



DataScientest

**COVINET
DÉTECTION DU COVID-19 À PARTIR
DE RADIOGRAPHIES PULMONAIRES**

**Projet de fin de cursus
DSJUL23**

**BURKE Hassan
GRECO Juliette
LANCMAN Nathan
KHAIRALLAH Matthieu**

Chef de projet : PENESSOT Gaël

TABLE DES MATIERES

1. CONTEXTE	6
2. OBJECTIFS	7
3. CADRE	8
4. MÉTHODOLOGIE	9
4.1. Etude Sain vs Malade -----	9
4.2. Etude des quatre catégories -----	9
5. PRE-PROCESSING ET FEATURE ENGINEERING.....	10
5.1. Reduction de dimension -----	10
5.1.1. Adaptation de la taille des images -----	10
5.2. Répartition et équilibrage par classe -----	10
5.3. Superposition des masques-----	11
6. DATAVIZ ET STATISTIQUES	12
6.1. Classes des images -----	12
6.2. Visualisation des images -----	12
6.2.1. Image moyenne -----	13
6.2.2. Contraste entre images moyennes -----	14
6.2.3. Analyse des composants principaux ----	Erreur ! Signet non défini.
7. RESULTATS	18
7.1. Étude sain vs malade sans masque -----	18
7.1.1. LeNet5 (modèle de référence) -----	18
7.1.2. EfficientNetB1 -----	18

7.1.3.	VGG16 -----	18
7.1.4.	Xception -----	19
7.1.5.	ResNet152 -----	19
7.2.	Étude sain vs malade avec masques -----	19
7.2.1.	EfficientNetB1 -----	19
7.2.2.	VGG16 -----	19
7.2.3.	ResNet152 -----	20
7.2.4.	Xception -----	20
7.3.	Étude des quatre classes sans masque -----	20
7.3.1.	EfficientNetB1 -----	20
7.3.2.	VGG16 -----	20
7.3.3.	Xception -----	21
7.3.4.	ResNet152 -----	21
7.4.	Étude des quatre classes avec masque -----	21
7.4.1.	EfficientNetB1 -----	21
7.4.2.	VGG16 -----	21
7.4.3.	Xception -----	22
7.4.4.	ResNet152 -----	22
8.	DISCUSSION.....	23
8.1.	Comparaison des études « avec masque » et « sans masque » -----	23
8.2.	Faux positifs et faux négatifs -----	23
8.3.	Compromis entre précision et puissance de calcul-----	24
8.4.	Optimisation des hyperparamètres -----	25
8.5.	Interprétabilité du modèle -----	25
8.6.	AXES D'AMÉLIORATION-----	26
9.	CONCLUSION.....	28

REFERENCES	29
LISTE DES TABLEAUX.....	31
LISTE DES FIGURES.....	32
LISTE DES ABREVIATIONS ET DES SYMBOLES	33
LISTE DES ANNEXES.....	34
ANNEXE I : RESULTATS DU LENET5 SUR LA CLASSIFICATION BINAIRE SANS MASQUE	35
ANNEXE II : RESULTATS DU EFFICIENTNETB1 SUR LA CLASSIFICATION BINAIRE SANS MASQUE	36
ANNEXE III : RESULTATS DU VGG16 SUR LA CLASSIFICATION BINAIRE SANS MASQUE	37
ANNEXE IV : RESULTATS DU RESNET152 SUR LA CLASSIFICATION BINAIRE SANS MASQUE	38
ANNEXE V : RESULTATS DU XCEPTION SUR LA CLASSIFICATION BINAIRE SANS MASQUE	39
ANNEXE VI : RESULTATS DU EFFICIENTNETB1 SUR LA CLASSIFICATION BINAIRE AVEC MASQUE	40
ANNEXE VII : RESULTATS DU VGG16 SUR LA CLASSIFICATION BINAIRE AVEC MASQUE	41
ANNEXE VII : RESULTATS DU RESNET152 SUR LA CLASSIFICATION BINAIRE AVEC MASQUE	42
ANNEXE IX : RESULTATS DU XCEPTION SUR LA CLASSIFICATION BINAIRE AVEC MASQUE	43

ANNEXE X : RESULTATS DU EFFICIENTNETB1 SUR LA CLASSIFICATION MULTI-CLASSE SANS MASQUE.....	44
ANNEXE XI : RESULTATS DU VGG16 SUR LA CLASSIFICATION MULTI-CLASSE SANS MASQUE	45
ANNEXE XII : RESULTATS DU XCEPTION SUR LA CLASSIFICATION MULTI-CLASSE SANS MASQUE.....	46
ANNEXE XIII : RESULTATS DU RESNET152 SUR LA CLASSIFICATION MULTI-CLASSE SANS MASQUE.....	47
ANNEXE XIV : RESULTATS DU EFFICIENTNETB1 SUR LA CLASSIFICATION MULTI-CLASSE AVEC MASQUE.....	48
ANNEXE XV : RESULTATS DU VGG16 SUR LA CLASSIFICATION MULTI-CLASSE AVEC MASQUE	49
ANNEXE IV : RESULTATS DU XCEPTION SUR LA CLASSIFICATION MULTI-CLASSE AVEC MASQUE	50
ANNEXE IV : RESULTATS DU XCEPTION SUR LA CLASSIFICATION MULTI-CLASSE AVEC MASQUE	51

1. CONTEXTE

L'année 2019 a été marquée par l'émergence rapide et inattendue du COVID-19, une maladie infectieuse causée par le coronavirus SARS-CoV-2¹. Cette pandémie mondiale a engendré des défis sans précédent pour les systèmes de santé, les économies et les sociétés dans leur ensemble. L'une des caractéristiques distinctives de cette maladie est son impact sur les systèmes respiratoires des individus, ce qui a suscité l'intérêt de la communauté scientifique pour développer des moyens de détection efficaces et rapides.

Le test PCR (Réaction en Chaîne par Polymérase) a été largement reconnu comme le "*gold standard*" pour la détection du COVID-19². Cependant, lors des pics de la pandémie, la demande accrue de tests PCR a parfois entraîné des ruptures d'approvisionnement, limitant ainsi la capacité de dépistage. Face à cette situation, l'imagerie médicale a constitué une alternative pour le diagnostic. Les radiographies pulmonaires et les tomodensitogrammes (CT-scans) sont devenus des outils essentiels pour identifier les signes distinctifs de l'infection pulmonaire causée par le virus^{3,4}.

La recherche en *deep learning*, une sous-discipline de l'intelligence artificielle, a connu un développement exponentiel au cours des dernières années. Les algorithmes de *deep learning* ont démontré leur capacité à apprendre des caractéristiques complexes à partir de données brutes, en particulier dans le domaine de la vision par ordinateur (*computer vision*)⁵. Dans le contexte médical, cette technologie a suscité un vif intérêt pour son potentiel à développer des modèles de classification automatisés d'images médicales, tels que les radiographies pulmonaires, pour la détection du COVID-19^{6,7,8,9,10}.

Cependant, la question clé persiste : peut-on vraiment entraîner un modèle de *deep learning* fiable et précis en utilisant des radiographies pulmonaires pour détecter le COVID-19 ? Cette interrogation soulève des considérations cruciales en termes de performance du modèle, de qualité des données, de généralisation et d'impact potentiel dans le domaine médical. À travers cette étude, nous plongerons dans cette problématique, en explorant les succès et les défis associés à l'application des techniques de *deep learning* à la classification des images pulmonaires liées au COVID-19.

2. OBJECTIFS

A l'aide d'une banque d'images de radiographies pulmonaires, nous souhaitons entraîner un modèle de *deep learning* afin de vérifier si celui-ci est capable de détecter efficacement la présence, ou non, de COVID-19.

Hassan B. et Nathan L. possèdent des connaissances déjà établies en *machine learning* mais sont novices dans le domaine du *deep learning*. Juliette G. et Matthieu K. sont des débutants dans le domaine de la Data Science.

Afin d'apporter plus de contexte autour de notre problématique nous avons effectué une brève revue de la littérature au sujet de la détection de pathologies à l'aide d'imageries médicales.

Des projets similaires au nôtre ont déjà été étudiés et publiés. Certains se sont intéressés à l'efficacité d'un modèle de *deep learning* dans la détection du COVID-19 à partir d'une base de données de radiographies pulmonaires, d'autres l'ont fait à partir de CT-scans.

A titre d'exemple, Yang et al. (2020)¹¹, ont développé une base de données publique d'images CT axée sur le COVID-19 dans le but de promouvoir le développement de méthodes d'intelligence artificielle visant à exploiter les CT-scans pour le dépistage et les tests des patients atteints de COVID-19. Cette base de données contient 349 images CT de patients atteints de COVID-19, provenant de 216 individus, ainsi que 463 images CT non liées au COVID-19 (utilisées comme exemples d'apprentissage négatifs). Ils ont réalisé des études expérimentales pour observer l'efficacité de cette base de données. En utilisant ces données, ils ont élaboré une approche basée sur l'apprentissage multitâche et l'apprentissage auto-supervisé contrastif, obtenant un score F1 de 0,90, une AUC de 0,98 et une précision de 0,89 lors de l'évaluation sur un ensemble de test composé d'images CT originales fournies par les hôpitaux. Selon les auteurs, de tels scores confirment que leurs modèles ont une utilité clinique avérée pour assister les radiologues dans le dépistage du COVID-19.

Le coût d'un CT-scan du thorax (plus précis qu'une radiographie¹²) étant plus élevé que celui d'une radiographie¹³, notre projet cherche à entraîner un modèle de *deep learning* sur des radiographies avec des performances au moins aussi satisfaisantes que celles obtenues avec le modèle entraîné sur CT-scan.

3. CADRE

Pour mener à bien notre projet, nous exploitons le jeu de données « COVID-19 Radiography Database »¹⁴. Ces données sont librement disponibles sur le site web [Kaggle](#).

Ce jeu de données contient des radiographies pulmonaires (au format *.png*) et leurs masques associés (au format *.png*) ainsi que leurs métadonnées (dans des fichiers *.xlsx*). Il a été créé par des chercheurs de l'Université du Qatar, à Doha, au Qatar, et de l'Université de Dhaka, au Bangladesh, ainsi que leurs collaborateurs du Pakistan et de Malaisie, en collaboration avec des médecins. Ce jeu de données comprend des cas de COVID-19, des cas normaux et d'autres infections.

Dans la première version, ils ont publié 219 images de radiographies pulmonaires pour les cas de COVID-19, 1341 pour les cas normaux et 1345 pour les cas de pneumonie virale. Dans la première mise à jour, le nombre d'images pour la classe COVID-19 a été augmenté à 1200.

Dans la deuxième mise à jour, la base de données a été étendue à 3616 cas de COVID-19 positifs, ainsi qu'à 10 192 cas normaux, 6012 cas d'opacité pulmonaire (infections pulmonaires non liées au COVID) et 1345 images de pneumonie virale, avec les masques pulmonaires correspondants. C'est cette dernière version que nous utilisons. La taille du dossier est de 816 Mo. La *Figure 1* illustre la provenance des images.

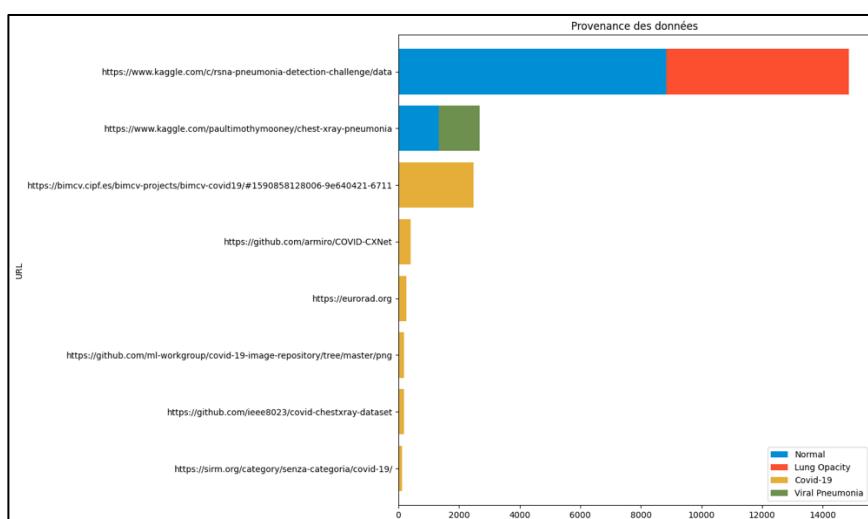


Figure 1 - Provenance des données

4. MÉTHODOLOGIE

Notre projet se décompose en deux grandes parties. Premièrement nous avons étudié la possibilité de mettre au point un modèle permettant de distinguer les patients sains des patients malades. Dans un deuxième temps, nous avons étudié la possibilité de discriminer la présence ou non de la COVID-19.

4.1. ETUDE SAIN VS MALADE

Nous avons dans un premier temps mis au points un modèle permettant de distinguer des radiographies pulmonaires saines (catégorie « Normal ») des radiographies malades (catégories « COVID », « Lung Opacity » et « Viral Pneumonia »).

Pour répondre à ce besoin nous nous sommes orientés vers les réseaux neuronaux convolutifs qui sont particulièrement efficaces en *computer vision* et dans la classification d'images.

Afin de sélectionner un modèle nous avons testé différents réseaux neuronaux convolutifs : LeNet, VGG16, ResNet152, Xception, EfficientNet, ...

4.2. ETUDE DES QUATRE CATEGORIES

Dans un deuxième temps nous avons focalisé notre travail sur la mise en place d'un modèle permettant non plus simplement de détecter la maladie mais également de pouvoir identifier la présence de la COVID à l'aide des imageries.

Nous avons également eu recours aux réseaux neuronaux convolutifs. Nous avons choisi le même que pour la problématique précédente.

5. PRE-PROCESSING ET FEATURE ENGINEERING

5.1. REDUCTION DE DIMENSION

5.1.1. Adaptation de la taille des images

Les images du jeu de données sont toutes de taille 299x299 pixels. Leurs masques associés sont de 256x256 pixels. Nous avons dû procéder à un ajustement de la taille des images et des masques pour les adapter aux différents modèles que nous avons testés.

Cette manipulation est également susceptible d'optimiser les temps d'entraînement et de réduire l'usage de la mémoire.

5.2. REPARTITION ET EQUILIBRAGE PAR CLASSE

Notre modèle final vise à détecter si un patient est atteint du COVID-19 ou non. Nous avons fait le choix d'entraîner quatre modèles que nous décrivons ci-après :

- Le premier modèle vise à déterminer si le patient est sain ou pathologique. En d'autres termes de savoir si l'image appartient à la classe « Normal » ou non.
- Le deuxième modèle a le même objectif que le premier mais en plus des images nous utiliserons leurs masques associés fournis dans le jeu de données. Nous souhaitons voir si le recours aux masques a une incidence significative sur la performance du modèle.
- Le troisième modèle vise à déterminer sur une image n'appartenant pas à la classe « Normal », si celle-ci appartient à la classe « COVID ».
- Le quatrième modèle est similaire au troisième mais sera réalisé avec la superposition des masques de chaque image.

Certaines catégories possèdent peu d'images. Ainsi, nous avons procédé à une *data augmentation* dans l'étude des quatre classes afin de pouvoir entraîner notre modèle sur un plus grand jeu de données.

```
1 train_datagen = ImageDataGenerator(  
2     rotation_range = 0.15,  
3     width_shift_range = 0.2,  
4     height_shift_range = 0.2,  
5     shear_range = 0.2,  
6     zoom_range = 0.2,  
7     fill_mode = 'nearest')
```

5.3. SUPERPOSITION DES MASQUES

Dans le cadre de notre projet, l'intégration de masques de la région pulmonaire constitue une stratégie pertinente que nous avons souhaité évaluer. Ces masques délimitent précisément la zone d'intérêt, permettant ainsi une focalisation sur la région pulmonaire tout en réduisant le bruit provenant d'éléments non pertinents.



Figure 2 - Exemple de superposition de masque

6. DATAVIZ ET STATISTIQUES

6.1. CLASSES DES IMAGES

Les radiographies du jeu de données sont réparties en quatre classes :

- Normal
- COVID
- Viral pneumonia
- Lung opacity (*lung opacity* peut faire référence à différentes pathologies)

La *Figure 3* illustre le nombre de radiographies de chaque classe.

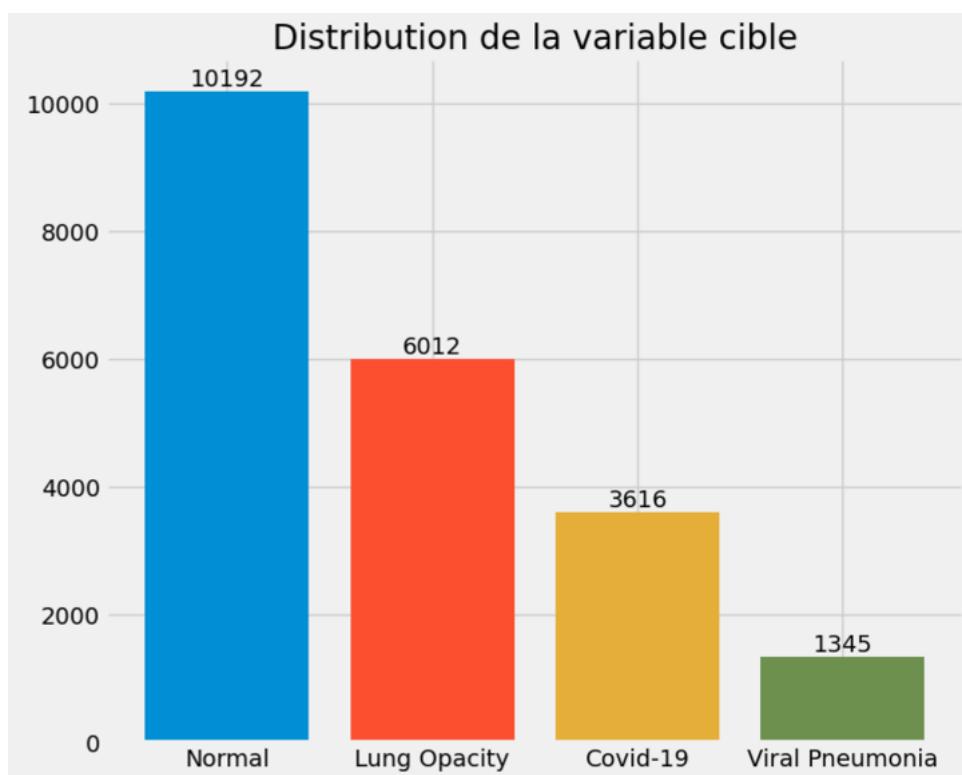


Figure 3 - Nombre de radiographies dans chaque classe

6.2. VISUALISATION DES IMAGES

Afin de se donner un aperçu des images qui constituent notre jeu de données nous avons rédigé un script en Python qui génère une figure avec 12 images sélectionnées au hasard à partir de leurs classes respectives (*Figure 4*).

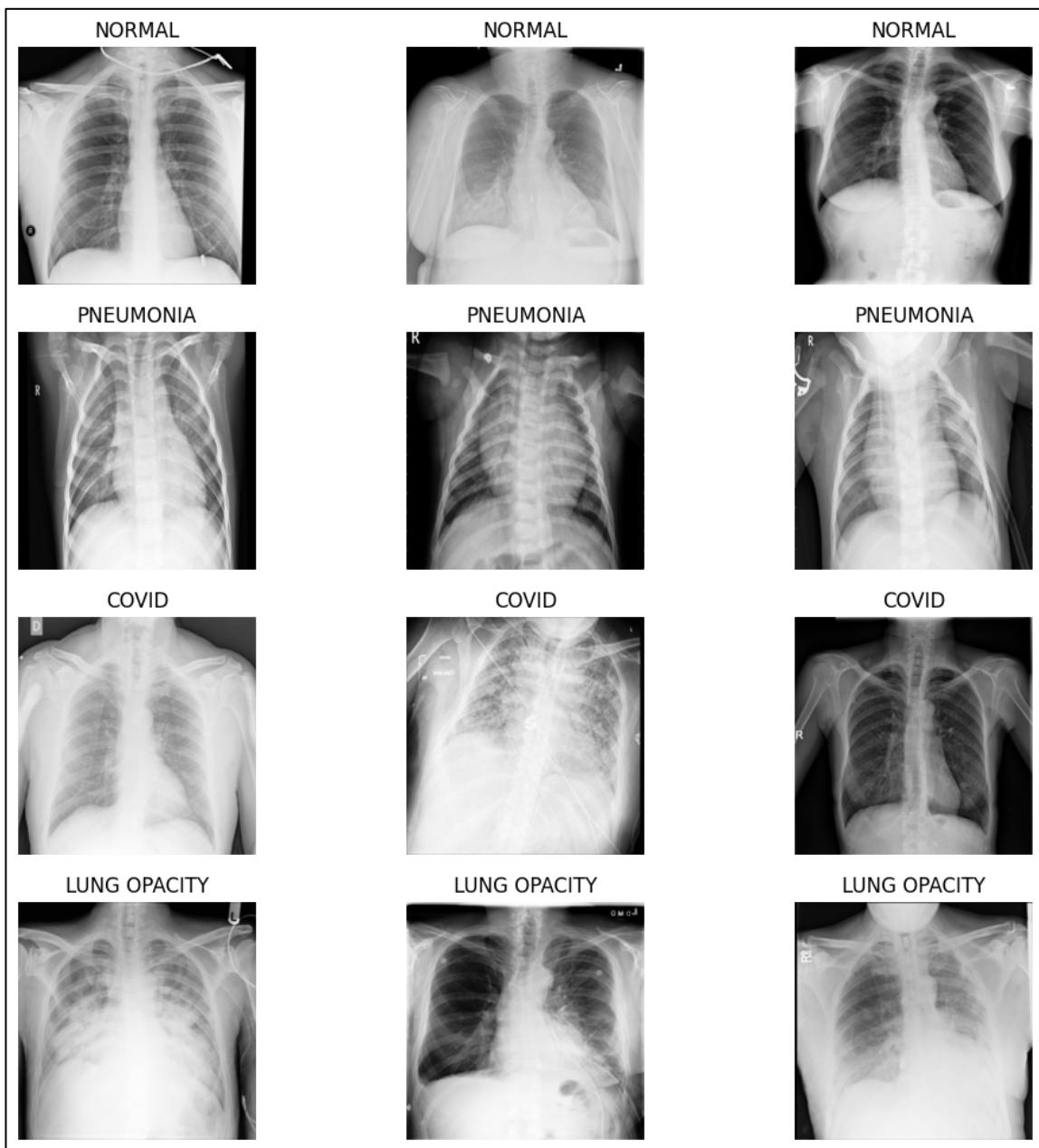


Figure 4 - Générateur d'images

6.2.1. Image moyenne

L'image moyenne générée pour chaque classe nous a permis d'observer certaines caractéristiques propres à chaque classe (*Figure 5*). Nous avons pu ainsi remarquer que les radiographies des sous-classes « COVID » et « Lung_Opacity » semblaient être les plus obstruées dans la région pulmonaire. Inversement, nous avons constaté une opacité nettement différente dans la région pulmonaire dans la sous-classe « Normal ».

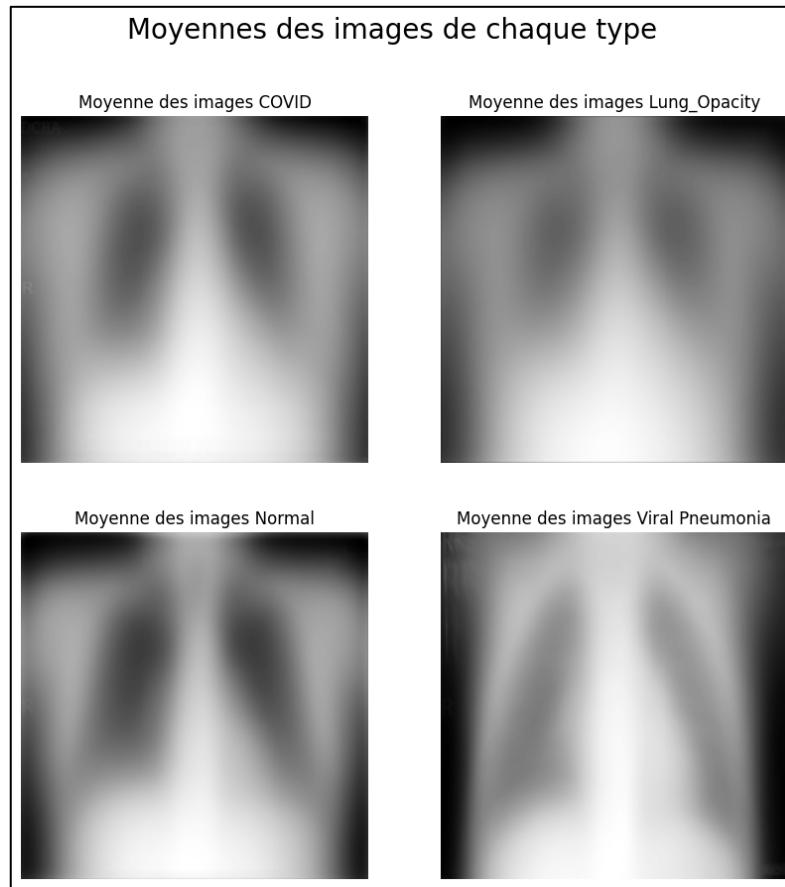


Figure 5 - Moyenne des images de chaque classe

6.2.2. Contraste entre images moyennes

A l'aide des images moyennes nous avons pu calculer les différences entre la sous-classe « COVID » et le reste (*Figure 6*). D'une part nous remarquons que face aux images « Normal », l'intensité des pixels dans la région pulmonaire des radiographies « COVID » a tendance être plus élevée. En imagerie, ceci pourrait se traduire, comme nous l'avons vu pour la moyenne des images, d'une obstruction au niveau des poumons. D'autre part, lorsque nous comparons « COVID » et « Viral pneumonia », le constat semble plus mitigé : les zones pulmonaires renvoient des pixels aux valeurs plus élevées dans certaines zones, ce qui pourrait impacter la compréhension des images par notre modèle.

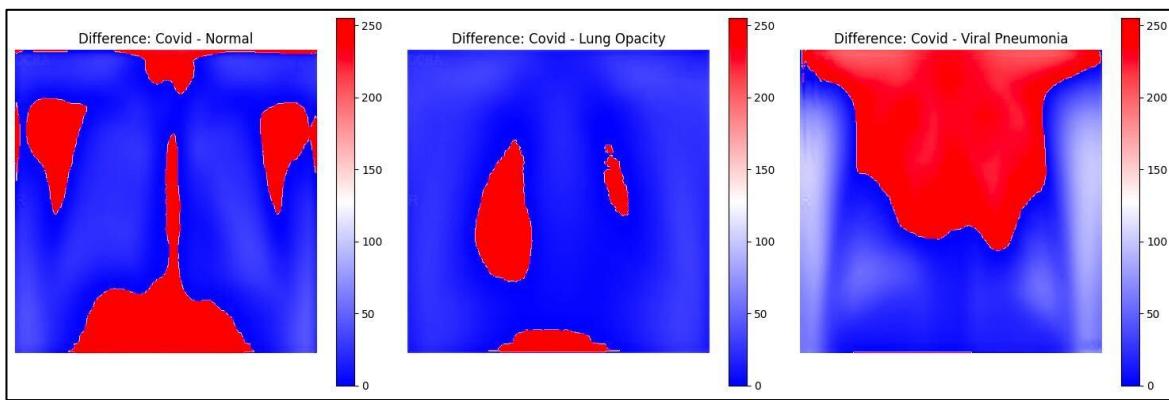


Figure 6 - Contraste entre l'image moyenne COVID et celles des autres classes

Nous avons procédé à la même manipulation pour observer le contraste entre l'image moyenne de la classe « Normal » à celles des autres classes (*Figure 7*).

On remarque dans les trois cas que la valeur des pixels dans la zone pulmonaire des imageries pathologiques est plus élevée que celle des images de la classe « Normal ».

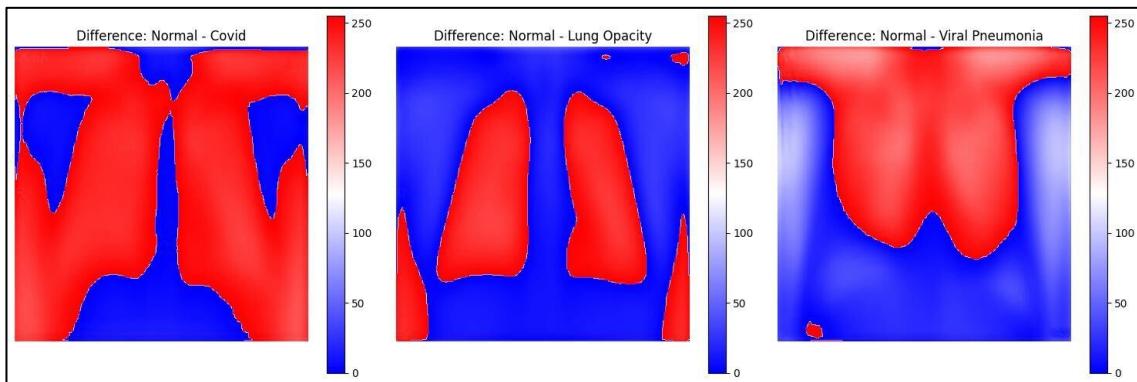


Figure 7 - Contraste entre l'image moyenne "Normal" et celles des autres classes

6.2.3. Eigenimages

Nous avons généré des eigenimages pour chaque image de chaque classe (*Figures 8, 9, 10 & 11*).

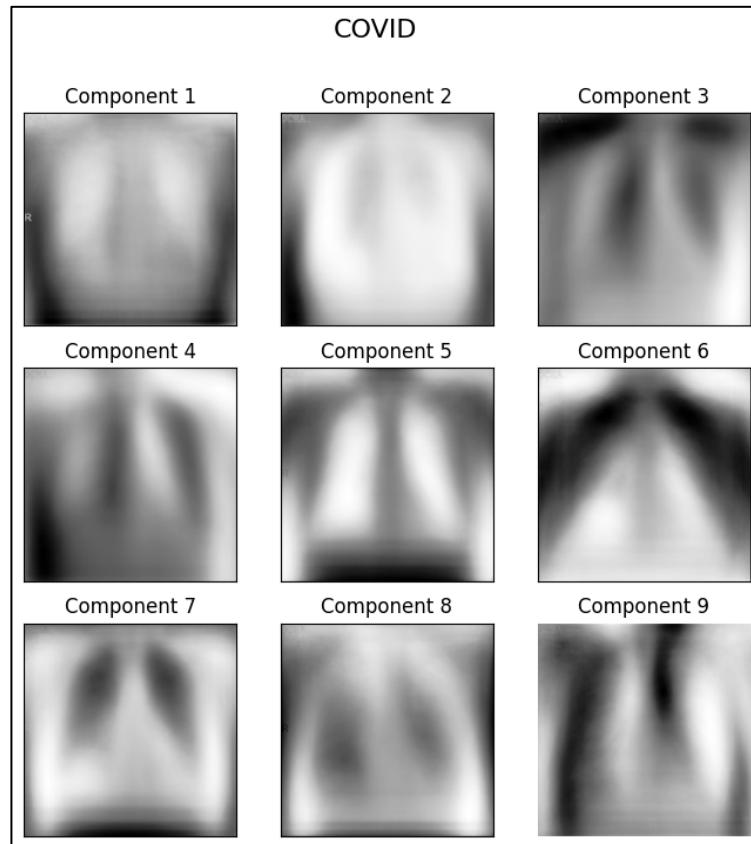


Figure 8 - Eigenimages de la classe COVID

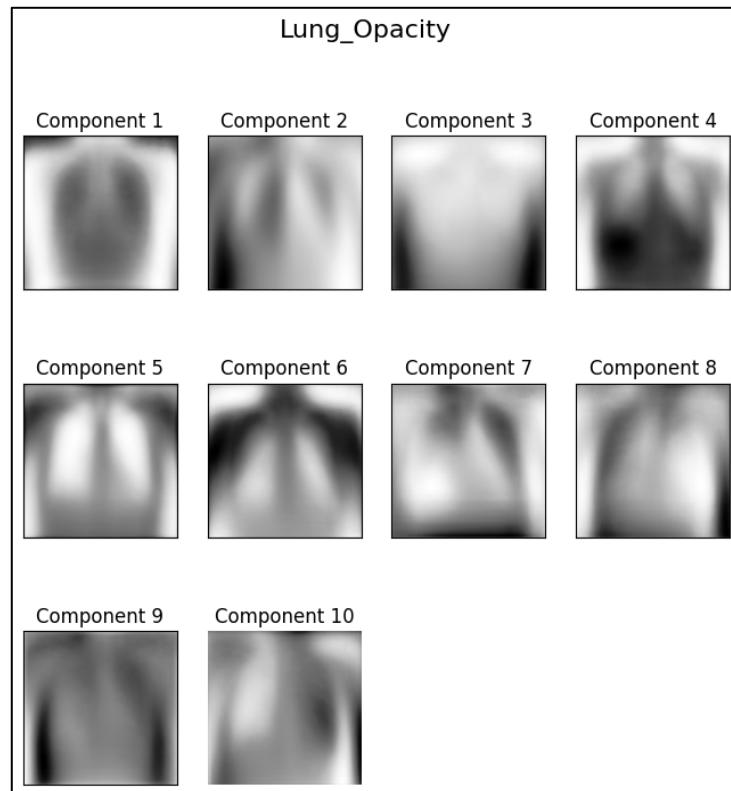


Figure 9 - Eigenimages de la classe Lung Opacity

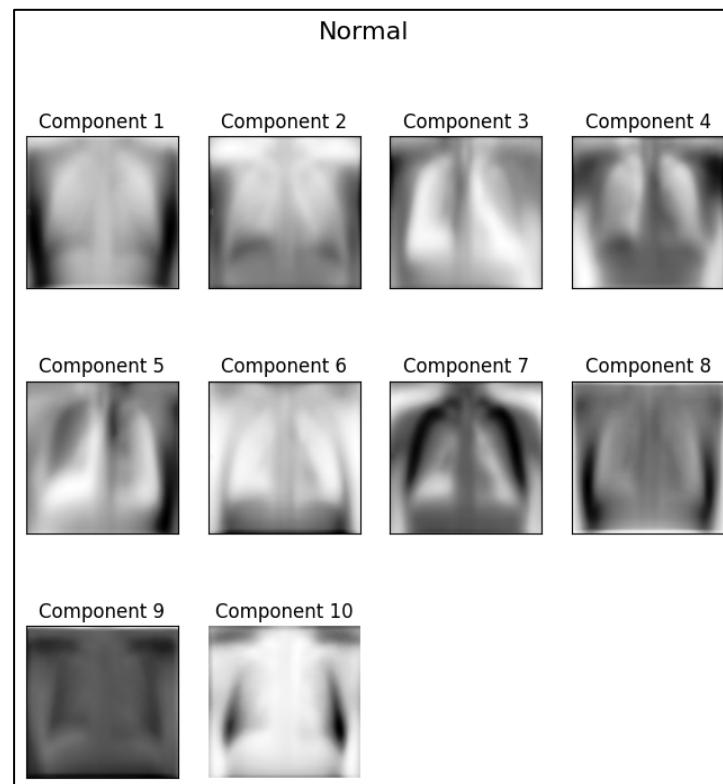


Figure 10 - Eigenimage des images de la classe Normal

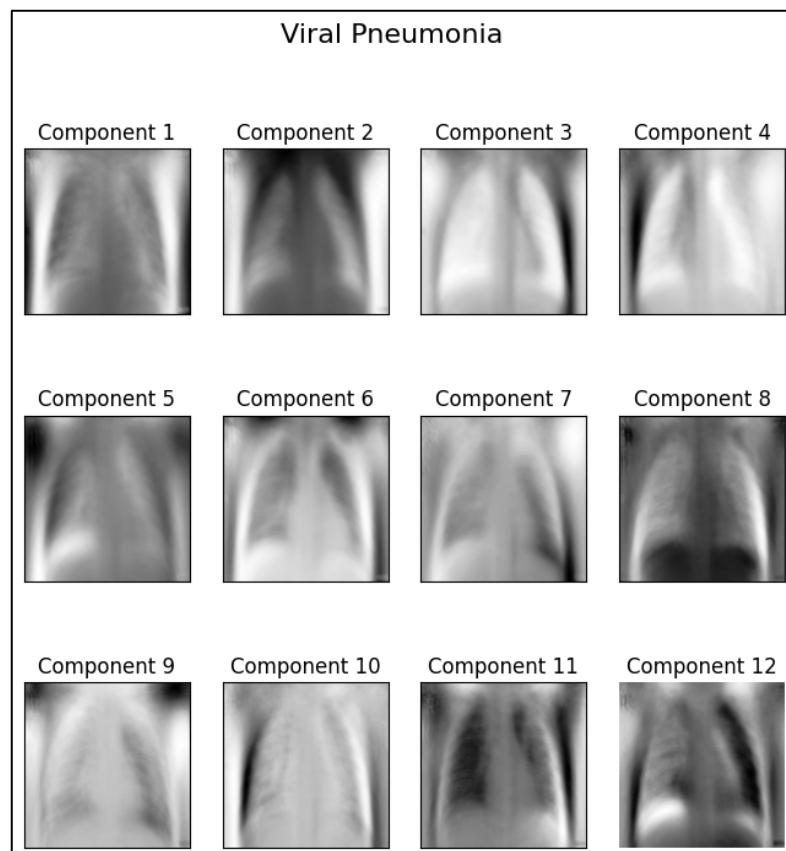


Figure 11 - Eigenimage des images de la classe Viral pneumonia

7. RESULTATS

7.1. ÉTUDE SAIN VS MALADE SANS MASQUE

Le tableau 1 récapitule l'ensemble des résultats pour chaque modèle.

Modèle	Précision	Recall	F1-Score	Commentaires
LeNet5	0.84	0.83	0.83	Référence
EfficientNetB1	0.91	0.91	0.91	Meilleure performance
VGG16	0.91	0.91	0.91	
Xception	0.82	0.82	0.82	
ResNet152	0.92	0.92	0.92	

Tableau 1 - Récapitulatif des performances des modèles sur deux classes sans masque

7.1.1. LeNet5 (modèle de référence)

Notre modèle LeNet5 nous a servi de référence pour comparer les autres modèles. Celui-ci a obtenu une précision de 83%. Le détail des résultats est disponible en Annexe I de ce document.

7.1.2. EfficientNetB1

Notre modèle EfficientNetB1 avec 31 couches dégelées a obtenu un score de 91%. Le détail des résultats est disponible en Annexe II de ce document.

7.1.3. VGG16

Notre modèle VGG16 a obtenu un score de 91%. Le détail des résultats est disponible en Annexe III de ce document.

7.1.4. Xception

Notre modèle Xception gelé a obtenu un score de 82%. Le détail des résultats est disponible en Annexe IV de ce document.

7.1.5. ResNet152

Notre modèle ResNet152 gelé a obtenu un score de 89%. Le détail des résultats est disponible en Annexe V de ce document.

7.2. ÉTUDE SAIN VS MALADE AVEC MASQUES

Le tableau 2 récapitule l'ensemble des résultats pour chaque modèle.

Modèle	Précision	Recall	F1-Score	Commentaires
EfficientNetB1	0.90	0.89	0.89	Meilleure performance
VGG16	0.84	0.83	0.83	
Xception	0.81	0.81	0.81	
ResNet152	0.89	0.89	0.89	Meilleure performance

Tableau 2 - Récapitulatif des performances des modèles sur deux classes avec masque

7.2.1. EfficientNetB1

Notre modèle EfficientNetB1 avec 3 couches dégelées a obtenu un score de 91%. Le détail des résultats est disponible en Annexe VI de ce document.

7.2.2. VGG16

Notre modèle VGG16 gelé a obtenu un score de 87%. Le détail des résultats est disponible en Annexe VII de ce document.

7.2.3. ResNet152

Notre modèle ResNet152 gelé a obtenu un score de 89%. Le détail des résultats est disponible en Annexe VIII de ce document.

7.2.4. Xception

Notre modèle Xception gelé a obtenu un score de 83%. Le détail des résultats est disponible en Annexe IX de ce document.

7.3. ÉTUDE DES QUATRE CLASSES SANS MASQUE

Le tableau 3 récapitule l'ensemble des résultats pour chaque modèle.

Modèle	Précision	Recall	F1-Score	Commentaires
EfficientNetB1	0.92	0.92	0.92	
VGG16	0.93	0.93	0.93	Meilleure performance
Xception	0.81	0.81	0.81	
ResNet152	0.91	0.91	0.91	

Tableau 3 - Récapitulatif des performances des modèles sur quatre classes sans masque

7.3.1. EfficientNetB1

Notre modèle EfficientNetB1 avec 31 couches dégelées a obtenu un score de 92%. Le détail des résultats est disponible en Annexe X de ce document.

7.3.2. VGG16

Notre modèle VGG16 avec 4 couches dégelées a obtenu un score de 93%. Les résultats sont disponibles en Annexe XI de ce document.

7.3.3. Xception

Notre modèle Xception avec 4 couches dégelées a obtenu un score de 81%. Le détail des résultats est disponible en Annexe XII de ce document.

7.3.4. ResNet152

Notre modèle ResNet152 gelé a obtenu un score de 91%. Le détail des résultats est disponible en Annexe XIII de ce document.

7.4. ÉTUDE DES QUATRE CLASSES AVEC MASQUE

Le tableau 4 récapitule l'ensemble des résultats pour chaque modèle.

Modèle	Précision	Recall	F1-Score	Commentaires
EfficientNetB1	0.87	0.86	0.87	Meilleure performance
VGG16	0.84	0.83	0.83	
Xception	0.75	0.76	0.75	
ResNet152	0.82	0.83	0.82	

Tableau 4 - Récapitulatif des performances des modèles sur quatre classes avec masque

7.4.1. EfficientNetB1

Notre modèle EfficientNetB1 avec 31 couches dégelées a obtenu un score de 87%. Le détail des résultats est disponible en Annexe XIV de ce document.

7.4.2. VGG16

Notre modèle VGG16 avec 4 couches dégelées a obtenu un score de 83%. Le détail des résultats est disponible en Annexe XV de ce document.

7.4.3. Xception

Notre modèle Xception gelé a obtenu un score de 76%. Le détail des résultats est disponible en Annexe XVI de ce document.

7.4.4. ResNet152

Notre modèle ResNet152 avec 4 couches dégelées a obtenu un score de 83%. Le détail des résultats est disponible en Annexe XVII de ce document.

8. DISCUSSION

Nous avons comparé différentes réseaux neuronaux convolutifs : VGG16¹⁵, ResNet152¹⁶, Xception¹⁷ et EfficientNet¹⁸. Notre choix de ces modèles est justifié par le fait qu'ils sont issus de la littérature scientifique et sont considérés comme des modèles « *State of the art* » dans le domaine de la vision par ordinateur.

8.1. COMPARAISON DES ETUDES « AVEC MASQUE » ET « SANS MASQUE »

Nos résultats indiquent une performance généralement inférieure des modèles lorsque des masques numériques sont appliqués pour cibler la région pulmonaire. Bien que ces masques soient destinés à isoler les zones d'intérêt et à potentiellement améliorer la classification, nos observations suggèrent que cette technique pourrait en fait réduire la performance des modèles. Une raison possible pourrait être la perte d'information contextuelle qui pourrait aider le modèle à différencier entre les diverses classes. En particulier, dans des conditions telles que la pneumonie virale ou l'opacité pulmonaire, des caractéristiques subtiles en dehors de la région pulmonaire masquée peuvent être cruciales pour une classification précise. Ces résultats soulèvent des questions importantes sur l'efficacité de l'utilisation de masques numériques dans la tâche de classification et suggèrent qu'une exploration plus approfondie est nécessaire pour comprendre leurs avantages et inconvénients.

8.2. FAUX POSITIFS ET FAUX NEGATIFS

Dans le contexte clinique, la précision du diagnostic est cruciale. Un faux positif, c'est-à-dire la classification erronée d'un patient sain comme étant malade, peut conduire à des traitements inutiles et induire un stress psychologique considérable pour le patient. De plus, cela peut engendrer des coûts de santé supplémentaires et contribuer à surcharger un système de soins déjà en tension. D'un autre côté, un faux négatif, où le modèle échoue à identifier une condition pathologique, peut avoir des conséquences encore plus graves. Une mauvaise classification peut retarder le diagnostic et le traitement, compromettant ainsi les chances de guérison ou de gestion efficace de la maladie. Dans notre étude, tous les modèles ont montré une faible proportion de faux positifs et de faux négatifs, mais il est crucial de prendre ces erreurs en compte. Des mesures complémentaires, telles que

des examens de suivi ou des tests diagnostiques supplémentaires, doivent être envisagées pour minimiser les risques associés à ces erreurs de classification.

A noter qu'un faux négatif peut être plus grave qu'un faux positif : une erreur de classification peut avoir des conséquences potentiellement graves, voire mortelles. En particulier, un faux négatif peut entraîner un retard de traitement crucial, ce qui pourrait agraver la condition du patient.

Dans l'étude binaire « Sain vs Malade », le modèle VGG16 a affiché le plus faible taux de faux négatifs, avec seulement 59 cas mal classés parmi les 996 réellement malades. Ce résultat pourrait faire de VGG16 un choix préférable dans un contexte clinique où minimiser les faux négatifs est une priorité. Cependant, il est important de considérer également d'autres métriques de performance et les spécificités de l'application clinique en question.

Une stratégie pourrait consister à utiliser VGG16 comme une sorte de "pré-filtrage" pour minimiser les faux négatifs, tout en utilisant d'autres tests ou modèles pour confirmer les diagnostics. Il est également vital de considérer le coût, tant en termes de temps de calcul que de ressources, pour déployer ces modèles à grande échelle.

8.3. COMPROMIS ENTRE PRECISION ET PUISSANCE DE CALCUL

Le choix de retenir le modèle EfficientNetB1 comme solution optimale pour notre étude s'est fondé sur un équilibre entre précision élevée et efficacité en termes de temps et de puissance de calcul. Dans un environnement clinique tel qu'un hôpital, où des décisions médicales critiques doivent être prises rapidement, la vitesse de calcul n'est pas seulement une commodité, mais une nécessité. EfficientNetB1 offre une précision respectable, en particulier avec un taux de faux négatifs relativement faible, tout en étant nettement moins gourmand en ressources par rapport à des modèles plus lourds comme ResNet152 ou VGG16. Ce compromis est crucial pour un déploiement à grande échelle, où des serveurs hospitaliers peuvent être limités en termes de puissance de calcul et où le temps d'attente pour le diagnostic doit être minimisé. Ainsi, en alliant des performances de haut niveau à une grande efficacité, EfficientNetB1 se présente comme une option viable et robuste pour des applications médicales en temps réel.

8.4. OPTIMISATION DES HYPERPARAMETRES

Pour affiner davantage la performance de notre modèle choisi, EfficientNetB1, nous avons employé Keras Tuner pour l'optimisation des hyperparamètres. Cette démarche a été guidée par un code de recherche aléatoire ('RandomSearch'), ciblant spécifiquement la minimisation de la perte de validation ('val_loss') sur un total de 50 essais ('max_trials'). L'usage de Keras Tuner permet non seulement de systématiser mais aussi de rigouriser la recherche d'une configuration optimale des paramètres du modèle, notamment la taille du lot, le taux d'apprentissage, et les unités de neurones dans les couches denses, entre autres. Cette optimisation automatisée se révèle particulièrement précieuse pour améliorer la robustesse et la fiabilité du modèle en contexte clinique. Elle permet de des améliorations en termes de précision et de réduction des faux négatifs et des faux positifs, ce qui, dans un cadre médical, peut se traduire par des gains significatifs en matière de diagnostic et de gestion des patients.

Cette étape d'optimisation s'inscrit dans une stratégie globale visant à produire un modèle non seulement performant mais aussi efficient, capable de répondre aux contraintes et aux exigences spécifiques du domaine médical. Il est ainsi impératif de souligner l'importance de cette optimisation dans le cycle de développement d'un modèle de deep learning destiné à des applications cliniques en temps réel.

8.5. INTERPRETABILITE DU MODELE

L'interprétabilité des modèles de deep learning est une préoccupation majeure, surtout dans le domaine médical où les décisions basées sur ces modèles peuvent avoir des conséquences directes sur la santé des patients. Afin d'offrir une meilleure compréhension des décisions prises par notre modèle EfficientNetB1, nous avons fait appel à la technique Grad-CAM (Gradient-weighted Class Activation Mapping). Cette méthode permet de visualiser les zones de l'image qui ont le plus contribué à la décision du modèle, en superposant une carte thermique sur l'image originale.

Utiliser Grad-CAM a plusieurs avantages. Premièrement, il offre aux cliniciens une perspective supplémentaire pour évaluer la fiabilité du diagnostic posé par le modèle. Par exemple, si la carte thermique se focalise sur la région pulmonaire lors de la classification d'une image comme indicative de pneumonie, cela peut rassurer le professionnel de santé sur la pertinence du diagnostic. Deuxièmement, ces cartes

d'activation peuvent servir à identifier d'éventuelles faiblesses dans le modèle. Si la carte thermique indique que le modèle se concentre sur des zones non pertinentes, cela peut signaler un besoin de recalibration ou de réentraînement du modèle.

Dans le cadre de notre étude, l'implémentation de Grad-CAM a révélé une focalisation cohérente et médicalement pertinente des zones d'intérêt, renforçant ainsi la crédibilité de nos résultats (Figure X). Cette approche augmente non seulement la confiance dans les prédictions du modèle, mais elle offre également un moyen pour les cliniciens de mieux comprendre le fonctionnement interne du modèle, ce qui est crucial pour son adoption dans des environnements médicaux où l'erreur est peu tolérable.

Prédictions correctes

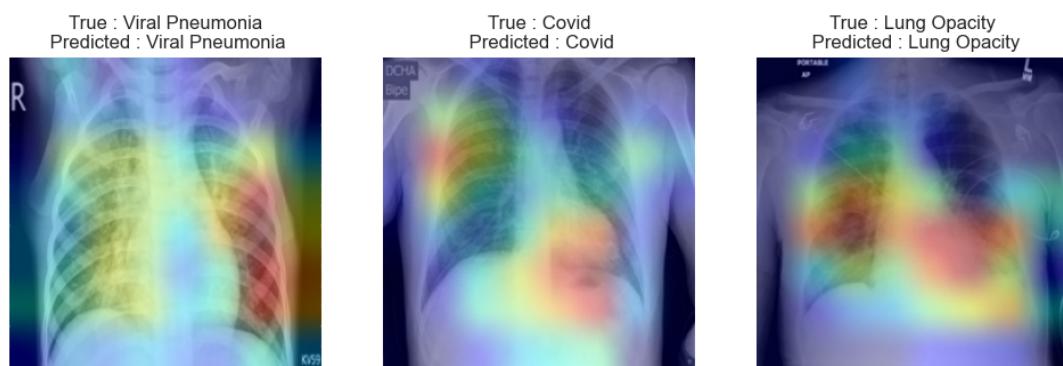


Figure 12 - GradCam de prédictions correctes

Mauvaises prédictions



Figure 13 - GradCam de prédictions incorrectes

8.6. AXES D'AMÉLIORATION

Bien que notre étude ait apporté des éclairages significatifs sur l'efficacité des différents modèles de réseaux neuronaux convolutifs dans la classification des images médicales, plusieurs axes d'amélioration demeurent. Premièrement, la robustesse du

modèle face à des images de mauvaise qualité n'a pas été testée. Dans un environnement clinique, il n'est pas rare de se trouver confronté à des images de faible résolution ou à des images altérées par du bruit ; ainsi, une étude ultérieure pourrait aborder ce point pour évaluer la performance du modèle dans ces conditions.

Deuxièmement, nous avons identifié tardivement que les couches "Dense" utilisées dans la comparaison initiale des modèles n'étaient pas uniformes. Ceci pourrait avoir introduit un biais dans nos résultats et mérite d'être corrigé dans des recherches futures.

L'ensemble de ces axes d'amélioration offre un potentiel considérable pour des travaux de recherche futurs, visant à maximiser à la fois la précision et la robustesse des modèles de classification en imagerie médicale.

9. CONCLUSION

Au terme de cette étude, nous avons réussi à comparer divers modèles de réseaux neuronaux convolutifs — VGG16, ResNet152, Xception et EfficientNetB1 — pour la classification de conditions médicales à partir d'images radiologiques. Les modèles ont été évalués en tenant compte de plusieurs critères essentiels en milieu clinique : la précision du diagnostic, le taux de faux positifs et de faux négatifs, ainsi que l'efficacité en termes de temps et de puissance de calcul.

Nous avons observé que l'application de masques numériques pour cibler la région pulmonaire pourrait avoir des effets négatifs sur la performance du modèle, soulevant des questions sur l'efficacité de cette technique. Notre analyse a aussi mis en lumière l'importance clinique des faux positifs et des faux négatifs, et nous avons proposé des stratégies pour minimiser ces erreurs.

Notre choix de retenir EfficientNetB1 comme modèle optimal s'est basé sur son équilibre entre une haute précision de classification et une grande efficacité en termes de ressources, ce qui est crucial dans un environnement clinique à forte contrainte temporelle.

Néanmoins, cette étude n'est pas exempte de limitations. La robustesse du modèle aux images de faible qualité n'a pas été testée, et des incohérences dans les couches denses entre les modèles ont été identifiées tardivement.

Dans l'ensemble, les résultats obtenus sont prometteurs et suggèrent que l'application de techniques d'apprentissage en profondeur dans le domaine médical est non seulement viable mais aussi extrêmement bénéfique. Notre modèle obtient de meilleures performances sur un support moins efficace dans la détection de la COVID-19, la radiographie, que celui de Yang et al. (2020) entraîné sur des CT-scans. La voie est donc ouverte pour des recherches futures visant à affiner ces méthodes et à étendre leur applicabilité à d'autres tâches médicales complexes.

REFERENCES

- ¹ Yesudhas D, Srivastava A, Gromiha MM. COVID-19 outbreak: history, mechanism, transmission, structural studies and therapeutics. *Infection*. 2021 Apr;49(2):199-213. doi: 10.1007/s15010-020-01516-2. Epub 2020 Sep 4. PMID: 32886331; PMCID: PMC7472674.
- ² Böger B, Fachi MM, Vilhena RO, Cobre AF, Tonin FS, Pontarolo R. Systematic review with meta-analysis of the accuracy of diagnostic tests for COVID-19. *Am J Infect Control*. 2021 Jan;49(1):21-29. doi: 10.1016/j.ajic.2020.07.011. Epub 2020 Jul 10. PMID: 32659413; PMCID: PMC7350782.
- ³ Aljondi R, Alghamdi S. Diagnostic Value of Imaging Modalities for COVID-19: Scoping Review. *J Med Internet Res*. 2020 Aug 19;22(8):e19673. doi: 10.2196/19673. PMID: 32716893; PMCID: PMC7468642.
- ⁴ Borakati A, Perera A, Johnson J, Sood T. Diagnostic accuracy of X-ray versus CT in COVID-19: a propensity-matched database study. *BMJ Open*. 2020 Nov 6;10(11):e042946. doi: 10.1136/bmjopen-2020-042946. PMID: 33158840; PMCID: PMC7650091.
- ⁵ Chollet, F. (2021). Deep learning with python. Manning Publications.
- ⁶ Ghaderzadeh M, Aria M, Asadi F. X-Ray Equipped with Artificial Intelligence: Changing the COVID-19 Diagnostic Paradigm during the Pandemic. *Biomed Res Int*. 2021 Aug 22;2021:9942873. doi: 10.1155/2021/9942873. PMID: 34458373; PMCID: PMC8390162.
- ⁷ Huang S, Yang J, Fong S, Zhao Q. Artificial intelligence in the diagnosis of COVID-19: challenges and perspectives. *Int J Biol Sci*. 2021 Apr 10;17(6):1581-1587. doi: 10.7150/ijbs.58855. PMID: 33907522; PMCID: PMC8071762.
- ⁸ Yang Y, Lure FYM, Miao H, Zhang Z, Jaeger S, Liu J, Guo L. Using artificial intelligence to assist radiologists in distinguishing COVID-19 from other pulmonary infections. *J Xray Sci Technol*. 2021;29(1):1-17. doi: 10.3233/XST-200735. PMID: 33164982; PMCID: PMC7990455.
- ⁹ Mohammad-Rahimi H, Nadimi M, Ghalyanchi-Langeroudi A, Taheri M, Ghafouri-Fard S. Application of Machine Learning in Diagnosis of COVID-19 Through X-Ray and CT Images: A Scoping Review. *Front Cardiovasc Med*. 2021 Mar 25;8:638011. doi: 10.3389/fcvm.2021.638011. PMID: 33842563; PMCID: PMC8027078.
- ¹⁰ Bhatele KR, Jha A, Tiwari D, Bhatele M, Sharma S, Mithora MR, Singhal S. COVID-19 Detection: A Systematic Review of Machine and Deep Learning-Based Approaches Utilizing Chest X-Rays and CT Scans. *Cognit Comput*. 2022 Dec 29:1-38. doi: 10.1007/s12559-022-10076-6. Epub ahead of print. PMID: 36593991; PMCID: PMC9797382.
- ¹¹ Yang, X., He, X., Zhao, J., Zhang, Y., Zhang, S., & Xie, P. (2020). COVID-CT-Dataset: A CT Scan Dataset about COVID-19 (Version 3). arXiv. <https://doi.org/10.48550/ARXIV.2003.13865>
- ¹² Martínez Chamorro E, Díez Tascón A, Ibáñez Sanz L, Ossaba Vélez S, Borruel Nacenta S. Radiologic diagnosis of patients with COVID-19. *Radiologí'a*. 2021 January-February;63(1):56–73. doi: 10.1016/j.rxeng.2020.11.001. Epub 2021 Jan 8. PMCID: PMC7791314.

- ¹³ National cost center CCRs for radiology, CT scan and MRI (2018). American College of Radiology. <https://www.acr.org/-/media/ACR/NOINDEX/Advocacy/Advocacy-News/ACR-CY-2019-IPPS-Proposed-Rule-Comment-Letter-Final-6-25-18.pdf?la=en>. (n.d.).
- ¹⁴ COVID-19 Radiography Database. (2022, 19 mars) Kaggle. Disponible à l'adresse : <https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database>
- ¹⁵ Simonyan, K. (2014, 4 septembre). Very deep convolutional networks for Large-Scale image recognition. arXiv.org. <https://arxiv.org/abs/1409.1556>
- ¹⁶ He, K. (2015, 10 décembre). Deep residual learning for image recognition. arXiv.org. <https://arxiv.org/abs/1512.03385>
- ¹⁷ Chollet, F. (2016, 7 octobre). Xception : Deep learning with depthwise separable convolutions. arXiv.org. <https://arxiv.org/abs/1610.02357>
- ¹⁸ Tan, M. (2019, 28 mai). EfficientNet : Rethinking model scaling for convolutional neural networks. arXiv.org. <https://arxiv.org/abs/1905.11946>

LISTE DES TABLEAUX

Tableau I : Exemple de tableau.....**Erreur ! Signet non défini.**

LISTE DES FIGURES

Figure 1 - Provenance des données	8
Figure 2 - Exemple de superposition de masque	11
Figure 3 - Nombre de radiographies dans chaque classe.....	12
Figure 4 - Générateur d'images.....	13
Figure 5 - Moyenne des images de chaque classe	14
Figure 6 - Contraste entre l'image moyenne COVID et celles des autres classes.....	15
Figure 7 - Contraste entre l'image moyenne "Normal" et celles des autres classes.....	15
Figure 8 - Eigenimages de la classe COVID	16
Figure 9 - Eigenimages de la classe Lung Opacity.....	16
Figure 10 - Eigenimage des images de la classe Normal	17
Figure 11 - Eigenimage des images de la classe Viral pneumonia.....	17

LISTE DES ABREVIATIONS ET DES SYMBOLES

ENetB1 : EfficientNetB1

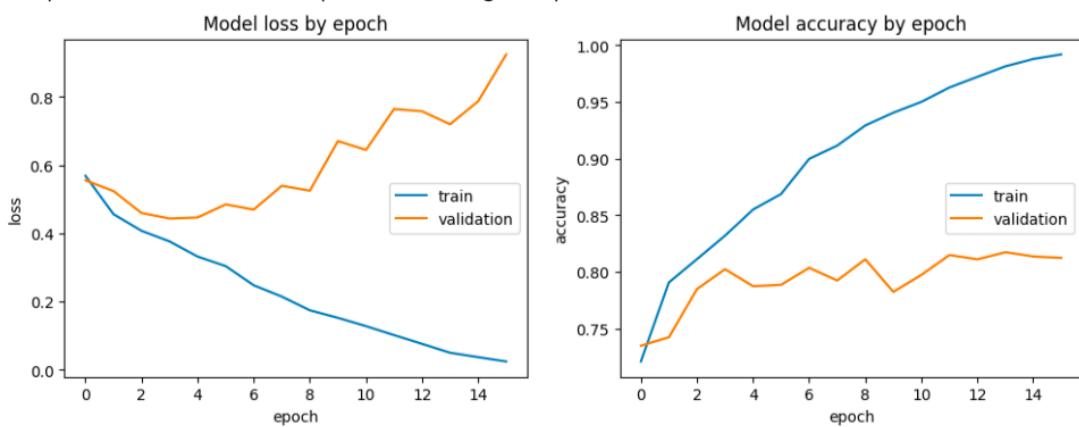
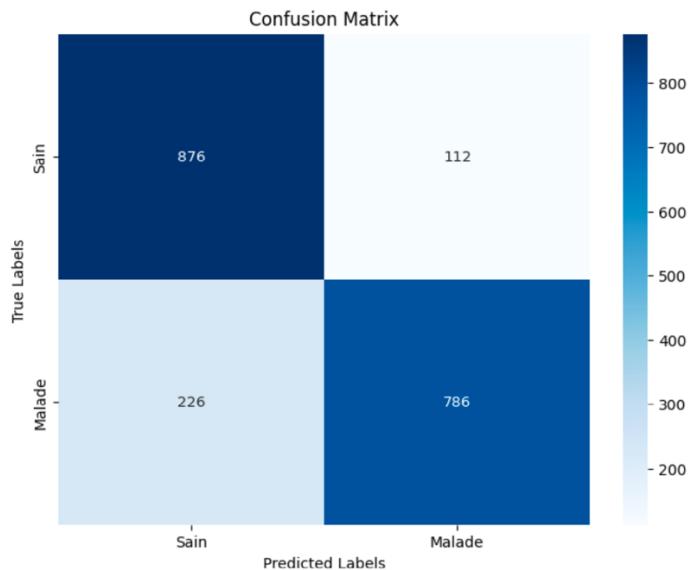
PCA : Analyse des composants principaux

RGB : Red-Green-Blue

LISTE DES ANNEXES

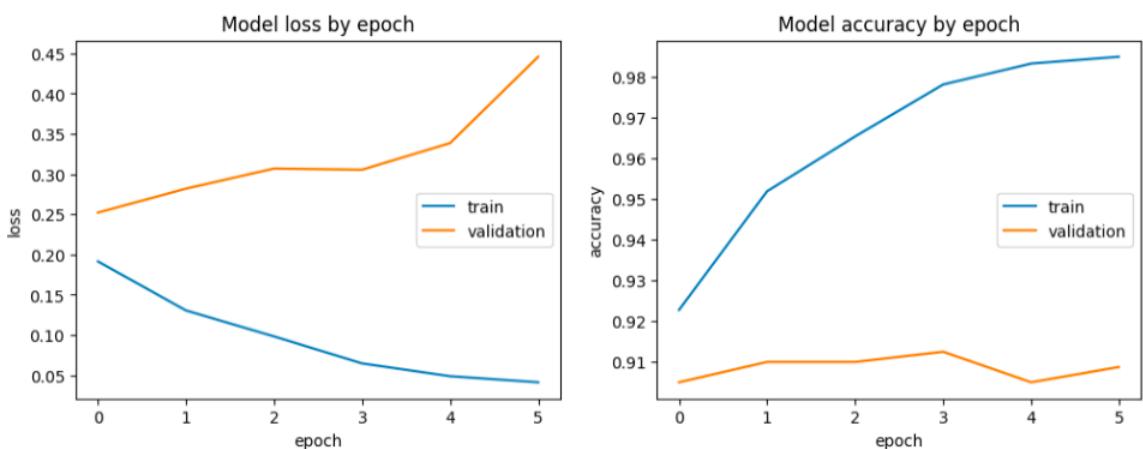
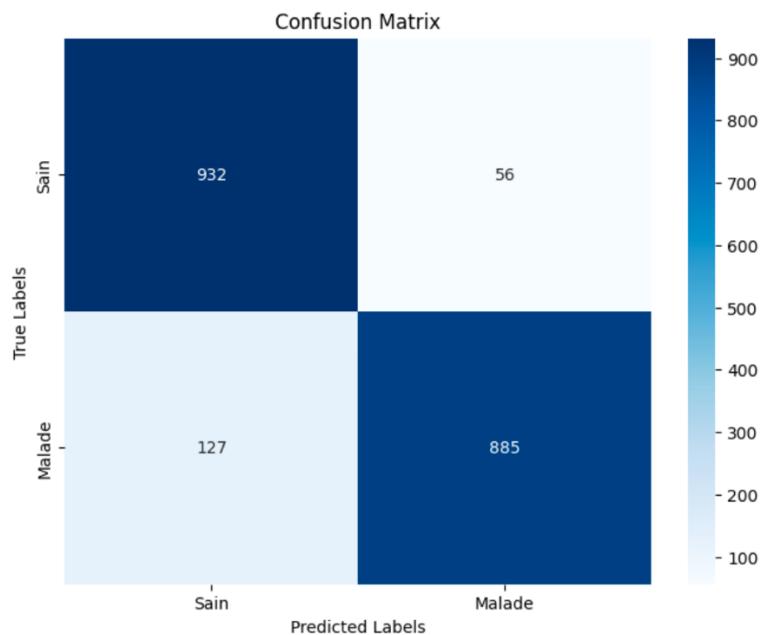
ANNEXE I : RESULTATS DU LENET5 SUR LA CLASSIFICATION BINAIRE SANS MASQUE

	Precision	Recall	F1-score	Support
0	0.79	0.89	0.84	988
1	0.88	0.78	0.82	1012
Accuracy			0.83	2000
Macro avg	0.84	0.83	0.83	2000
Weighted avg	0.84	0.83	0.83	2000



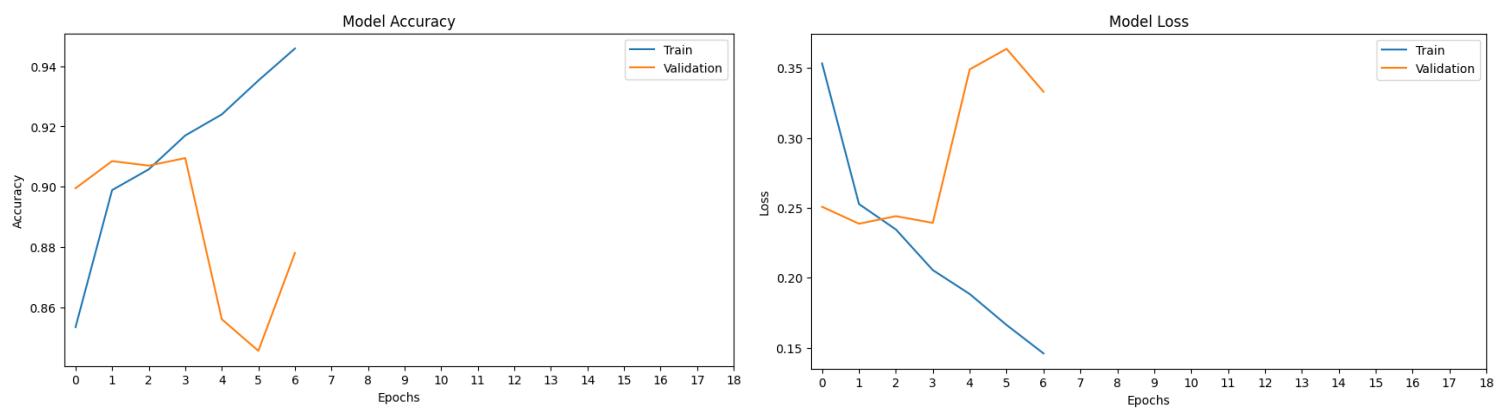
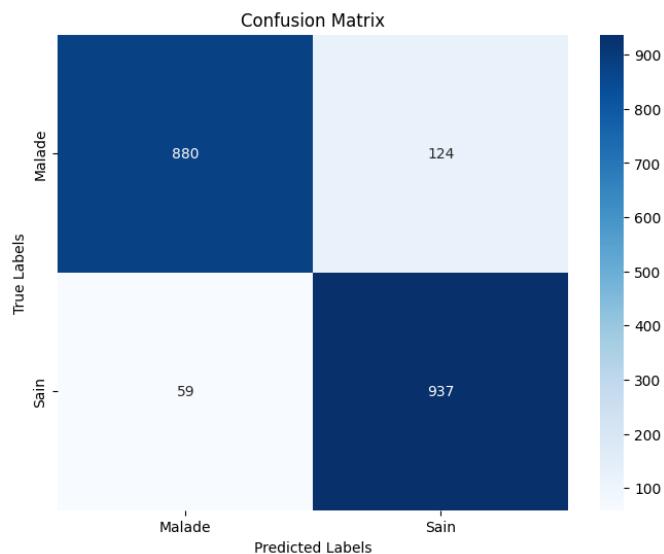
ANNEXE II : RESULTATS DU EFFICIENTNETB1 SUR LA CLASSIFICATION BINAIRE SANS MASQUE

	Precision	Recall	F1-score	Support
0	0.88	0.94	0.91	988
1	0.94	0.88	0.91	1012
Accuracy			0.91	2000
Macro avg	0.91	0.91	0.91	2000
Weighted avg	0.91	0.91	0.91	2000



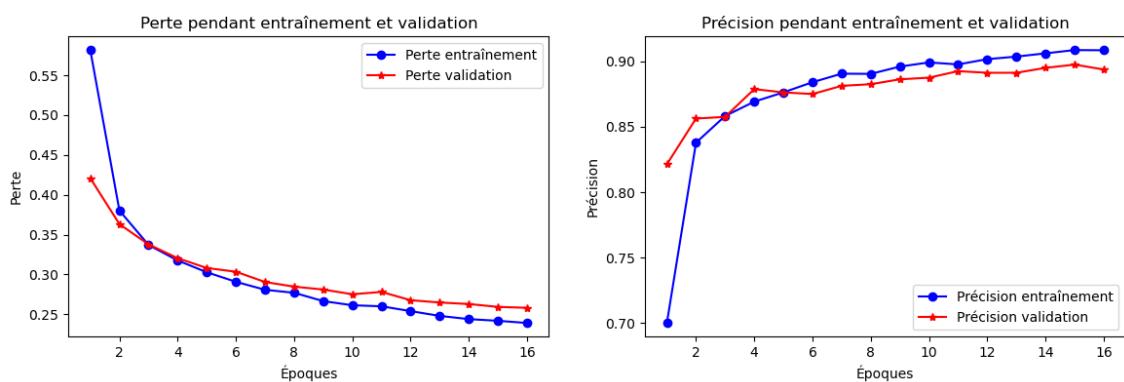
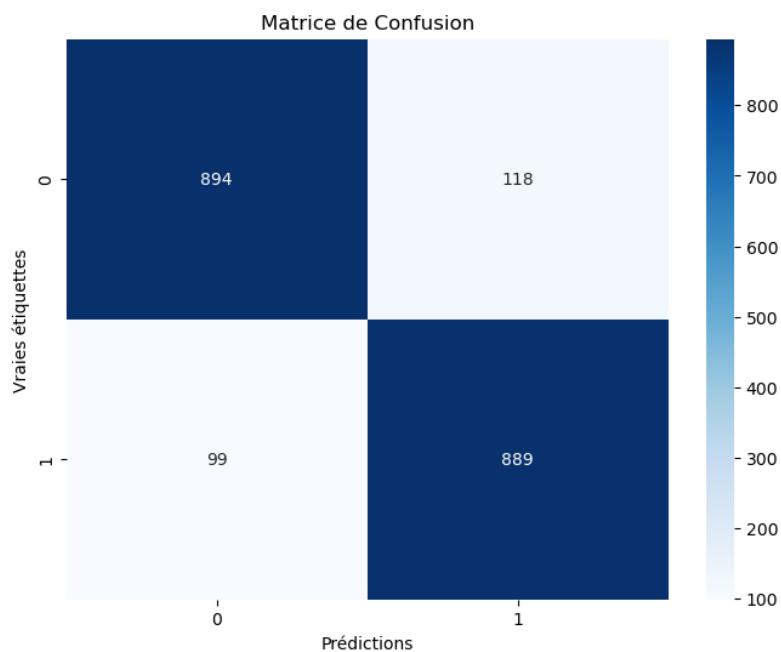
ANNEXE III : RESULTATS DU VGG16 SUR LA CLASSIFICATION BINAIRE SANS MASQUE

	Precision	Recall	F1-score	Support
0	0.94	0.88	0.91	1004
1	0.88	0.94	0.91	996
Accuracy			0.91	2000
Macro avg	0.91	0.91	0.91	2000
Weighted avg	0.91	0.91	0.91	2000



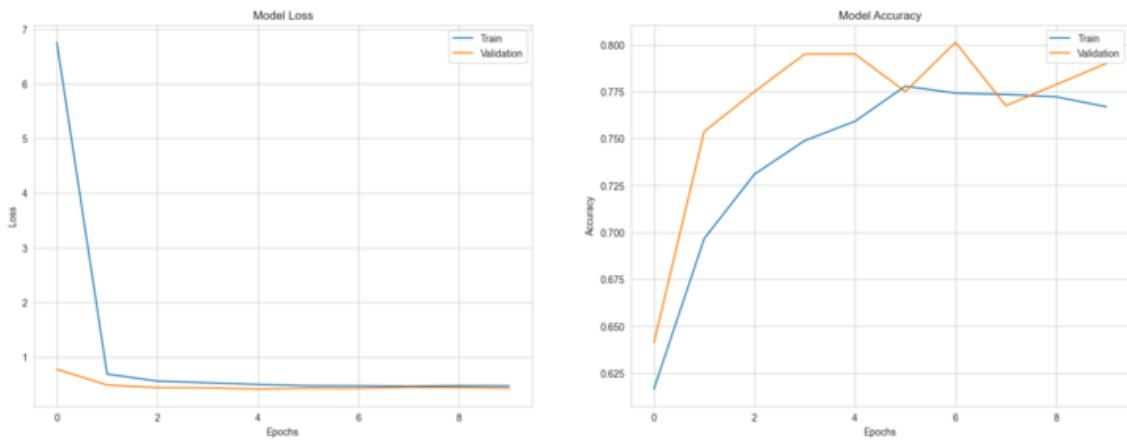
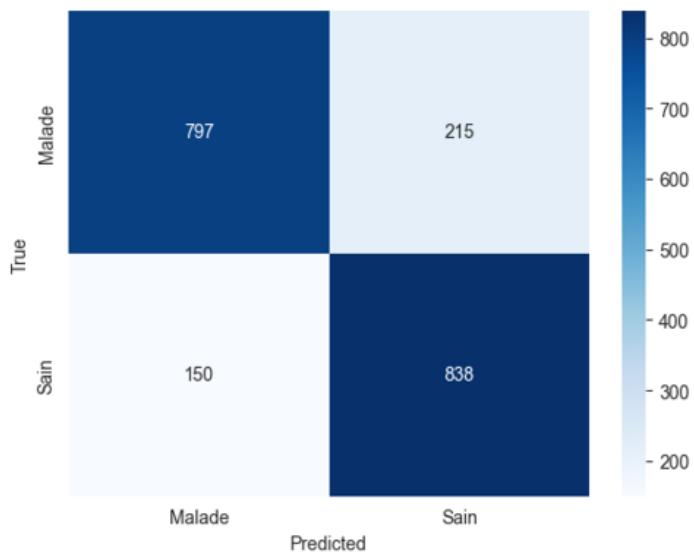
ANNEXE IV : RESULTATS DU RESNET152 SUR LA CLASSIFICATION BINAIRE SANS MASQUE

	Precision	Recall	F1-score	Support
0	0.91	0.93	0.92	1012
1	0.93	0.90	0.92	988
Accuracy			0.92	2000
Macro avg	0.92	0.92	0.92	2000
Weighted avg	0.92	0.92	0.92	2000



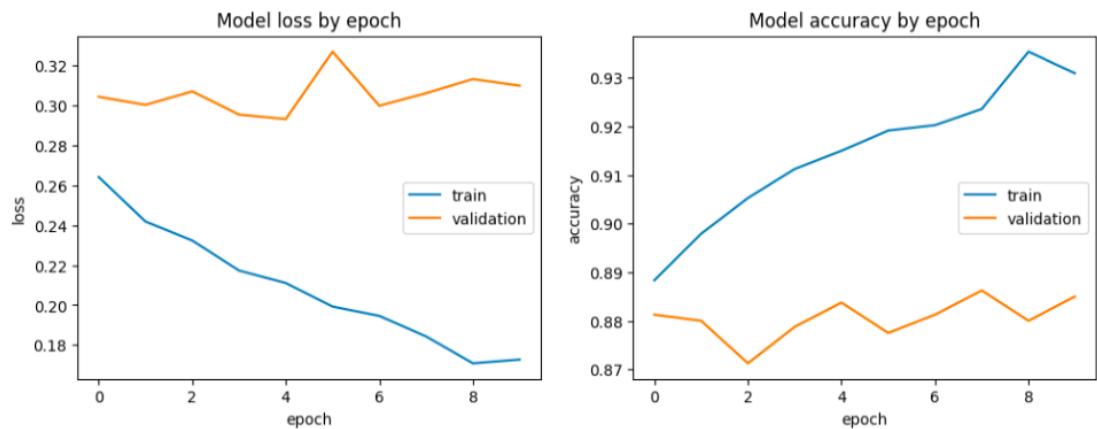
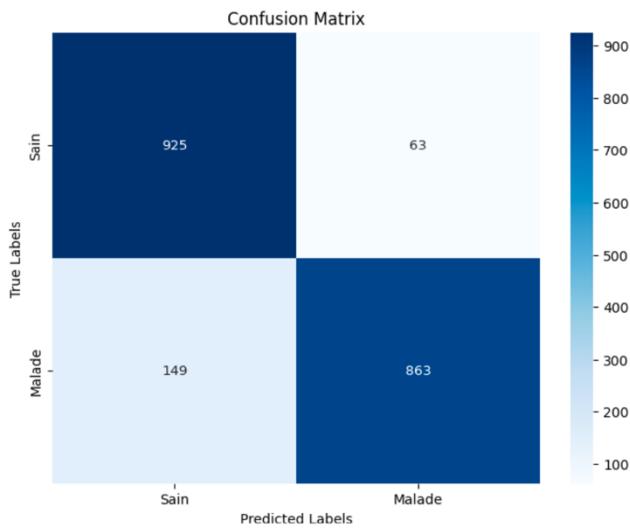
ANNEXE V : RESULTATS DU XCEPTION SUR LA CLASSIFICATION BINAIRE SANS MASQUE

	Precision	Recall	F1-score	Support
0	0.84	0.79	0.81	1012
1	0.80	0.85	0.82	988
Accuracy			0.82	2000
Macro avg	0.82	0.82	0.82	2000
Weighted avg	0.82	0.82	0.82	2000



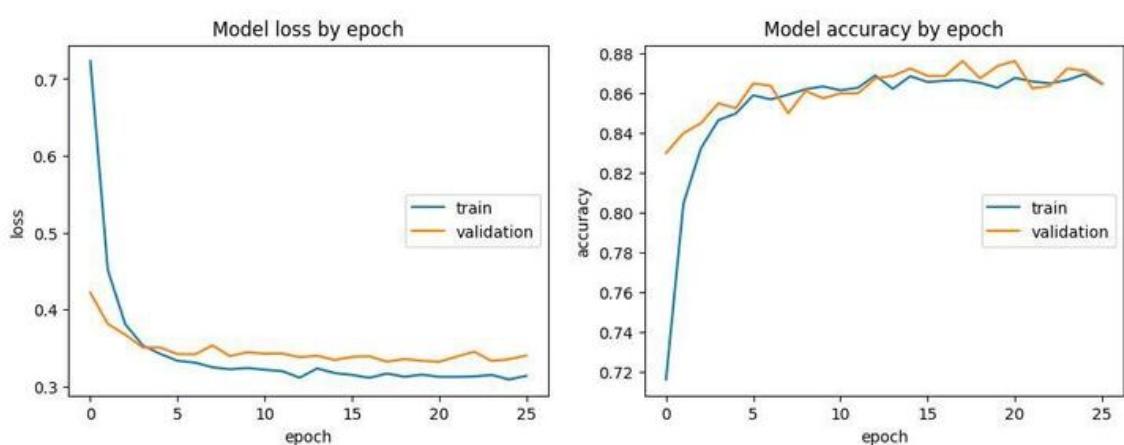
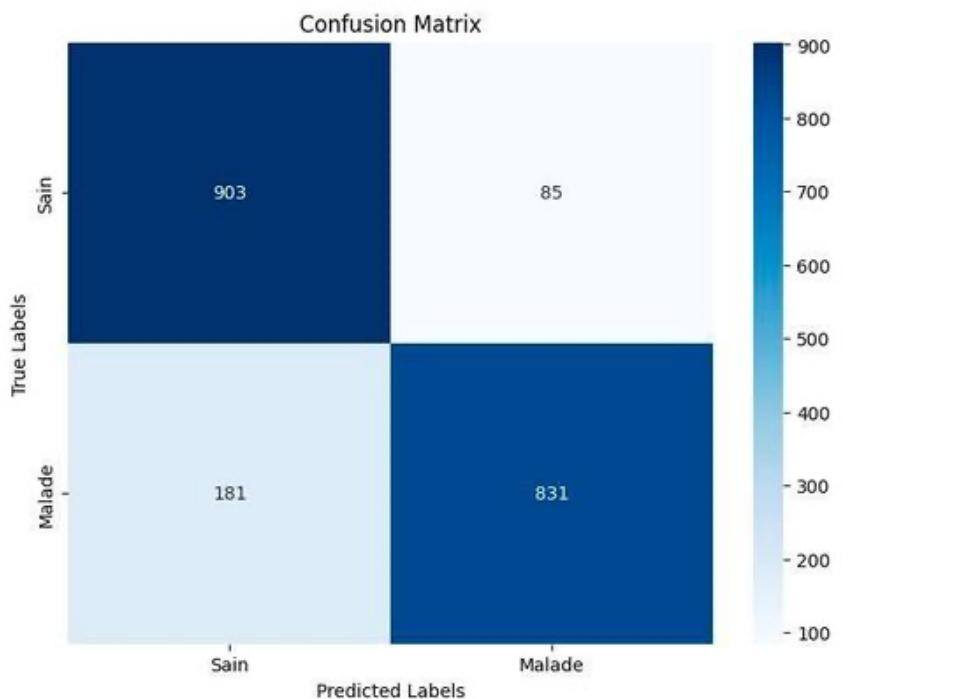
ANNEXE VI : RESULTATS DU EFFICIENTNETB1 SUR LA CLASSIFICATION BINAIRE AVEC MASQUE

	Precision	Recall	F1-score	Support
0	0.86	0.94	0.90	988
1	0.93	0.85	0.89	1012
Accuracy			0.89	2000
Macro avg	0.90	0.89	0.89	2000
Weighted avg	0.90	0.89	0.89	2000



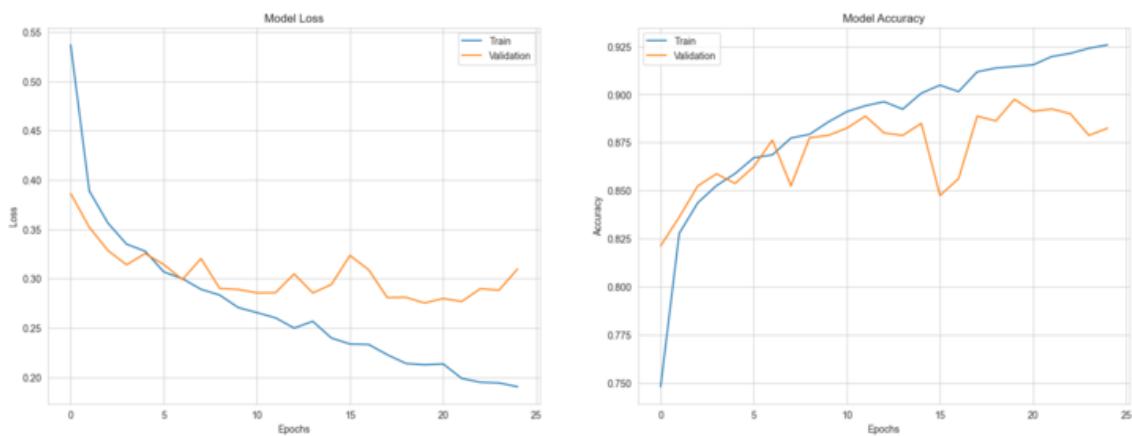
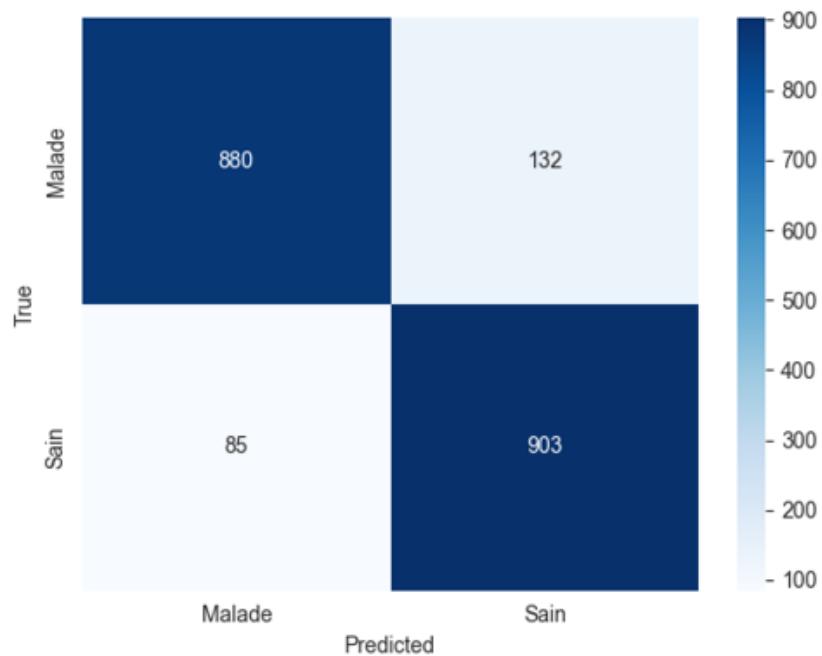
ANNEXE VII : RESULTATS DU VGG16 SUR LA CLASSIFICATION BINAIRE AVEC MASQUE

	Precision	Recall	F1-score	Support
0	0.83	0.91	0.87	988
1	0.91	0.82	0.86	1012
Accuracy			0.87	2000
Macro avg	0.87	0.87	0.87	2000
Weighted avg	0.87	0.87	0.87	2000



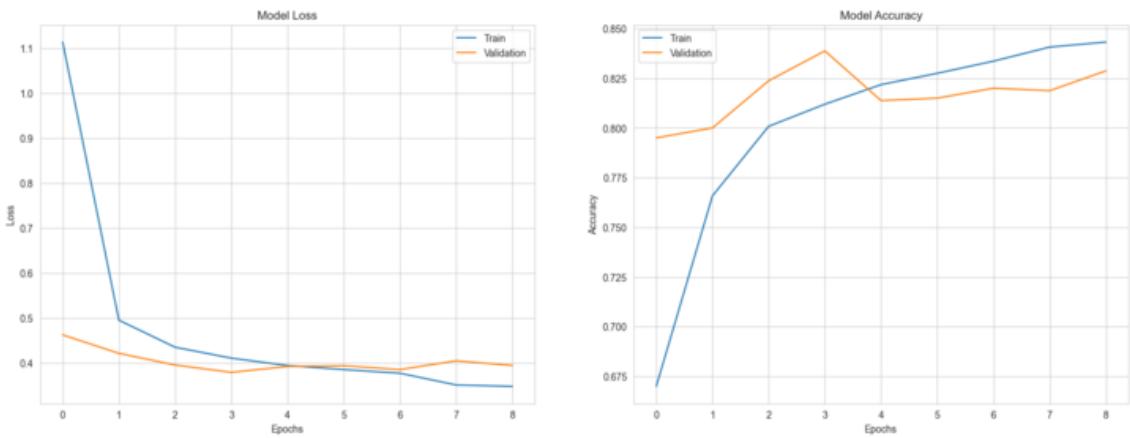
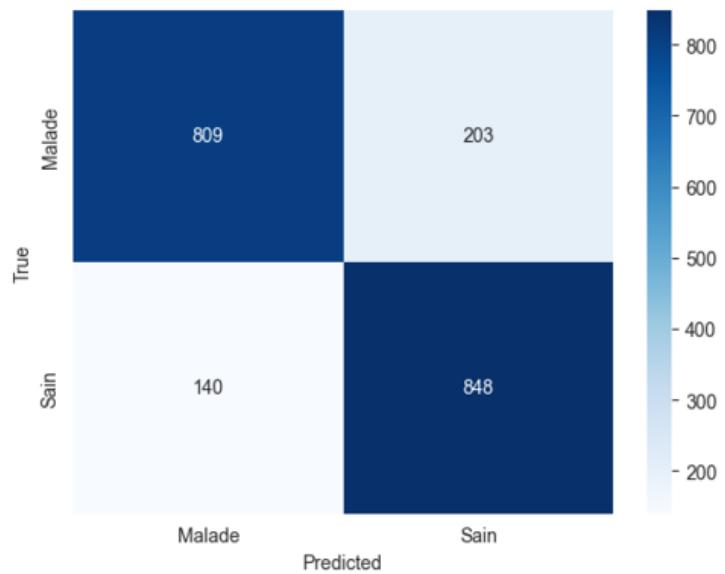
ANNEXE VII : RESULTATS DU RESNET152 SUR LA CLASSIFICATION BINAIRE AVEC MASQUE

	Precision	Recall	F1-score	Support
0	0.91	0.87	0.89	1012
1	0.87	0.91	0.89	1012
Accuracy			0.89	2000
Macro avg	0.89	0.89	0.89	2000
Weighted avg	0.89	0.89	0.89	2000



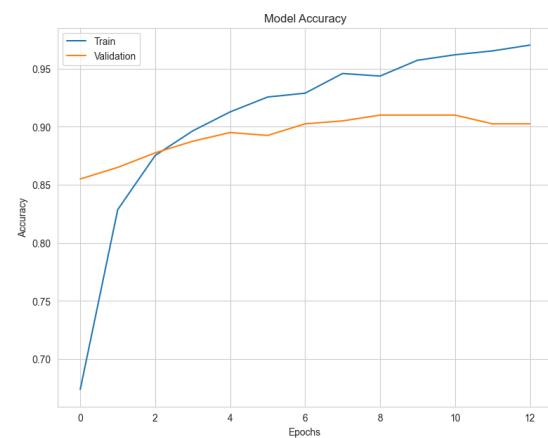
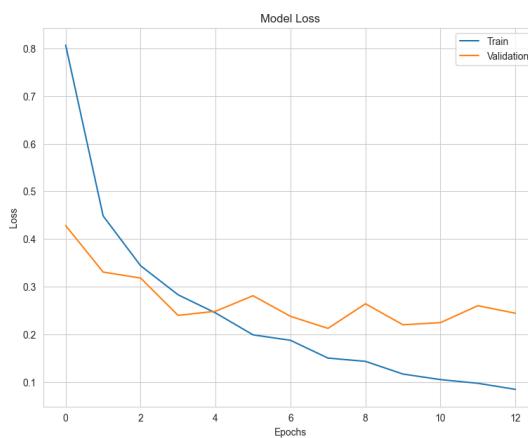
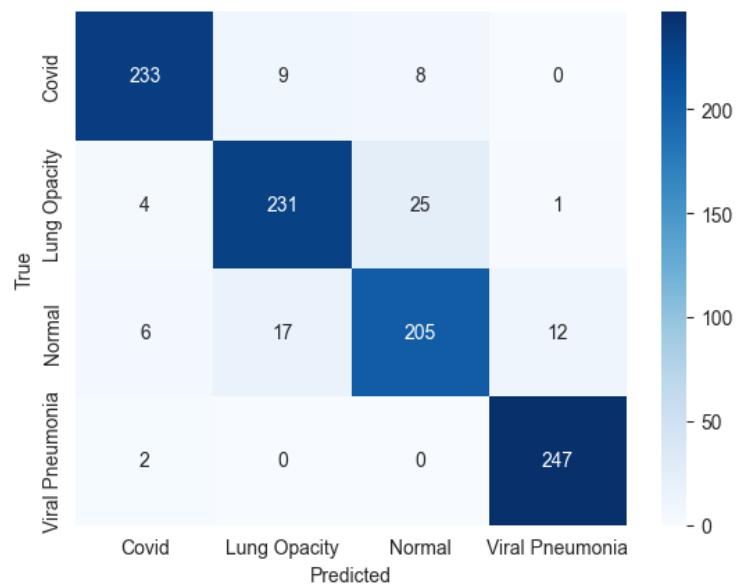
ANNEXE IX : RESULTATS DU XCEPTION SUR LA CLASSIFICATION BINAIRE AVEC MASQUE

	Precision	Recall	F1-score	Support
0	0.85	0.80	0.83	1012
1	0.81	0.86	0.83	988
Accuracy			0.83	2000
Macro avg	0.83	0.83	0.83	2000
Weighted avg	0.83	0.83	0.83	2000



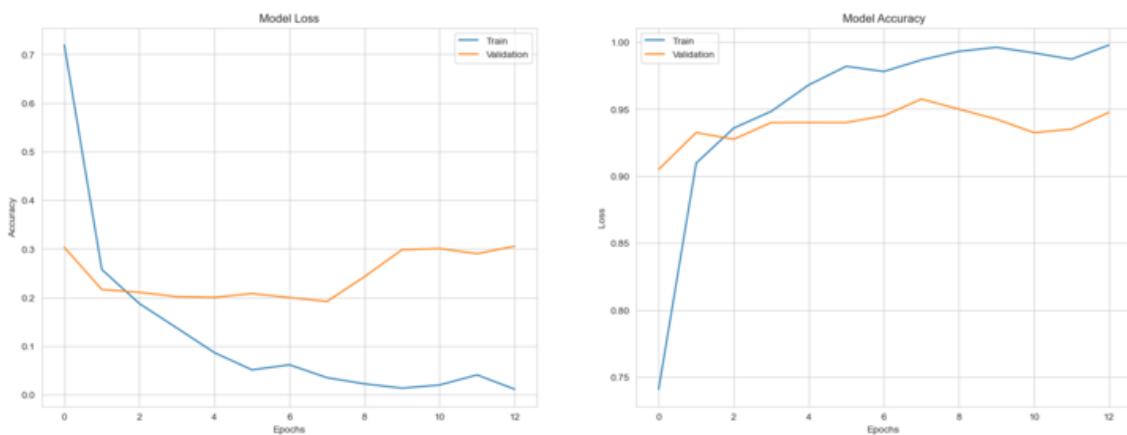
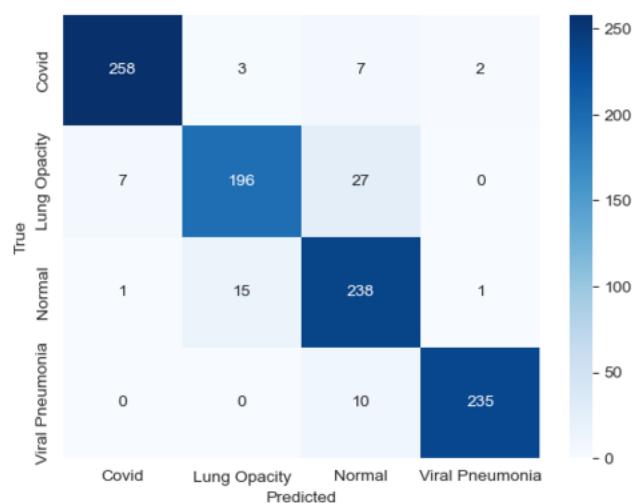
ANNEXE X : RESULTATS DU EFFICIENTNETB1 SUR LA CLASSIFICATION MULTI-CLASSE SANS MASQUE

	Precision	Recall	F1-score	Support
0	0.95	0.93	0.94	250
1	0.90	0.89	0.89	261
2	0.86	0.85	0.86	240
3	0.95	0.99	0.97	249
Accuracy			0.92	1000
Macro avg	0.92	0.92	0.92	1000
Weighted avg	0.92	0.92	0.92	1000



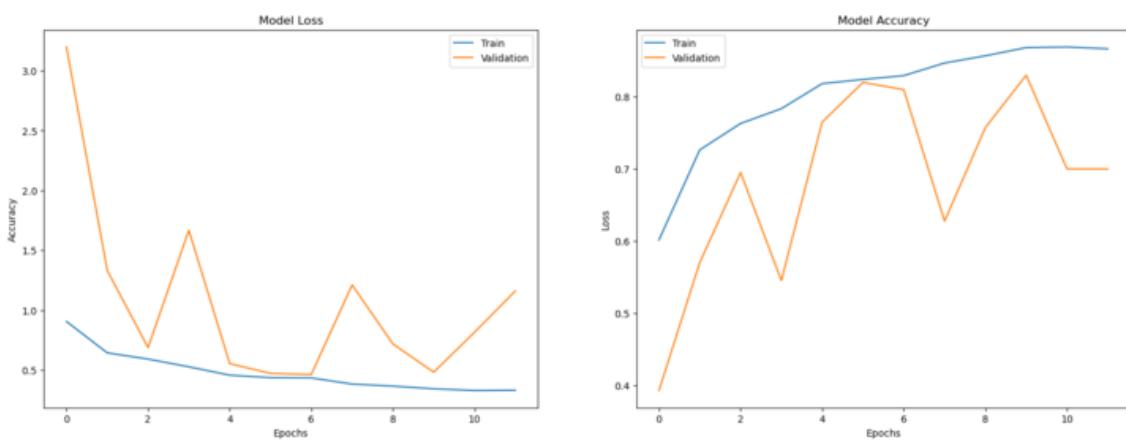
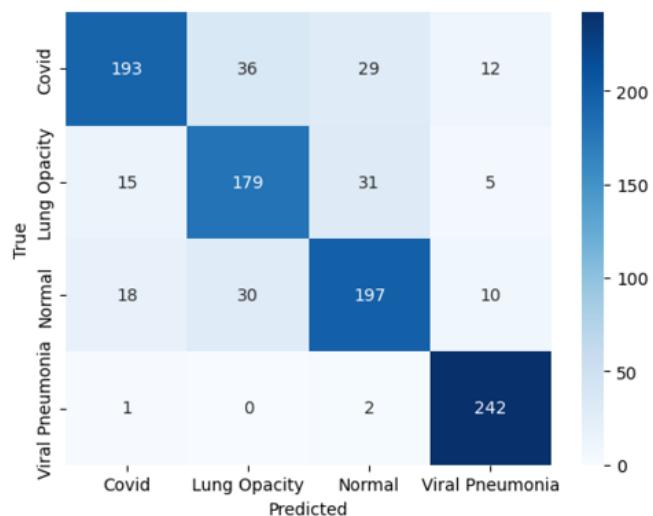
ANNEXE XI : RESULTATS DU VGG16 SUR LA CLASSIFICATION MULTI-CLASSE SANS MASQUE

	Precision	Recall	F1-score	Support
0	0.97	0.96	0.96	270
1	0.92	0.85	0.88	230
2	0.84	0.93	0.89	255
3	0.99	0.96	0.97	245
Accuracy			0.93	1000
Macro avg	0.93	0.93	0.93	1000
Weighted avg	0.93	0.93	0.93	1000



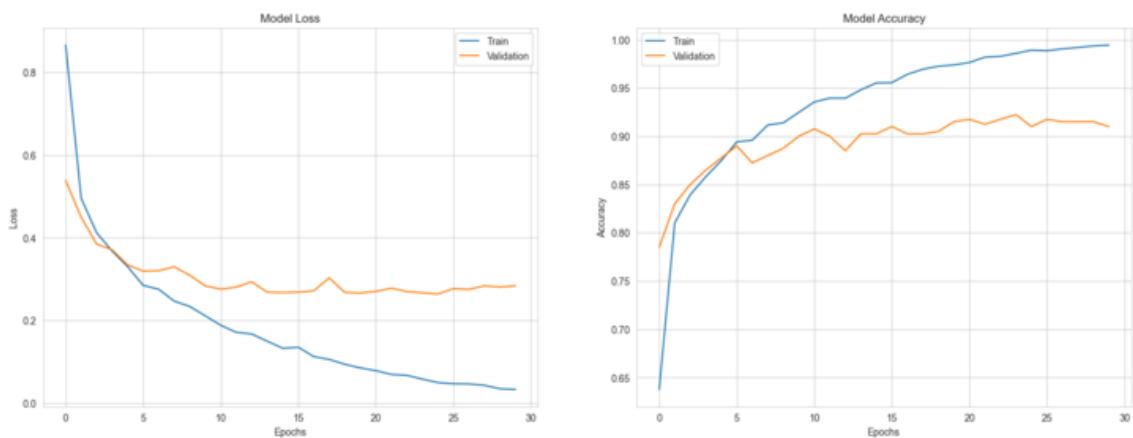
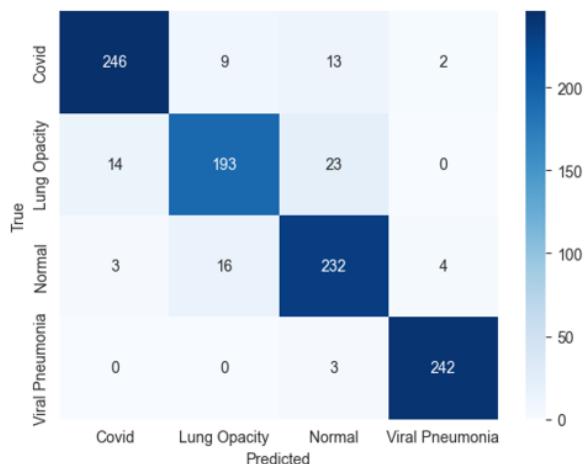
ANNEXE XII : RESULTATS DU XCEPTION SUR LA CLASSIFICATION MULTI-CLASSE SANS MASQUE

	Precision	Recall	F1-score	Support
0	0.85	0.71	0.78	270
1	0.73	0.78	0.75	230
2	0.76	0.77	0.77	255
3	0.90	0.99	0.94	245
Accuracy			0.81	1000
Macro avg	0.81	0.81	0.81	1000
Weighted avg	0.81	0.81	0.81	1000



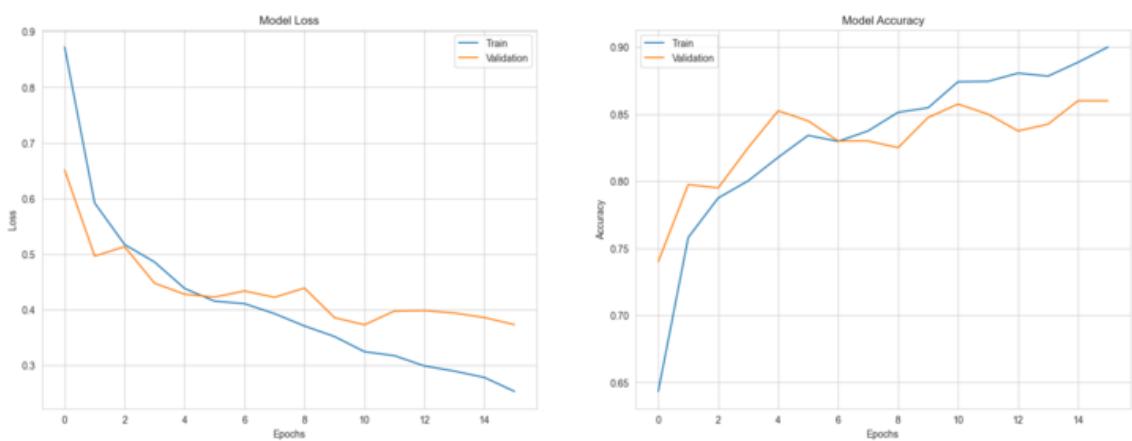
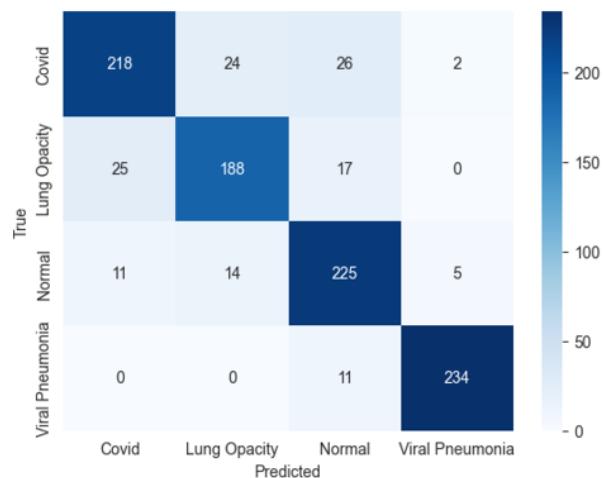
ANNEXE XIII : RESULTATS DU RESNET152 SUR LA CLASSIFICATION MULTI-CLASSE SANS MASQUE

	Precision	Recall	F1-score	Support
0	0.94	0.91	0.92	270
1	0.89	0.84	0.86	230
2	0.86	0.91	0.88	255
3	0.98	0.99	0.98	245
Accuracy			0.91	1000
Macro avg	0.91	0.91	0.91	1000
Weighted avg	0.91	0.91	0.91	1000



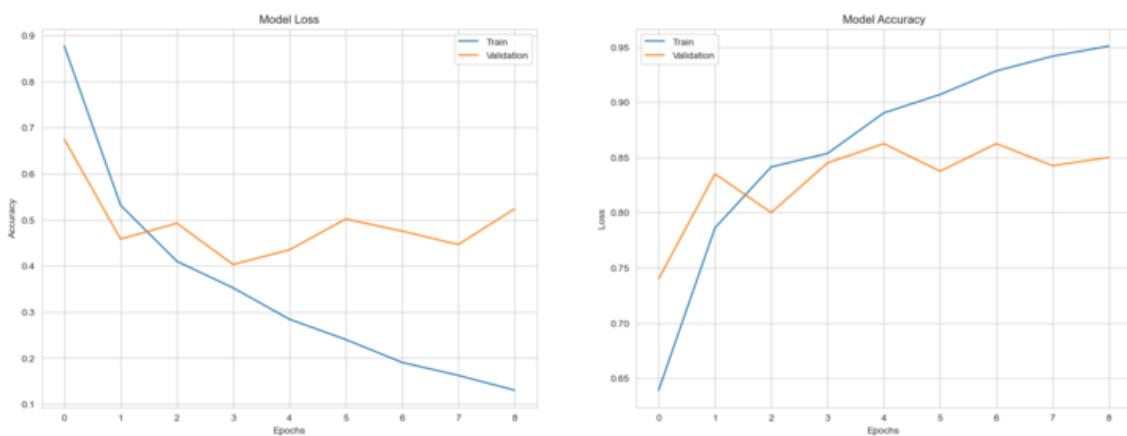
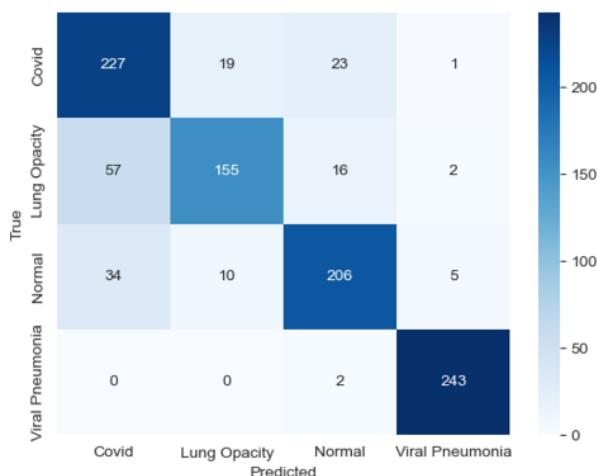
ANNEXE XIV : RESULTATS DU EFFICIENTNETB1 SUR LA CLASSIFICATION MULTI-CLASSE AVEC MASQUE

	Precision	Recall	F1-score	Support
0	0.86	0.81	0.83	270
1	0.83	0.82	0.82	230
2	0.81	0.88	0.84	255
3	0.97	0.96	0.96	245
Accuracy			0.86	1000
Macro avg	0.87	0.87	0.87	1000
Weighted avg	0.87	0.87	0.87	1000



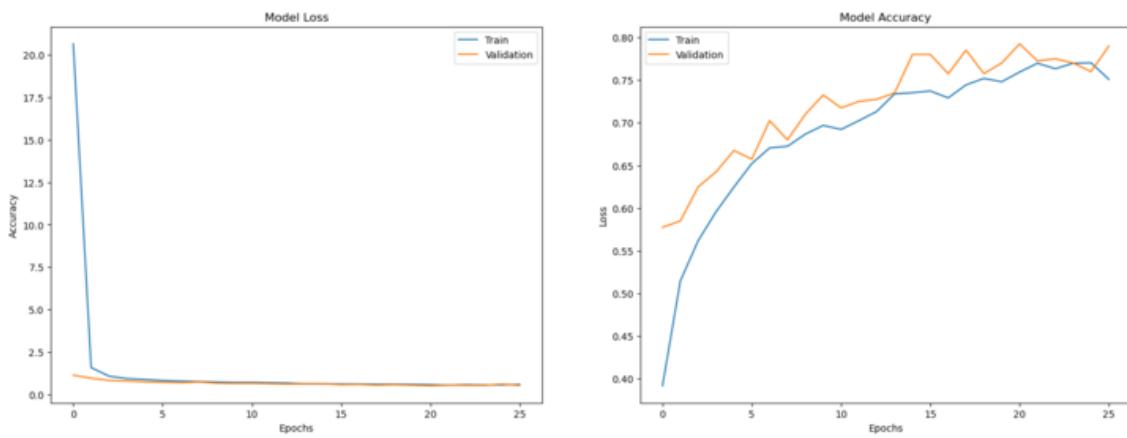
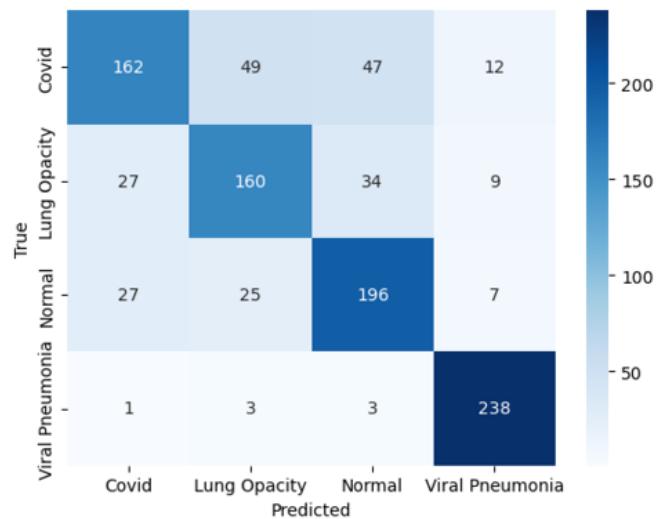
ANNEXE XV : RESULTATS DU VGG16 SUR LA CLASSIFICATION MULTI-CLASSE AVEC MASQUE

	Precision	Recall	F1-score	Support
0	0.71	0.84	0.77	270
1	0.84	0.67	0.75	230
2	0.83	0.81	0.82	255
3	0.97	0.99	0.98	245
Accuracy			0.83	1000
Macro avg	0.84	0.83	0.83	1000
Weighted avg	0.84	0.83	0.83	1000



ANNEXE IV : RESULTATS DU XCEPTION SUR LA CLASSIFICATION MULTI-CLASSE AVEC MASQUE

	Precision	Recall	F1-score	Support
0	0.75	0.60	0.67	270
1	0.68	0.70	0.69	230
2	0.70	0.77	0.73	255
3	0.89	0.97	0.93	245
Accuracy			0.76	1000
Macro avg	0.75	0.76	0.75	1000
Weighted avg	0.75	0.76	0.75	1000



ANNEXE IV : RESULTATS DU XCEPTION SUR LA CLASSIFICATION MULTI-CLASSE AVEC MASQUE

	Precision	Recall	F1-score	Support
0	0.78	0.79	0.78	270
1	0.78	0.76	0.77	230
2	0.81	0.78	0.79	255
3	0.93	0.98	0.96	245
Accuracy			0.83	1000
Macro avg	0.83	0.83	0.83	1000
Weighted avg	0.82	0.83	0.82	1000

