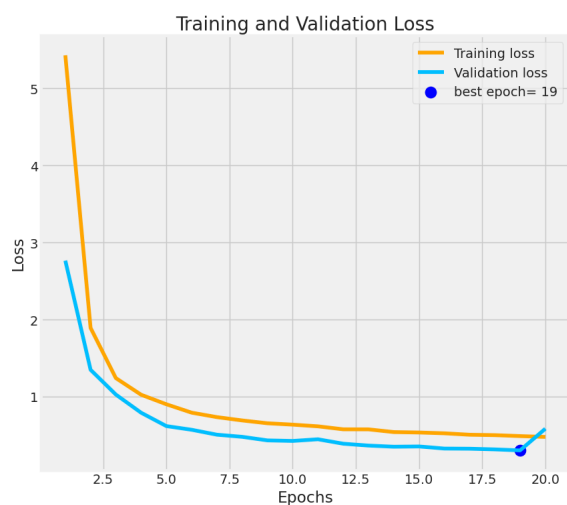
COVID	0.96	0.98	0.97	362
Lung_Opacity	0.93	0.90	0.92	602
Normal	0.94	0.96	0.95	1019
Viral_Pneumonia	1.00	0.94	0.97	134
accuracy			0.94	2117
macro avg	0.96	0.94	0.95	2117
weighted avg	0.94	0.94	0.94	2117

## Modèle EfficientNetB0 (237 couches)



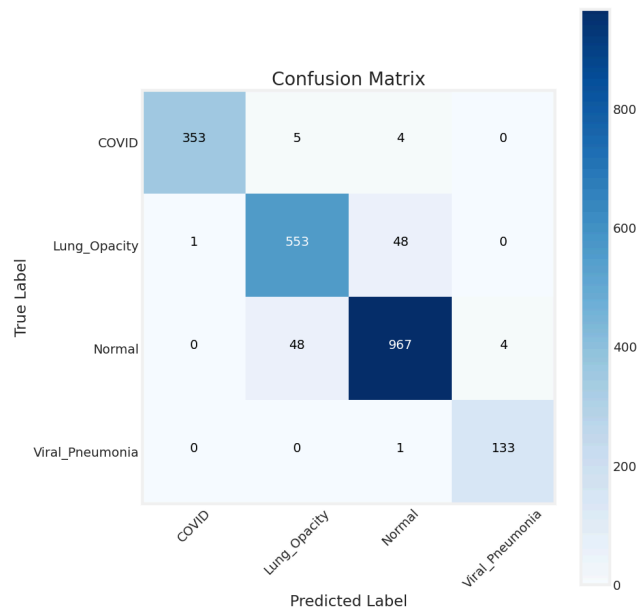
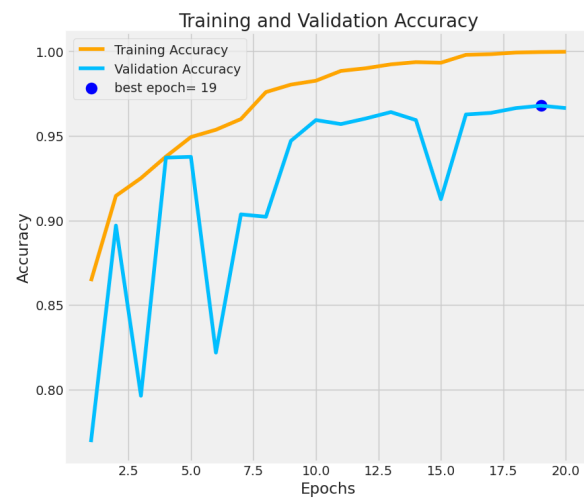
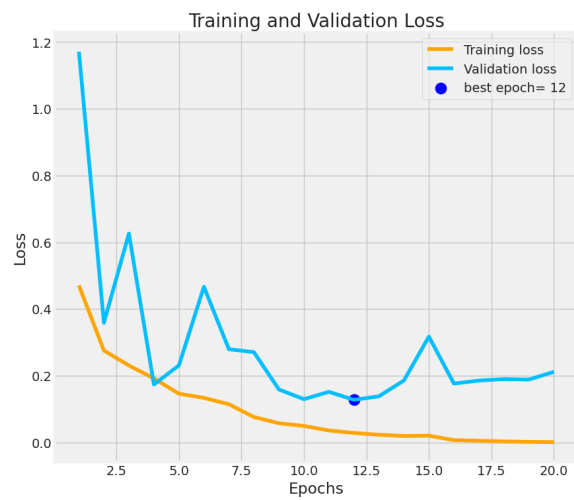
	precision	recall	f1-score	support
COVID	1.00	0.98	0.99	362
Lung_Opacity	0.94	0.91	0.92	602
Normal	0.94	0.96	0.95	1019
Viral_Pneumonia	0.99	0.99	0.99	134
accuracy			0.95	2117
macro avg	0.97	0.96	0.96	2117
weighted avg	0.95	0.95	0.95	2117

## Modèle EfficientNetB4 (381 couches)



	precision	recall	f1-score	support
COVID	0.99	0.99	0.99	362
Lung_Opacity	0.95	0.92	0.93	602
Normal	0.95	0.97	0.96	1019
Viral_Pneumonia	0.98	0.99	0.98	134
accuracy			0.96	2117
macro avg	0.97	0.96	0.97	2117
weighted avg	0.96	0.96	0.96	2117

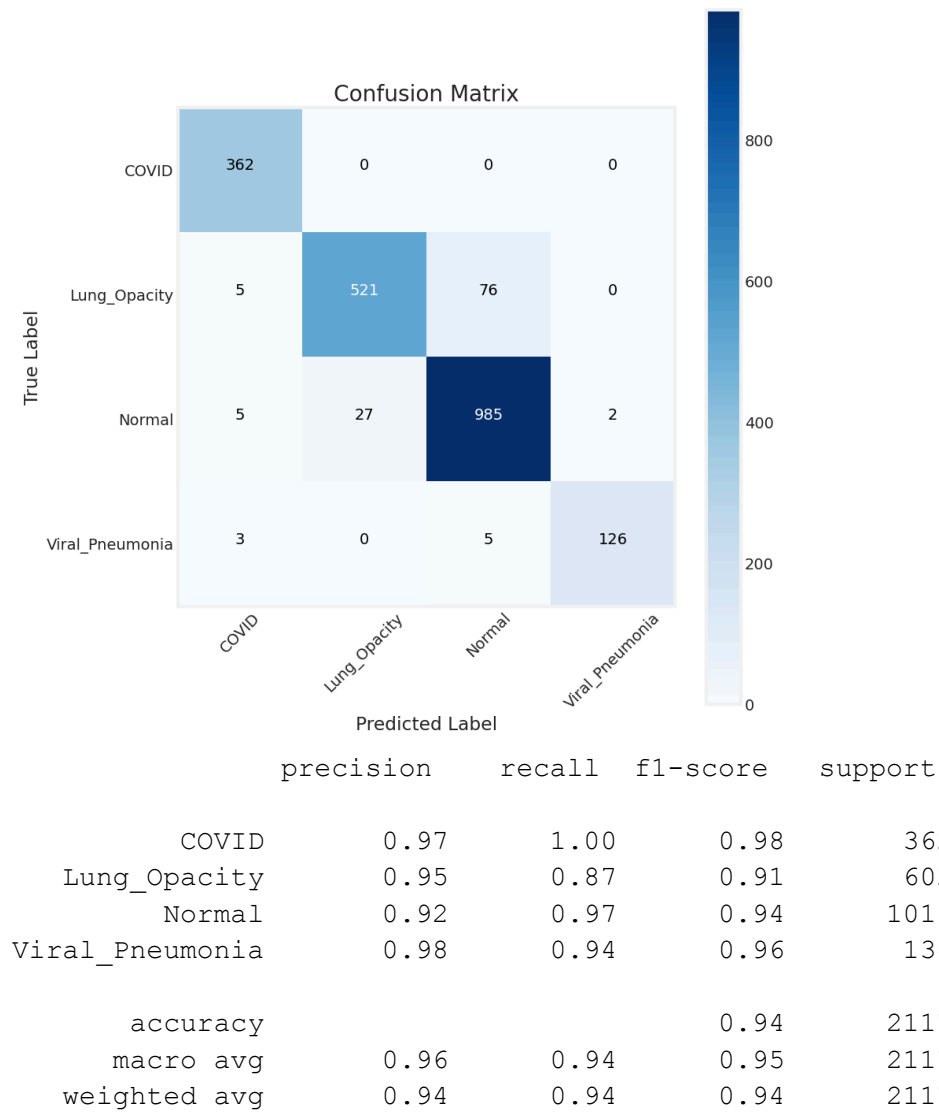
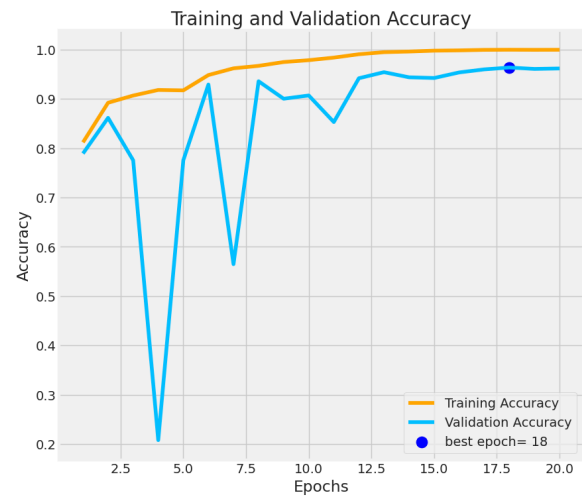
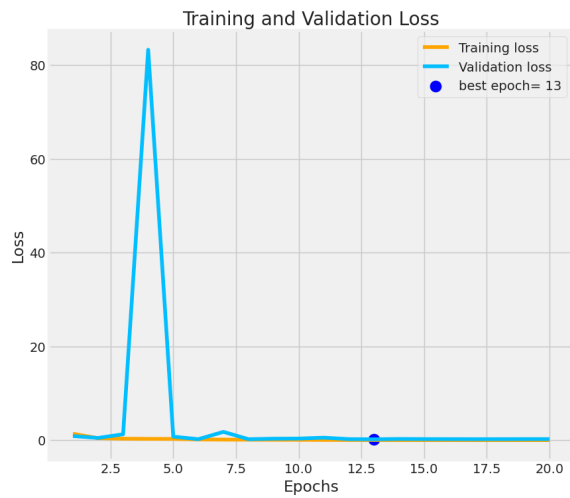
## Modèle DenseNet169 (82 couches<sup>1</sup>)



	precision	recall	f1-score	support
COVID	1.00	0.98	0.99	362
Lung_Opacity	0.91	0.92	0.92	602
Normal	0.95	0.95	0.95	1019
Viral_Pneumonia	0.97	0.99	0.98	134
accuracy			0.95	2117
macro avg	0.96	0.96	0.96	2117
weighted avg	0.95	0.95	0.95	2117

<sup>1</sup><https://amaarora.github.io/posts/2020-08-02-densenets.html#densenet-architecture-as-a-collection-of-denseblocks>

## Modèle Resnet101 (x couches)



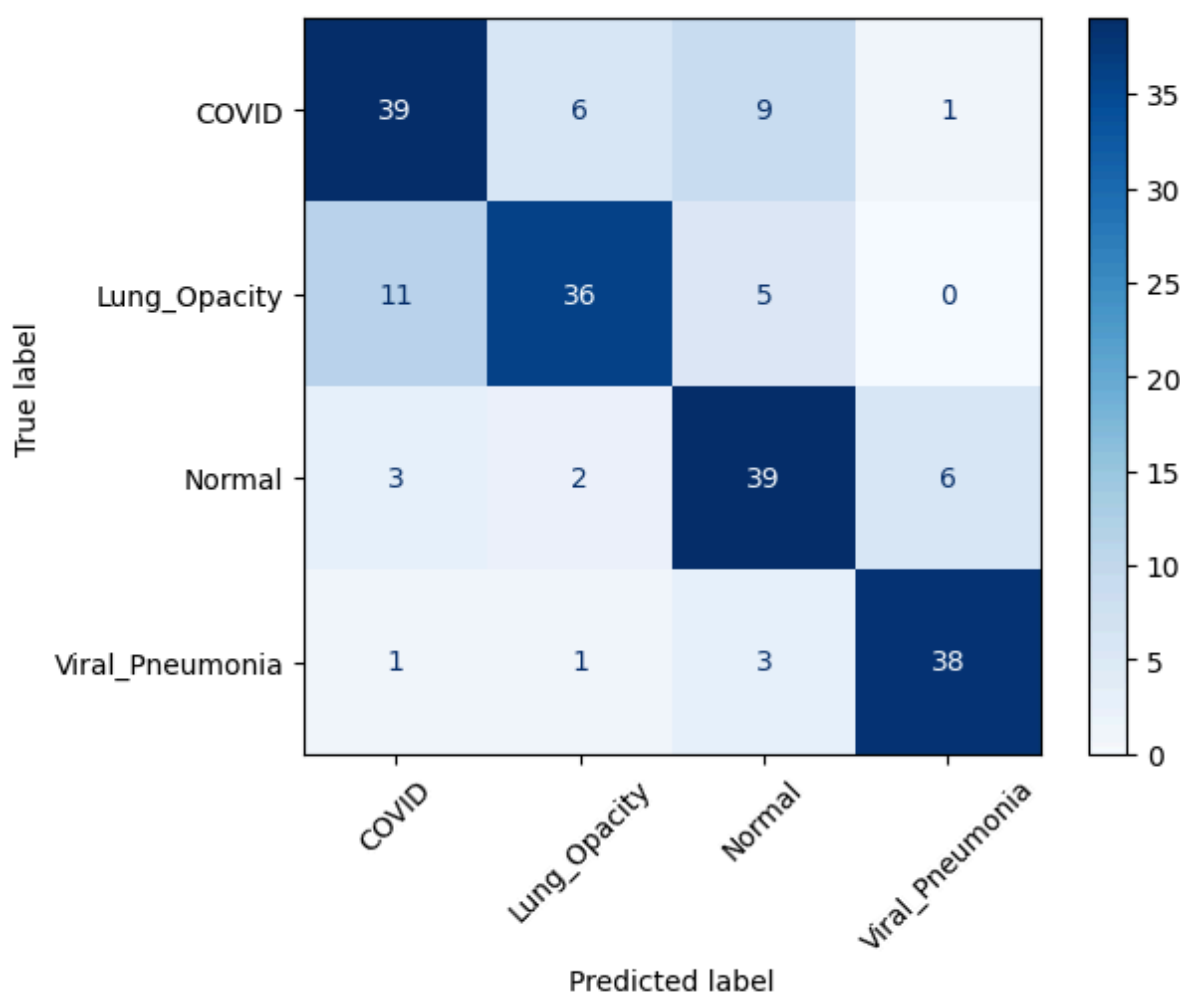
## Conclusion

Les précisions obtenues en estimant l'ensemble des paramètres sont très élevées quel que soit le modèle. Le meilleur modèle est Efficient Net B4 avec une précision globale de 0.96 et de 0.99 pour la catégorie COVID aussi bien en accuracy qu'en recall.

La meilleure précision du modèle où l'ensemble des couches ont été réentraînées peut s'expliquer par la spécificité des données de radiographies par rapport aux images qui ont été utilisées dans image\_net.

## Ouverture : test Model de type Transformers ViT

Un modèle ViT a été essayé sur un jeu de données réduit de 2000 images et 2 epochs avec les résultats suivants pour la matrice de confusion et le rapport de classification.





	precision	recall	f1-score	support
COVID	0.72	0.71	0.72	55
Lung_Opacity	0.80	0.69	0.74	52
Normal	0.70	0.78	0.74	50
Viral_Pneumonia	0.84	0.88	0.86	43
accuracy			0.76	200
macro avg	0.77	0.77	0.76	200
weighted avg	0.76	0.76	0.76	200

Les résultats sont encourageants en particulier si on les met en regard du faible nombre d'epoch et de la taille du dataset. Il est intéressant de voir que l'accuracy est relativement similaire en fonction des classes. Les temps de calcul sont cependant rédhibitoires avec environ 30mn par epoch. Il faudra étudier plus profondément le sujet des transformers pour voir si la piste mérite ou non d'être creusée.