



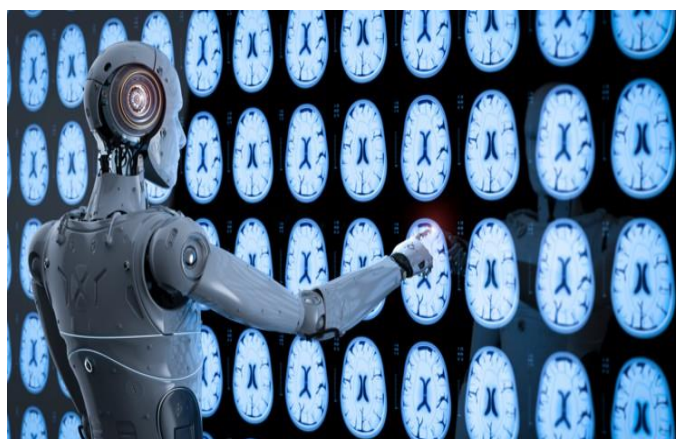
ÉCOLE CENTRALE CASABLANCA

## **RAPPORT DE PROJET ENJEU**

---

# **Traitement d'image à usage de diagnostic médical**

---



### **Réalisé par :**

Amine NAIT CHARIF  
Doha FEDDOUL  
Florian BOUBOUSSI DONGMO  
Youssef CHOUHAIDI

### **Tuteur Entreprise :**

M. Fayçal NOUSHI

### **Tuteur école :**

M. Mohamed-Hassan KHALILI

**Juin 2020**

# Table des matières

<b>Remerciements</b>	3
<b>I. Résumé</b>	4
<b>II. Introduction</b>	5
<b>III. Développement du projet</b>	6
1. Planification du projet	6
2. Contexte	9
3. Problématique	11
4. La méthodologie	11
5. Les résultats obtenus	17
6. Validation des résultats	30
<b>IV. Conclusion générale</b>	32
<b>V. Références</b>	34
<b>VI. Annexes</b>	35

# Remerciements

Nous tenons à remercier dans un premier temps toute l'équipe pédagogique de l'École Centrale Casablanca qui veille sur la qualité de la formation. Nous adressons un merci particulier à notre tuteur école M. Hassan KHALILI pour son aide, sa bonne humeur, ses précieux conseils dans les moments les plus délicats et le soutien qu'il nous a apporté durant notre réalisation des livrables. Nous remercions également le tuteur entreprise M. Fayçal NOUSHI pour son engagement à l'égard de notre équipe de projet enjeu et sa générosité en matière de partage d'informations et de connaissances ainsi que pour les échanges fructueux que nous avons eus lors de la réunion du kick-off. Un très grand merci à Mme. EL HAMDANI KAWTAR qui s'est montrée très à l'écoute de nos difficultés en termes de cohésion d'équipe, nous a aidés à surmonter nos crises, à savoir exploiter notre intelligence collective ainsi qu'à trouver un terrain d'entente entre les membres de l'équipe.

# I. Résumé

Le secteur sanitaire, comme tous les autres secteurs, est révolutionné par les nouvelles technologies. L'intelligence artificielle et particulièrement le traitement d'images est un outil informatique prometteur capable d'améliorer la qualité des soins médicaux et d'expandre la couverture sanitaire. En effet, l'intervention de l'intelligence artificielle dans ce domaine connu pour la quantité gigantesque de données à gérer, promet une métamorphose en matière du diagnostic médical. Dans ce sens, l'objectif de notre projet est de trouver un algorithme performant, rapide et précis permettant de détecter les anomalies sur des images médicales. Etant donné que le Machine Learning est une technique puissante de l'intelligence artificielle, nous avons décidé de déployer ses prouesses pour affiner la classification de nos images tout en nous focalisant sur deux anomalies : le covid-19 et le cancer du cerveau. Avant de se lancer dans la réalisation de nos algorithmes, une familiarisation avec le domaine sanitaire et le traitement d'images s'est avérée nécessaire, ce qui a donné naissance par la suite à nos deux synthèses bibliographiques. Il est enfin à signaler qu'on pourrait toujours améliorer la précision de nos programmes implémentés sur python et qu'ils devraient passer par des validations cliniques avant d'être employés dans le diagnostic médical.

## II. Introduction

D'année en année, l'intelligence artificielle s'enracine de plus en plus profondément dans notre monde, un monde en pleine révolution digitale. Les promesses de cet outil informatique s'étendent à tous les domaines de la vie et ses potentiels atouts sont l'objet de maintes recherches. En effet, l'émergence des technologies se basant sur l'intelligence artificielle s'est prouvée avantageuse et fructueuse, contribuant ainsi à la croissance de nombreux secteurs. Etant donné que le secteur sanitaire est de plus en plus touché par la numérisation et que celui-ci se caractérise par une quantité gigantesque de données, l'emploi de ces technologies dans la sphère médicale est en plein essor. En effet, les médecins sont contraints à gérer une énorme quantité de données provenant des articles scientifiques, des essais cliniques et des caractéristiques propres à chaque patient afin de fournir des services précis, performants et sans dangers. C'est donc ce qui justifie le déploiement de ces nouvelles technologies dans divers domaines de la santé comme le diagnostic médical, la chirurgie, la prédiction et le traitement des patients. Ces dernières années, l'expansion de cet outil dans le secteur médical va crescendo et ceci dans le but de pouvoir expandre la couverture sanitaire, améliorer la prise en charge des patients, personnaliser les traitements médicaux et trouver des remèdes aux maladies redoutables. C'est pour ces raisons que plusieurs géants du domaine informatique, tels que IBM, Google et Microsoft, mobilisent des millions de dollars pour élaborer des outils médicaux efficaces s'appuyant sur l'intelligence artificielle.

De toutes les branches de l'informatique employées dans le domaine médical, le traitement d'images s'est avéré particulièrement important et prometteur, notamment en ce qui concerne les diagnostics médicaux. En effet, par ses algorithmes, le traitement d'images est capable de détecter des anomalies sur des imageries médicales tels que les IRMs, les radiographies et les endoscopies. Sa précision peut même dans certains cas surpasser celle de l'homme.

Dans ce rapport, nous allons restituer les différentes étapes qui ont marqué le développement de notre projet enjeu, en abordant dans un premier temps son contexte et sa problématique. Ensuite, nous allons expliquer la méthodologie du travail que notre équipe a adoptée en suivant la planification que nous nous sommes fixées au début de l'année scolaire. Enfin, nous présenterons les résultats auxquels nous avons abouti.

# III. Développement du projet

Dans le but de développer à bien notre projet, nous avons mis en place un plan d'action dont le respect s'est avéré essentiel.

## Chapitre 1 : Planification du projet

Afin de bien mener notre projet, nous avons mis en place une feuille de route dans laquelle nous avons mentionné et détaillé nos différents objectifs à atteindre, les moyens nécessaires pour y parvenir, les étapes à suivre progressivement dans la réalisation du projet et les méthodes de suivi de celles-ci.

### I- LES OBJECTIFS À ATTEINDRE

Dès le début de notre travail, nous avons établi l'objectif à atteindre dans notre charte de projet. Nous avons également rédigé des sous-objectifs afin de mieux parvenir à notre but principal.

#### a- L'objectif du projet

Notre objectif principal dans ce projet est de répondre aux besoins de notre entreprise en lui livrant un algorithme de détection des cancers sur des imageries médicales.

Cet objectif a été modifié par la suite avec l'apparition du covid-19, une pandémie qui a bouleversé la stabilité du monde entier au vue des conséquences qu'elle a engendrées.

L'objectif sur lequel nous nous sommes basés à la fin du projet (soumis par notre tuteur entreprise) était alors celui de réaliser un algorithme de détection du covid-19 sur des radiographies pulmonaires. Compte tenu de notre avancement, nous avons également réalisé un programme de détection du cancer du cerveau.

#### b- Les sous-objectifs

Pour parvenir à nos fins, nous avons mis sur pied des sous-objectifs :

- Connaitre les fondamentaux sur les cancers, les tumeurs et leurs détections : par notre manque de connaissances sur cette branche de la médecine, il nous paraissait important de débiter par une familiarisation avec tous les concepts et principes tournant autour des cancers. De plus,

nous ne saurons élaborer un dispositif de détection numérique si nous ne connaissons initialement comme s'y prend un médecin pour le faire.

- Connaitre les fondamentaux sur le traitement d'images numériques : étant donné que nous devions avoir recours à des procédés et techniques de traitement d'images, nous avons jugé nécessaire de nous instruire sur toutes les notions et principes relevant du traitement d'images sous Python.
- Connaitre les fondamentaux sur le Machine Learning : cet objectif, initialement facultatif, visait à nous préparer au cas où l'on devait être emmené à recourir au Machine Learning pour améliorer le programme. En effet, le tuteur entreprise avait déjà eu à le mentionner dans la fiche projet.

## II- LES OUTILS NÉCESSAIRES POUR RÉUSSIR NOTRE PROJET

Vu que nos sous-objectifs sont principalement liés à des recherches bibliographiques, nous avons eu besoin de nombreux documents et articles de recherches spécialisés dans ces différents domaines.

La recherches de ces articles n'a pas été évidente au début mais nous avons pu, avec l'appui de nos deux tuteurs, obtenir un bon nombre de documents via Google Scholar, PyImageSearch, Github et bien d'autres sites. En plus de cela, nous avons bénéficié de l'aide de certains médecins que nous avons contactés.

## III- LES ÉTAPES CHRONOLOGIQUES À SUIVRE

### Planification du projet



Travail réalisé

Novembre



#### Compréhension du projet

- Rédaction de la charte projet
- Mini-soutenance
- Réunion quick-off avec le tuteur entreprise

#### Recherche bibliographique

- Familiarisation avec les techniques de traitement d'images, les imageries médicales et différentes tumeurs
- Rédaction du premier livrable : synthèse bibliographique
- Soutenance

Janvier



Février-Mars



#### Finalisation des synthèses bibliographiques

- Rédaction du deuxième livrable : état des lieux
- Familiarisation avec le Machine Learning
- Soumission des deux livrables

#### Réalisation des algorithmes

- Réunion avec les tuteurs
- Réalisation d'un algorithme de dépistage du Covid-19
- Réalisation d'un algorithme capable de détecter le cancer du cerveau

Avril-Mai



Juin



#### Finalisation du projet

- Soumission des résultats
- Soutenance finale

Les objectifs fixés et les documents dont nous aurons besoin à notre possession, nous avons élaboré un plan chronologique à suivre pour être dans les temps :

**Septembre-Décembre :** Afin de répondre à nos deux premiers sous-objectifs, nous avons, durant le premier semestre, effectuer une importante recherche bibliographique sur les cancers et le traitement d'images en général et en particulier sur python. Cette recherche bibliographique a fait l'objet de notre premier livrable : LA SYNTHÈSE BIBLIOGRAPHIQUE (voir annexe).

- Septembre-Octobre : Premier contact avec le projet clôturé par une mini-soutenance où nous avons présenté, en élaborant une charte, notre perception sur ce dernier.
- Novembre-Décembre : recherche bibliographie et rédaction de LA SYNTHÈSE BIBLIOGRAPHIQUE.

**Janvier-Juin :** la première documentation terminée, nous trouvions toujours quelques difficultés avec les techniques utilisées par le corps médical pour diagnostiquer un cancer. Nous avons donc, afin de contourner ce problème, débuté une étude du Machine Learning sous python. L'apparition du covid-19 au Maroc a poussé le tuteur entreprise à proposer de nouveaux objectifs pour le projet, ce qui nous a conduit de facto à revoir notre travail.

Janvier-Mars : Correction du premier livrable et soutenance de fin de semestre. Rédaction du second livrable portant sur le Machine Learning et les logiciels existant dans ce domaine à travers le monde : ÉTAT DES LIEUX (voir annexe). Soumission des deux livrables aux tuteurs.

Mars-Mai : Développement d'un algorithme de détection du covid-19 et du cancer du cerveau.

Juin : Finalisation du projet et soumission des résultats.

#### IV- SUIVI DE NOTRE ÉVOLUTION

Afin de nous assurer une évolution dans les temps et d'une façon optimale et efficace, nous avons mis en place au sein de l'équipe un certain nombre de règles :

- Tout d'abord à travers les conseils donnés durant des cours de gestion d'équipe par l'école, nous avons procédé à une répartition des postes et des fonctions entre les membres de l'équipe. Cette répartition nous a permis de gagner en efficacité et en temps.
- De plus, nous avons mené régulièrement des réunions avec nos différents tuteurs pour avoir des retours et directives en continu de leurs parts.



# Chapitre 2 : Contexte

## 1. Les projets enjeux à l'Ecole Centrale Casablanca

L'école Centrale Casablanca a mis la réalisation des projets au cœur de sa stratégie en matière de formation de l'élève-ingénieur. La réussite de ces projets offre un avantage capital aux lauréats de cette école, fruit d'une convention étatique entre le Maroc et la France en termes d'acquisition de compétences intrinsèques à l'ingénieur à savoir le leadership, la prise d'initiatives et la bonne gestion du relationnel. Ces qualités lui permettent de faciliter le travail en équipe, assimiler les directives et mettre ses compétences à profit. Parmi les projets soumis aux élèves tout au long de leur cursus, nous avons le projet enjeu. Le projet enjeu est un projet de première année qui tourne autour d'un problème réel soumis par une entreprise. En effet il a été conçu dans l'objectif de familiariser l'étudiant avec le monde de l'entreprise, de s'ouvrir sur ses aléas, ses enjeux et ses exigences. Il s'étale sur toute l'année académique sous la supervision d'un tuteur école. Notre progression sur le projet est régulièrement évaluée à travers des mini-soutenances, à l'issue desquelles nous recevons des retours constructifs des membres du jury et qui nous seront utiles afin de présenter des résultats conformes aux attentes de l'entreprise en question. Ce rapport présente à cet effet, un récapitulatif sur notre état d'avancement concernant notre projet enjeux intitulé TRAITEMENT D'IMAGES À USAGE DE DIAGNOSTIC MÉDICAL proposé par l'entreprise ZEN-NETWORK.

## 2. ZEN NETWORKS



Fondée en 2012 par M. Fayçal Noushi, Zen Networks est une entreprise marocaine spécialisée dans les domaines du IoT, Monitoring, Security, Cloud et DevOps. Par sa flexibilité et son efficacité, Zen Networks propose des services très profitables à ses clients et des solutions innovantes qui sont déployées dans les réseaux d'opérateurs mobiles à travers le monde ainsi que dans d'autres secteurs. Zen Networks intervient donc surtout dans l'intégration et la création de solutions généralement axées sur la surveillance, de l'automatisation et du Devops pour les réseaux.

Par ailleurs, cette entreprise est celle qui a proposé le sujet de notre projet enjeu intitulé « Traitement d'images à usage de diagnostic médical ».

## Chapitre 3 : La problématique

L'augmentation progressive et perpétuelle des cas de cancers pousse un bon nombre de chercheurs à améliorer les dispositifs de diagnostic et de traitement de cette maladie. Compte tenu du fait que cette dernière reste encore à ce jour, malgré les énormes progrès de la médecine, très difficile à prendre en charge surtout lorsqu'elle est détectée assez tardivement, l'accent de la recherche est essentiellement mis sur l'amélioration des dispositifs permettant de diagnostiquer ces pathologies à leurs débuts. Avec l'essor et l'efficacité que montre l'informatique dans de nombreux domaines, notamment celui du traitement d'images, l'entreprise Zen Network cherche alors à jumeler prouesses informatiques et médicales pour fournir un dispositif rapide, performant et efficace de détection des cancers.

## Chapitre 4 : Méthodologie

Notre travail a été principalement marqué par trois rendus. Il s'agit entre autres des livrables portant sur LA SYNTHÈSE BIBLIOGRAPHIQUE, L'ÉTAT DES LIEUX et enfin LA RÉDACTION DU PROGRAMME.

### I- LA SYNTHÈSE BIBLIOGRAPHIQUE

LA SYNTHÈSE BIBLIOGRAPHIQUE est notre premier livrable à travers lequel nous avons présenté nos premières recherches sur le projet.

#### 1- Le contenu

La SYNTHÈSE BIBLIOGRAPHIQUE comporte trois axes de recherche principaux à savoir le traitement d'images, les tumeurs et cancers et la détection de celles-ci.

#### a- Le traitement d'images

Dans cette partie, nous avons tout d'abord étudié les images numériques et leurs caractéristiques ; ensuite, nous avons recensé et étudié les principales méthodes et techniques portant sur le traitement d'images numériques et les bibliothèques sous Python permettant de les faire.

#### b- Les tumeurs et cancers

Il a été question ici pour nous d'effectuer un premier contact avec cette maladie. En effet, il s'agissait là pour nous d'un concept nouveau que nous devions impérativement comprendre pour avancer efficacement dans le projet. Nous avons donc dans cette partie présenté les différentes catégories d'images médicales (sur lesquelles nous serions emmenés à travailler) et de cancers.

#### c- La détections des cancers

Pour finir, afin de mieux être en mesure de développer un programme de détection des cancers, nous avons jugé bon de connaître les procédés actuels employés dans les cliniques et centres médicaux spécialisés. Nous avons donc à cet effet étudié de nombreux articles sur le sujet et contacté certains membres du corps médical pour plus d'informations.

### 2- Les difficultés rencontrées

Durant cette phase de recherche et d'apprentissage, nous avons eu à faire face à une difficulté majeure portant sur la dernière partie du livrable. En effet, il a été difficile pour nous de trouver des articles ou autres documents présentant d'une façon concrète les méthodes de diagnostic de cancers.

## II- L'ÉTAT DES LIEUX

L'usage du Machine Learning pour le développement de notre programme n'était pas au début une nécessité mais une simple éventualité. Toutefois compte tenu de notre avancement et des difficultés rencontrées durant nos premières recherches, nous avons finalement décidé de recourir à ce dernier. Pour cela nous

avons donc effectué une seconde recherche bibliographique qui a fait l'objet de notre second livrable : l'ÉTAT DES LIEUX.

## **1- Le contenu**

Tout comme le premier livrable, le contenu de celui-ci a été divisé suivant trois grands axes : L'Intelligence Artificielle et le Machine Learning, le Machine Learning sous Python et les logiciels de traitement d'images médicales dans le monde et au Maroc.

### **a- L'Intelligence Artificielle et le Machine Learning**

Dans cette partie, nous avons présenté le concept d'Intelligence Artificielle 'I.A' (visant le développement des facultés humaines de compréhension, de mémorisation, de prise de décisions, d'expression de sentiments et bien d'autres sur une machine) et de Machine Learning 'M.L' (visant le développement de méthodes d'apprentissage sur des machines).

### **b- Le Machine Learning sous Python**

Après une première introduction à l'I.A et au M.L dans la première partie, nous avons présenté les bibliothèques et quelques algorithmes python permettant l'usage du M.L.

### **c- Les logiciels de traitement d'images médicales dans le monde et au Maroc**

Nous avons en dernier lieu effectué une étude sur les logiciels de diagnostic de cancers existants à travers le monde et au Maroc afin d'une part de savoir les recherches et travaux effectués dans ce domaine et d'autre part afin de pouvoir, dans la mesure du possible, étudier ces programmes pour la conception du notre.

## **2- Les difficultés rencontrées**

Lors de cette étape nous avons eu du mal à comprendre certaines notions liées au M.L en générale et sous Python en particulier. De plus, nous n'avons pas pu avoir connaissance des logiciels employés au sein des cliniques marocaines pour la détection de cancers.

### III- DÉVELOPPEMENT DES PROGRAMMES

Au début du mois de mars, nous avons commencé la rédaction de nos premières esquisses de programmes. En effet, afin de débiter concrètement la mise en pratique, nous avons débuté par des programmes simples de détection de couleurs, contours et bords sur une image. Après une réalisation et une étude de concepts plus poussés en machine Learning, nous sommes passés à la rédaction de programmes faisant recours à des techniques un peu plus avancées et complexes.

#### 1- Développement du programme de détection du covid-19

Pour la rédaction du programme de détection du covid-19, nous sommes partis d'une première ébauche de programme fourni par notre tuteur entreprise. À travers ce dernier, nous avons pu davantage comprendre le mode d'application de certaines fonctionnalités de python (usant du Machine Learning).

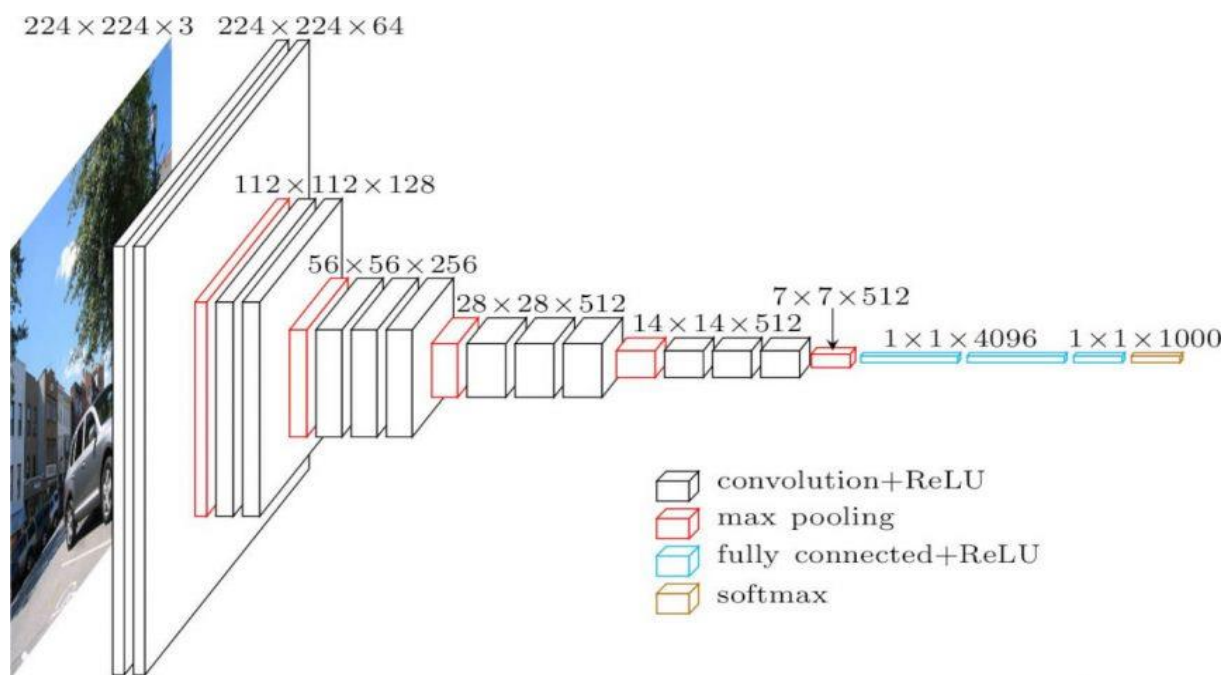
##### a- Étude de l'ébauche de programme

Cette étude nous a présenté une architecture basique de fonctionnement d'un programme d'apprentissage et nous a permis de soulever certains points importants.

Tout d'abord, nous nous sommes familiarisés avec différents réseaux de neurones employés par python. Ceci nous a permis de sélectionner le plus efficace à savoir VGG16 dans notre cas. VGG16 est un réseau de neurones convolutifs pour la classification et la détection d'éléments sous une image, préprogrammé sous Python. Il a été proposé pour la première fois en 2014 par Karen Simonyan et Andrew Zisserman, deux chercheurs de l'université d'Oxford. VGG16 a été développé et entraîné pendant une semaine sur une bibliothèque de plus de 15 millions d'images (obtenus d'ImageNet, un ensemble d'images de plus de 20 000 catégories de haute résolution et étiquetées.). Son architecture (voir image suivante) révèle une succession de cinq blocs de couches de neurones dont la première débute sur un traitement d'une image RGB de taille 224\*224 pixels (c'est

pour cela que dans un programme faisant recours à VGG16, les images subissent un prétraitement où elles sont converties en images RGB  $224 \times 224$ ).

VGG16 présente toutefois deux inconvénients majeurs : Il est assez lent à entraîner et nécessite énormément d'espace mémoire en comparaison avec d'autres réseaux de neurones. Néanmoins, ces performances et sa précision restent au-dessus des autres réseaux convolutifs ce qui fait de lui le réseau par excellence pour des travaux de précision.



### ARCHITECTURE DE VGG16

Ensuite, nous avons pu remarquer certaines nuances dans l'usage classique de certains programmes de détection. En effet, si nous rédigeons un simple programme de détection ou non du covid-19 sur une radiographie, il est assez difficile de savoir si en réalité le programme ne détecte pas en fait un patient malade (d'une quelconque maladie pulmonaire autre que covid) d'un patient sain. Pour qu'un programme s'avère plus efficace quant à la détection du covid-19, il faudrait non pas seulement l'entraîner avec des radiographies de patients sains et atteints du covid-19, mais il faudrait également l'entraîner avec des radiographies



de patients atteints d'autres maladies pour que le programme puisse faire la différence entre les maladies. Il pourrait donc être pertinent d'insérer deux étapes au sein du programme : la première consistant à diagnostiquer si le patient est sain ou malade ; la seconde consistant, au cas où le patient serait malade, à déterminer s'il s'agit ou non du covid-19.

En plus, le premier point mentionné ci-dessus sur le temps nécessaire pour le fonctionnement de réseaux de neurones tels que VGG16 nous limite considérablement sur le développement d'un programme hautement efficace. En effet, un tel programme pouvant, selon la complexité, prendre des heures voire des jours pour fonctionner sur nos machines, il nous est difficile compte tenu des machines à disposition de développer un tel programme et de savoir si oui ou non il est effectivement en état de marche et ne contient pas des erreurs de ligne de code (le temps de vérification serait assez long et fastidieux). Nous privilégierons donc un programme assez complexe mais à la hauteur des machines employées.

#### **b- Rédaction du programme final**

Après une étude et des essais, nous avons pu rédiger le programme final. Il ne se limite pas à diagnostiquer juste le covid-19 sur une radiographie mais fait la différence entre cette maladie et une autre comme mentionné plus haut. Les différentes mesures de précision de celui-ci sont assez encourageantes. Il s'avère toujours être difficile à lancer sur nos machines compte tenu du Machine Learning employé. Cette limite des capacités de nos machines nous impose certaines contraintes quant à l'amélioration du programme.

Il n'est toutefois pas encore assez développé pour pouvoir servir de test officiel et reconnu pour le covid-19. En effet, pour qu'il en est autrement, le programme devrait respecter les règles **des rapports transparents d'un modèle de prédiction multi-variable pour le pronostic ou le diagnostic individuel** (en anglais **Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD)**) ; le programme devrait également être l'objet des tests scientifiques rigoureux et être validé par la communauté scientifique.



## 2- Développement du programme de détection des tumeurs cérébrales

Nous avons bénéficié de nos différents travaux menés dans le cadre du programme de détection du covid-19 en le réadaptant pour détecter des tumeurs cérébrales. La précision du programme et ses caractéristiques restent à peu de chose près identique au programme portant sur le covid-19. Pour les mêmes raisons que précédemment, ce dernier ne saurait être employé dans un cadre médical effectif et reste juste un outil de premier test pour un utilisateur.

# Chapitre 5 : Les résultats obtenus

Le cadre méthodologique de ce projet de traitement d'images reposera sur une approche de classification par apprentissage automatique (Machine Learning ML). Le principe est d'extraire des caractéristiques de la tumeur puis d'élaborer un modèle de classification à partir d'une base de données (Dataset). Dès lors, l'essor du Machine Learning a redonné un fort intérêt au traitement et à l'analyse des grandes masses de données, notamment à des fins d'analyses prédictives.

Dans ce contexte, la comparaison des algorithmes et programmes utilisés représente un enjeu important pour mesurer l'impact des avancées dans le domaine. En effet, les benchmarks qui visent à comparer les performances des différents algorithmes d'apprentissage dans des tâches spécifiques, surtout dans le domaine de traitement d'images médicales, ont pris une place prépondérante chez les spécialistes du Machine Learning. Ces dernières années, l'idée de l'évaluation comparative a également gagné en popularité dans le domaine de l'analyse d'imageries médicales. De tels benchmarks, appelés "défis", ont tous en commun le fait que différents groupes optimisent leurs propres méthodes d'apprentissage pour améliorer la précision de leurs modèles. De nombreux

articles sont ainsi publiées, ce qui permet aux autres Data scientists d'investir plus d'efforts dans l'optimisation de certains algorithmes que d'autres.

Ainsi, nous avons lancé nos recherches, tout d'abord sur **Kaggle** une communauté de data scientists qui organise des compétitions de Machine Learning où il existe une variété de Datasets intéressants. En effet, chaque année, **Kaggle** organise le **Data Science Bowl**. En 2017, ce défi a réuni près de 10 000 participants du monde entier. Collectivement, ils ont soumis près de 18 000 algorithmes, tous visant à aider les professionnels de la santé à détecter le cancer du poumon plus tôt et avec une meilleure précision. Les résultats étaient incroyablement prometteurs.

Dans le cadre de notre étude dans laquelle nous aurons besoin d'un ensemble de dataset pour réaliser notre algorithme, nous avons en plus de **kaggle**, lancé nos recherches sur **Github** (un site web et un service de cloud qui aide les développeurs à stocker et à gérer leur code, ainsi qu'à suivre et contrôler les modifications qui lui sont apportées) qui représente une référence majeure dans le développement informatique ainsi que dans la disponibilité des ressources. Après un benchmark des algorithmes existants sur ces deux sites, nous avons trouvé des ressources et des articles détaillant certaines démarches pour la détection d'une tumeur/cancer qui se résument comme suit :

**Prétraitement : cette étape consiste à éliminer les bruits que contient l'image ainsi qu'à renforcer son contraste.**

Les images médicales contiennent du bruit. Par exemple, une image contenant des poils avec la lésion peut provoquer une mauvaise classification. De ce fait, la première étape de la procédure d'extraction des tumeurs est le prétraitement, c'est-à-dire l'amélioration de la qualité de l'image par la suppression des informations indésirables, ce que l'on appelle les bruits d'image. Plusieurs inexactitudes peuvent se produire dans la classification si ce problème n'est pas traité correctement. Ainsi, de nombreux filtres peuvent être appliqués pour

éliminer les bruits (le bruit gaussien, le bruit de Speckle et le bruit de Poisson.) tels que le filtre médian, le filtre médian adaptatif, le filtre gaussien et le filtre de Wiener adaptatif.

En outre, les bruits d'image sont censés être supprimés ou ajustés en effectuant des techniques de traitement d'image telles que le renforcement du contraste, le lissage de l'image, la normalisation et la localisation de l'organe concerné. A titre d'exemple, les IRM du cerveau sont d'abord converties en niveaux de gris puis subissent un ajustement de contraste à l'aide d'une opération de lissage afin de réduire le bruit qu'elles contiennent.

### **Segmentation d'image : cette étape consiste à extraire la tumeur.**

La division de l'image d'entrée en régions, où les informations nécessaires pour un traitement ultérieur peuvent être extraites, est connue sous le nom de segmentation. Celle-ci consiste essentiellement en la séparation d'une région d'intérêt « region of interest ROI » de l'arrière-plan de l'image. Le ROI est la partie de l'image que nous voulons utiliser.

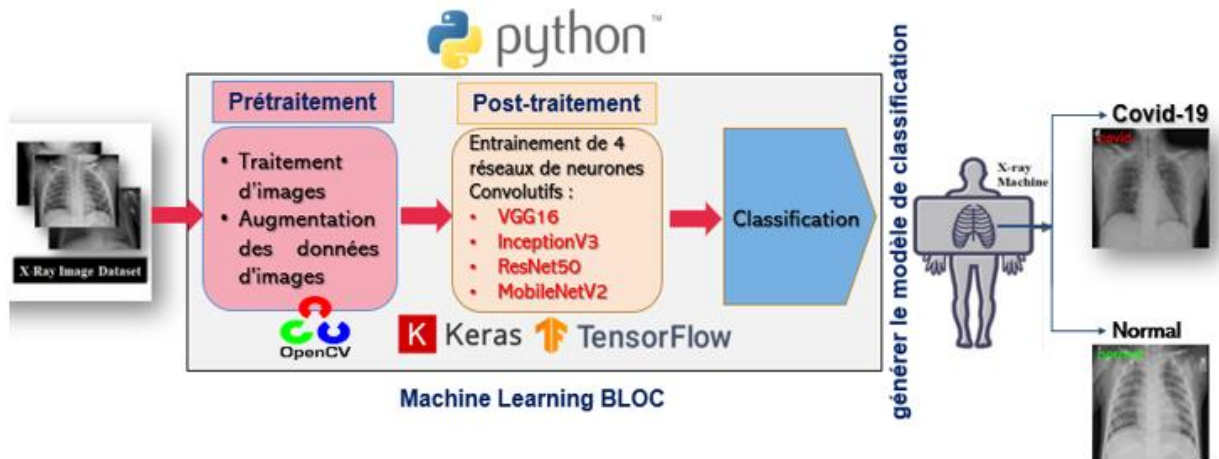
Dans le cas d'images cancéreuses, nous avons besoin de la partie lésionnelle (tumeur) pour extraire les caractéristiques de la partie malade. Il existe de nombreuses méthodes de segmentation d'image : seuillage d'histogramme, seuillage adaptatif, vecteur de flux de gradient, regroupement et croissance de régions statistiques et détection/extraction des contours.

### **Post-traitement : cette étape consiste à caractériser la tumeur selon certains paramètres.**

Après avoir terminé les étapes du prétraitement et la segmentation de l'image, il y a un post-traitement qui consiste à saisir les caractéristiques de la tumeur et à les classifier. En fait, cette étape de classification consiste à extraire les caractéristiques géométriques de la lésion. Cette approche de classification se base essentiellement sur des algorithmes de Machine Learning.

## 1. 1<sup>er</sup> cas : Covid-19

### Description du programme COVID19-NET



Le programme « COVID19-NET » repose sur une intelligence artificielle qui a « appris » à distinguer des images de poumons atteints par le Covid-19.

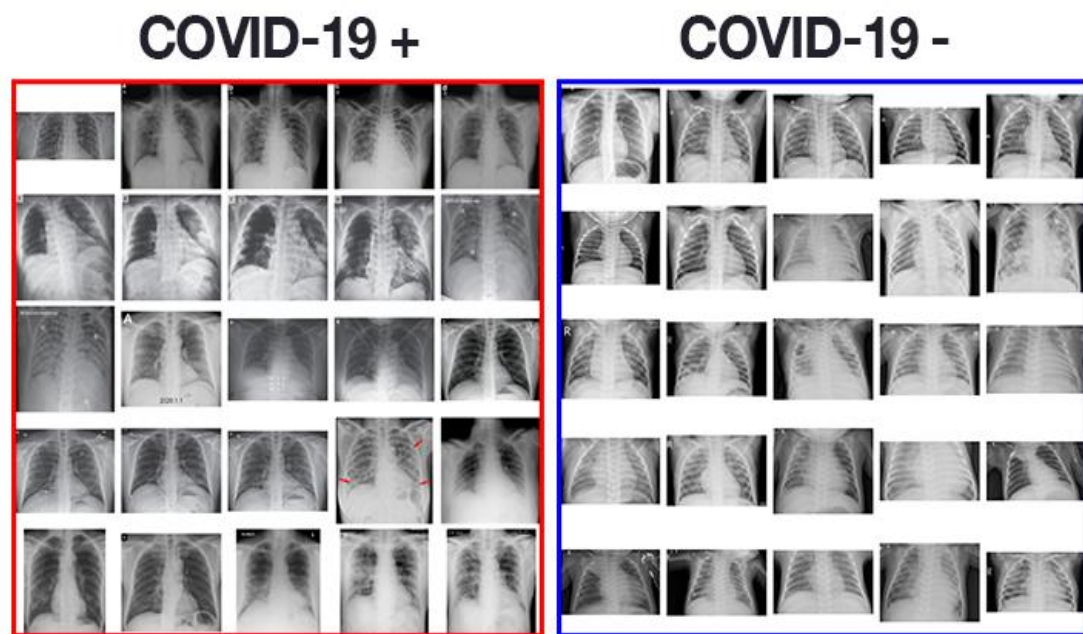
Pour entraîner le logiciel, nous avons utilisé une série d'images de radiographies pulmonaires de personnes ayant contractées la maladie et de personnes saines. Par conséquent, nous avons utilisé **les bibliothèques TensorFlow et Keras** pour former/entraîner, évaluer et tester notre programme. Quatre réseaux de neurones à convolution d'apprentissage profond très performants ont été testés à savoir **VGG16, InceptionV3, ResNet50 et MobileNetV2**.

La première étape suivie par le programme est l'amélioration de la qualité de l'image qui consiste à éliminer les bruits que contient l'image ainsi qu'à renforcer son contraste. Vient ensuite, en deuxième phase du prétraitement, l'augmentation de la base de données. Cette étape vise à augmenter artificiellement la taille de notre dataset en créant des versions modifiées des images. La bibliothèque de Deep Learning **Keras** offre la possibilité d'augmenter la taille de notre Dataset via la classe **ImageDataGenerator**. Dans cette étape du

prétraitement, toutes les images du Dataset ont été redimensionnées à la taille de 224x224 pixels afin d'implémenter nos réseaux de neurones à convolution.

Par la suite, il y a un post-traitement qui consiste à saisir et à extraire les caractéristiques de l'image. Dans le cadre de notre étude, nous nous sommes basés sur des algorithmes de Deep Learning. Ainsi, Nous avons implémenté notre code avec 4 réseaux de neurones à convolution : **VGG16**, **InceptionV3**, **ResNet50** et **MobileNetV2** afin de tester leurs précisions et générer notre modèle de classification. Dès lors, il a été possible de fournir des images de poumons de personnes malades afin d'être classifiées par le programme.

#### Dataset d'étude



Dans cette étude, les radiographies pulmonaires COVID-19 ont été obtenues à partir du DATASET partagé par le **Dr Joseph Cohen sur GitHub**, tandis que, les radiographies pulmonaires normales ont été sélectionnées via le **projet Kaggle "Chest X-Ray Images (Pneumonia)"**. Dans un premier temps, nos expériences ont été basées sur un ensemble de données créées avec des images de radiographie pulmonaire de **25 patients atteints par le COVID-19 et 25 radiographies de patients normaux** (50 images au total).

## Travail réalisé

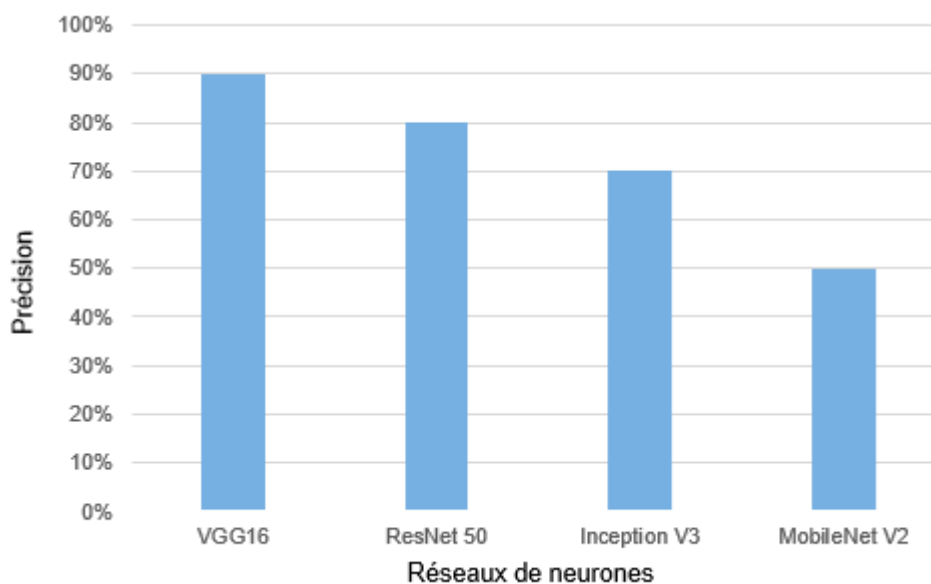
L'objectif de notre étude est de proposer un outil très performant basé sur le Deep Learning qui permettrait aux radiologues de dépister automatiquement le COVID-19 à l'aide d'une image radiographique de poumon.

Jusqu'à présent, nous avons :

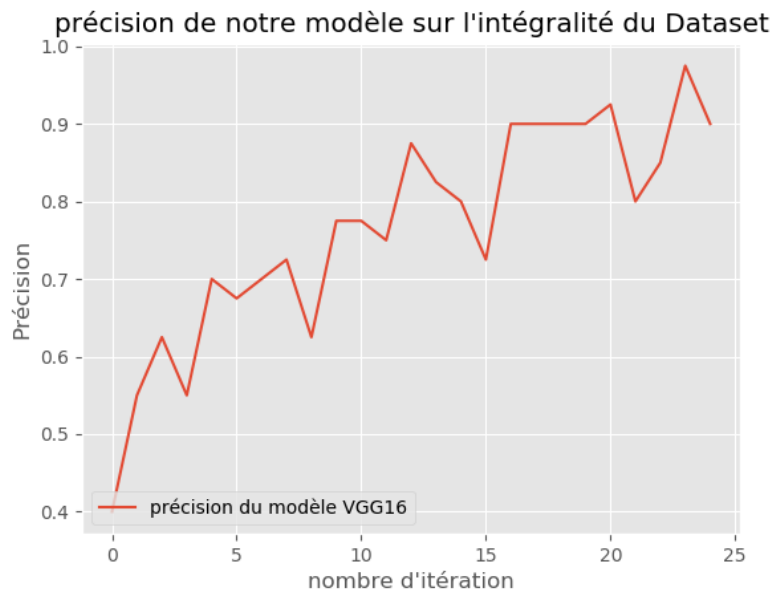
- Élaborer un modèle de classification (COVID-Net) qu'on a implémenté avec les quatre réseaux de neurones cités ci-dessus afin d'évaluer leurs précisions.
- Réaliser un rapport comparatif des performances entre différents modèles d'apprentissage profond avec des interprétations et des améliorations pour montrer celui le plus précis pour la classification COVID-19.
- Tester la fiabilité de notre modèle sur le dataset Test.
- Finalement, pour concrétiser notre étude, nous avons réalisé une application web dans laquelle nous avons inséré notre programme.

## Evaluation de la performance

Pour entrainer notre programme, 80 % des images radiographiques, y compris celles des patients malades et sains, ont été choisies au hasard. Les 20% restantes ont été utilisées pour évaluer le modèle.



Par la suite, nous avons décidé de nous focaliser sur le modèle entraîné par le réseau de neurones **VGG16** vu sa bonne précision par rapport aux autres modèles.



Nous avons tracé également une courbe permettant de présenter la précision du modèle VGG16 sur l'intégralité de notre Dataset. Comme vous le voyez sur la figure ci-dessus, lors de chaque itération de notre modèle, sa précision augmente. Elle fluctue entre 90% et 100% jusqu'à ce qu'elle se stabilise sur 90%.

Pour tester la fiabilité de notre modèle de classification, nous avons évalué sa précision définie comme étant :

$$\text{Précision} = \frac{\text{Nombre d'images correctement prédites}}{\text{Nombre total d'images testées}} \times 100\%$$

Le calcul montre que notre détecteur automatique COVID-19 obtient une précision de 90-92% sur notre dataset. Nous obtenons également **100% de sensibilité et 80% de spécificité** ce qui implique que :

- Parmi les patients qui sont touchés par le Covid-19 (vrais positifs), nous avons pu les identifier avec une précision de 100% comme « Covid-19 positif » grâce au modèle.
- Parmi les patients qui n'ont pas de COVID-19 (vrais négatifs), nous avons pu les identifier avec précision une précision de 80% comme étant « COVID-19 négatifs ».

### Qualité de notre modèle de classification – matrice de confusion

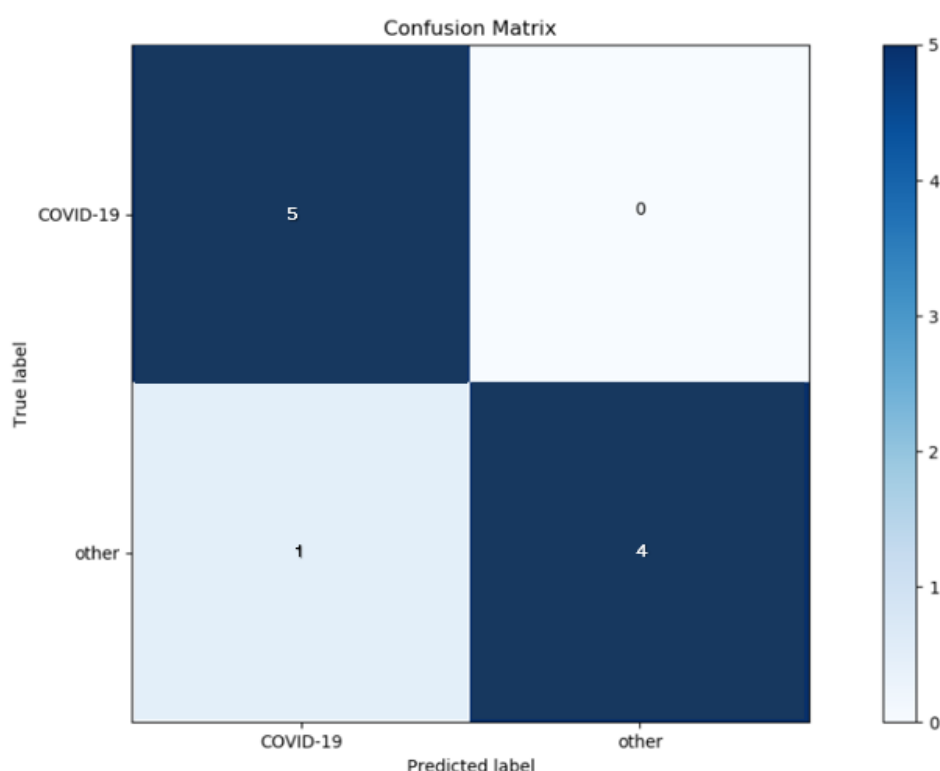
La matrice de confusion est un outil permettant de mesurer les performances d'un modèle de Machine Learning en vérifiant notamment à quelle fréquence ses prédictions sont exactes par rapport à la réalité dans des problèmes de classification. Cette matrice est présentée comme suit :

	Malade	Non malade
Test positif	VP	FP
Test négatif	FN	VN

- VP (vrais positifs) représente le nombre d'individus malades avec un test positif,
- FP (faux positifs) représente le nombre d'individus non malades avec un test positif,
- FN (faux négatifs) représente le nombre d'individus malades avec un test négatif,
- VN (vrais négatifs) représente le nombre d'individus non malades avec un test négatif.



Ainsi, la matrice de confusion de notre modèle est la suivante :



La matrice de confusion de notre modèle se lit alors comme suit :

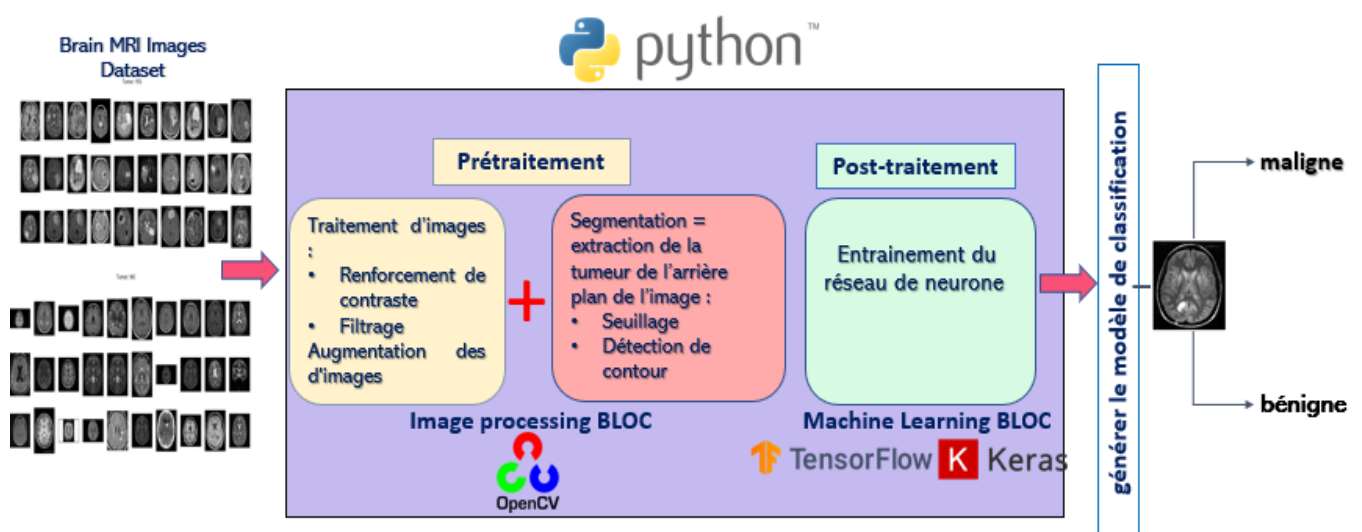
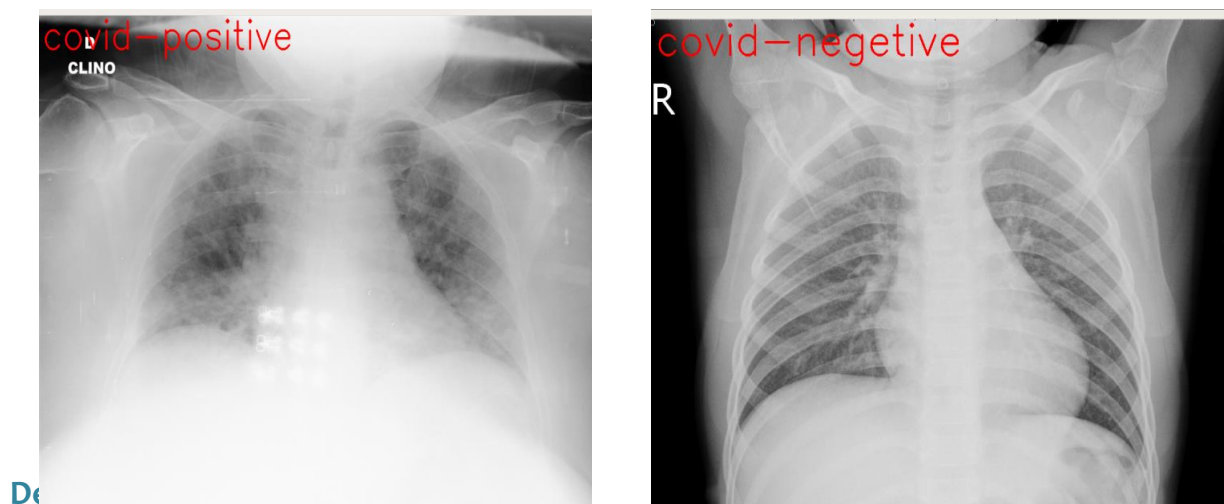
- **1ère ligne**, sur les 5 radiographies pulmonaires de personnes qui ont contracté le Covid-19, 5 ont été estimées par le système de classification comme tels et 0 ont été estimées comme faux-négatifs
- **2ème ligne**, sur les 5 radiographies de patients sains du covid-19, 4 ont été estimés comme tels et 1 vrais-négatifs

Afin d'améliorer la précision du programme, nous avons enrichi notre dataset à partir d'un ensemble d'images prises du Projet-Kaggle "**Chest X-Ray Images (Pneumonia)**" et également sur Github qui ont été partagées par le Docteur Cohen. Notre dataset contient ainsi **70 images de radiographie pulmonaire de patients malades du COVID-19** et **68 radiographies de patients normaux** (soit 138 images au total).

Notre modèle de classification a eu pour précision 96-97%. Nous avons obtenu également 100% de sensibilité et 92% de spécificité. Par ailleurs, la matrice de confusion de notre modèle de classification est la suivante :  $\begin{pmatrix} 68 & 2 \\ 5 & 63 \end{pmatrix}$ .

De plus, pour tester la fiabilité de notre modèle de classification, nous avons réalisé un code de détection du Covid-19 à partir de notre modèle de classification. Les deux images ci-dessous, dont on connaît antérieurement le résultat des IRM, ont bel et bien affirmé l'efficacité et la fiabilité de notre modèle.

Voici donc le résultat de notre code appliqué à deux images de notre TESTSET (le code de détection-Covid19 est mis en annexe) :

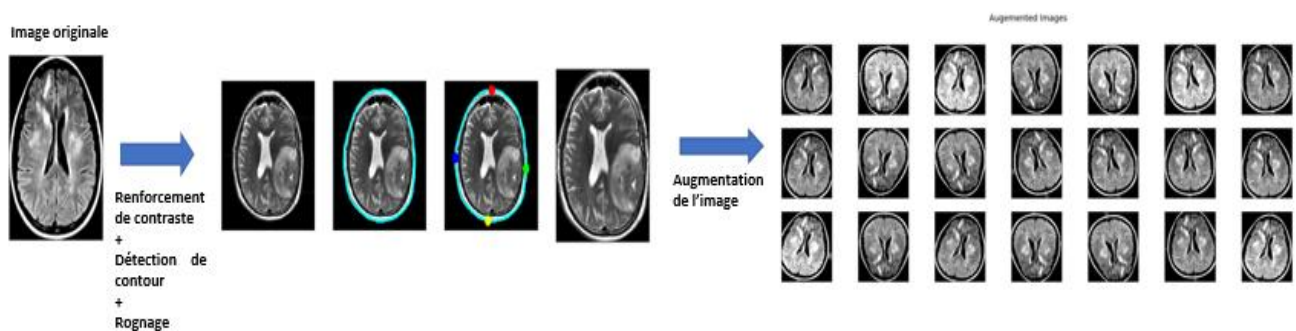


Le présent travail met en œuvre un système efficace de détection du cancer sur une IRM cérébrale donnée. Pour entraîner le programme, nous avons utilisé une série d'images d'IRM cérébrales comportant des tumeurs malignes (cancer) et des tumeurs bénignes.

La première étape de la procédure de classification est comme dans le cas du covid-19, le prétraitement se faisant en deux étapes : élimination du bruit et amélioration du contraste.

Cette première étape du traitement entraîne la transformation de l'ensemble des données vers un format plus interprétable et rend les images plus adaptées à un traitement ultérieur. Vient ensuite, l'augmentation de la base de données. Comme nous l'avons évoqué précédemment, toutes les images du Dataset ont été redimensionnées à la taille de 224x224 pixels.

La deuxième étape du prétraitement consiste en la segmentation des IRM du dataset. Ceci permet d'effectuer un traitement sur l'ensemble des images afin de pouvoir mettre en avant certaines caractéristiques de la tumeur.



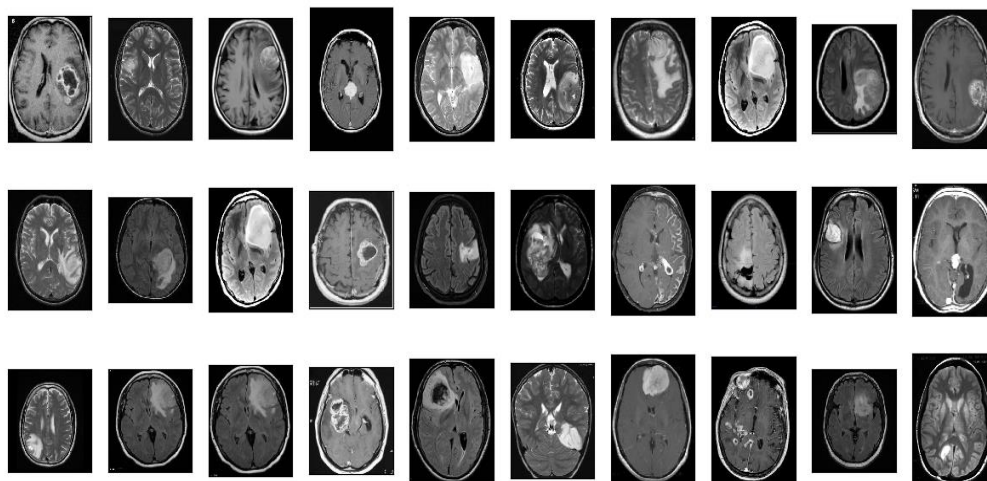
### Exemple de prétraitement

Après avoir terminé le processus du prétraitement, il y a un post-traitement qui consiste à extraire les caractéristiques de la tumeur. Ainsi, nous avons implémenté notre code avec le réseau de neurones **VGG16** afin de générer notre

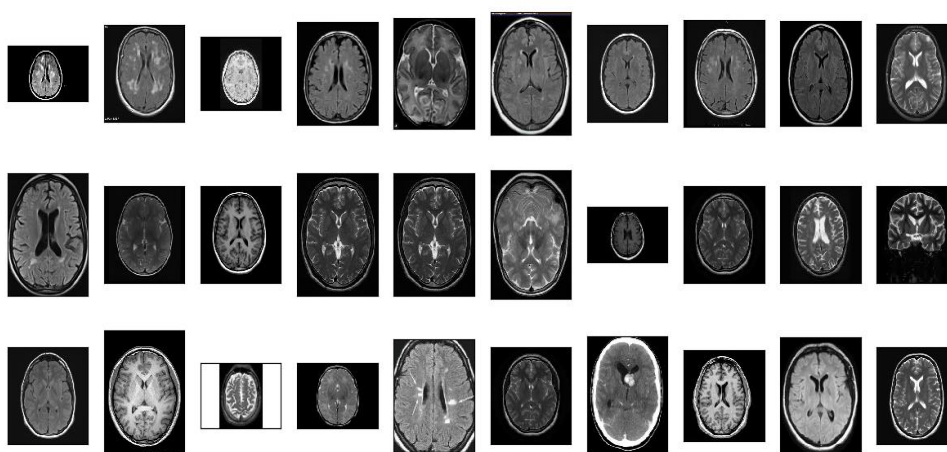
modèle de classification. Dès lors, il serait donc possible de lui fournir une IRM du cerveau, et le modèle pourrait alors déterminer la nature de la tumeur cérébrale.

## Dataset d'étude

Tumor: YES



Tumor: NO



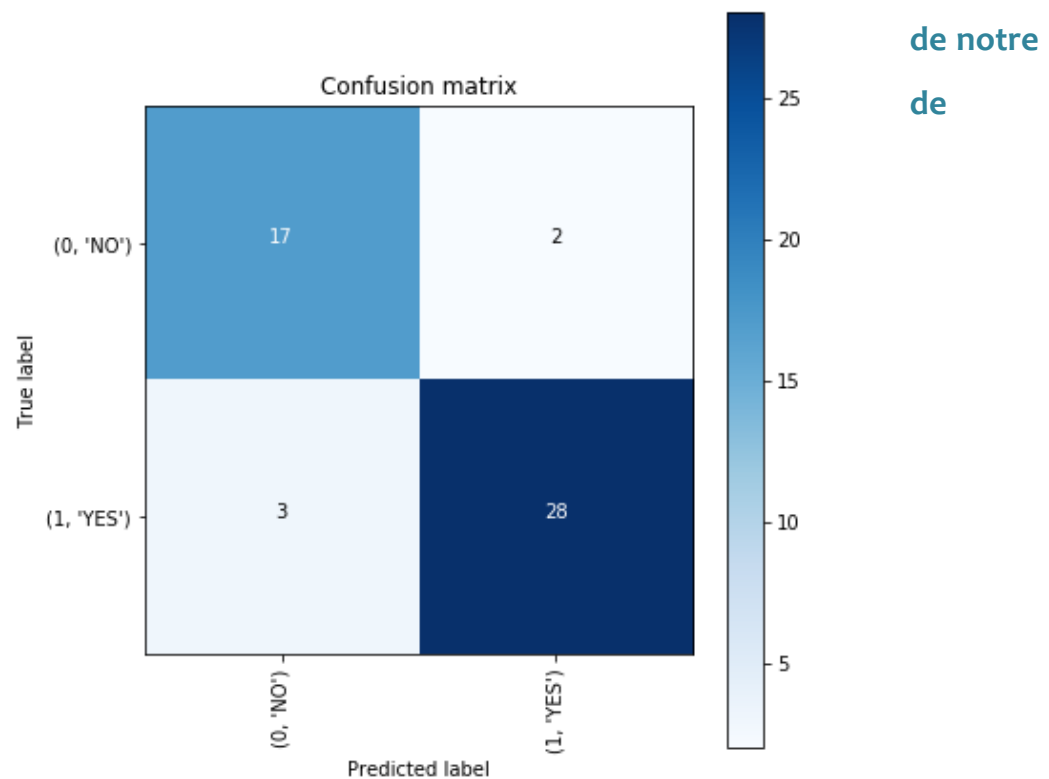
Dans cette étude, les IRM cérébrales ont été collectées à partir du dataset partagé par le professeur Navoneel Chakrabarty dans son projet-kaggle intitulé « **Brain MRI Images for Brain Tumor Detection** ». Nos expériences ont été basées sur un dataset créé avec des IRM 255 patients, dont 155 saines et 100 malades.

## Evaluation de la performance

Comme pour le cas Covid19, 80 % des IRM sont choisies au hasard pour entraîner notre réseau de neurones VGG16. Les 20% restantes sont utilisées pour évaluer le modèle. La précision trouvée a été égale à 92%. Nous obtenons également **85% de sensibilité et 93% de spécificité** ce qui implique que :

- Parmi les IRM de patients sains, (vrais positifs) 85% ont bien été bien identifiés par le programme.
- Parmi ceux malades (vrais négatifs), 92% ont bien été identifiés.

## Qualité modèle



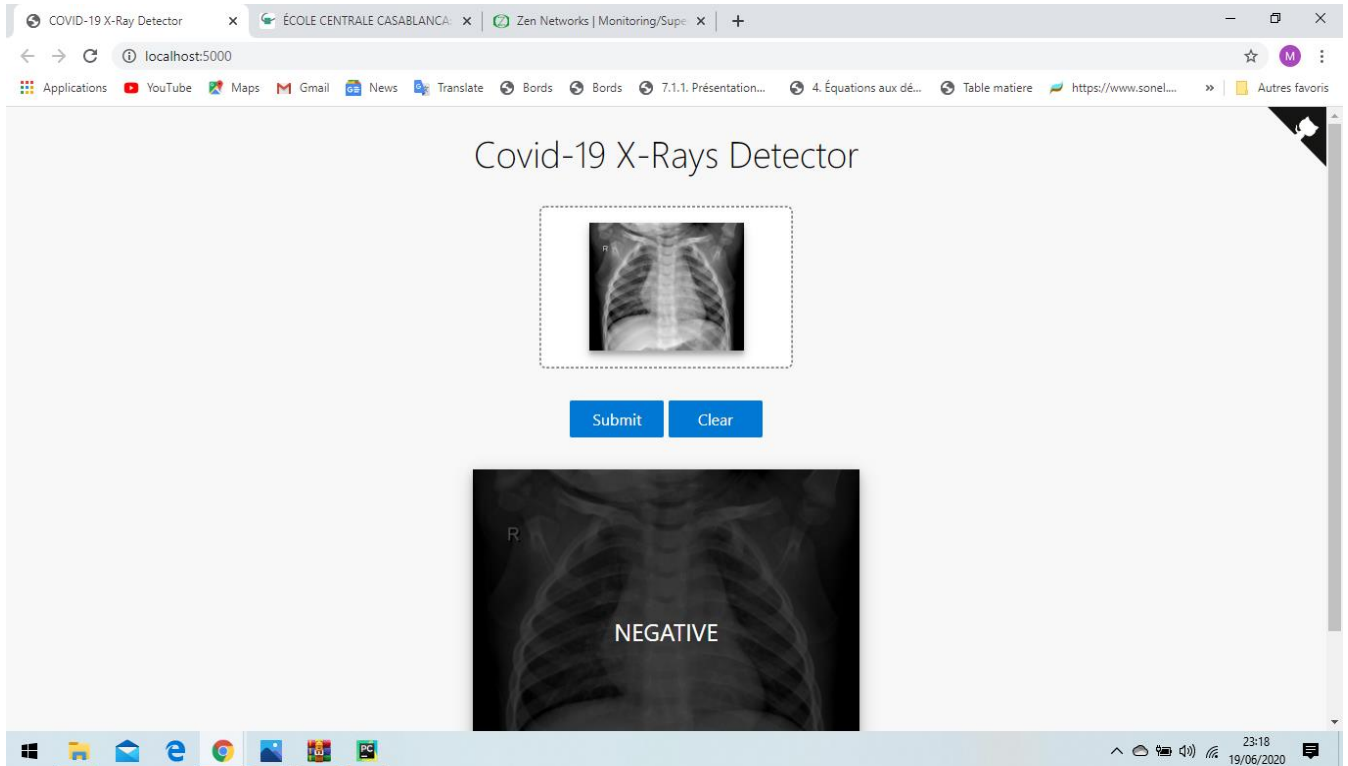
## classification – matrice de confusion

Par ailleurs, la matrice de confusion de notre modèle de classification est la suivante :

### Matrice de confusion du programme de détection du cancer du cerveau

La matrice de confusion se lit exactement comme dans le cas du programme de détection du covid-19.

Enfin, pour rendre nos programmes accessibles à n'importe quelle personne qui désire les utiliser pour classer une image médicale, nous les avons implémentés dans une page web en utilisant Python Flask.



Essai du programme covid-19 sur la page web

## Chapitre6 : Validation des résultats

Comme a été mentionné, le premier semestre était principalement dédié à la recherche bibliographique. Nous avons à plusieurs reprises modifié nos deux synthèses bibliographiques en tenant en considération les remarques instructives de notre tuteur école, et par la suite nous avons soumis le travail à notre tuteur entreprise M.Noushi Fayçal. Ce dernier, durant une réunion sur teams, nous a félicité pour notre engagement et pour la qualité de nos livrables qui ont fait

preuve d'un grand travail de recherche. En outre, après la lecture de notre dernier rapport, notre tuteur école M.Khalili et aussi M.El Khou Majda ont clairement applaudit le travail réalisé dans le cadre de notre projet enjeu. En effet, notre tuteur école avait clairement exprimé sa satisfaction quant à notre implication et notre dévouement et nous a appelé à garder le même rythme de travail. Par ailleurs, nous avons été encouragés par M.Noushi pour participer dans la compétition Covid Challenge organisée à l'Ecole Centrale Casablanca, une compétition dans laquelle nous avons fini en première place. Dans ce sens, nous avons pu bénéficier des remarques des membres du jury et les utiliser pour améliorer l'efficacité de notre travail. Il faut aussi signaler que nous avons eu des retours positifs et des directives d'amélioration de nos deux tuteurs lors d'une dernière réunion avant la soutenance finale et qui a eu lieu le 19 juin 2020. Car en effet, nous avons pu réaliser les résultats attendus, les programmes de détection d'anomalies, et soumettre à temps les livrables de notre projet enjeu.

## IV. Conclusion générale

En définitive, l'avancée spectaculaire de l'intelligence artificielle est déconcertante pour tous les domaines, y compris le secteur médical. En effet, l'utilisation de ces nouvelles technologies favorise l'expansion de la couverture sanitaire et offre un champ très large de finalités thérapeutiques ou diagnostiques, allant jusqu'à la prévention et la prédiction des maladies. Dans ce sens, notre équipe s'est concentrée sur le traitement d'images médicales pour détecter la pandémie covid-19 et aussi le cancer du cerveau. En utilisant les techniques du Machine Learning nous avons pu établir des diagnostics médicaux et différencier entre des patients sains et d'autres atteints d'anomalies. Certes, les algorithmes réalisés et les études comparatives menées ont pu faire preuve d'une grande précision et importance, mais le travail pourrait être amélioré. En effet, plus on enrichit notre base de données, plus la précision et la fiabilité de nos programmes va augmenter. Cependant, pour traiter une grande masse de données, il faut avoir une puissante carte graphique, car ceci peut endommager les ordinateurs. Il faut souligner que d'autres pathologies pulmonaires ou cérébrales peuvent être ajoutées à nos algorithmes de façon de l'enrichir et s'assurer de la fiabilité de sa classification une fois la base de données contient différentes anomalies et non pas juste des images médicales de patients malades et sains.

D'autre part, en pratique, l'usage de ces algorithmes nécessite la validation des professionnels du secteur sanitaire, la certification des autorités concernées et le passage par plusieurs études cliniques, ce qui n'est pas évident. En outre, le déploiement de l'intelligence artificielle dans la médecine soulève de nombreux problèmes liés à l'éthique et à la régulation. D'où l'importance de s'assurer que ces technologies ne mettent pas la vie des populations en danger.

Nos livrables sont le fruit d'un travail acharné et d'un engagement particulier de la part de tous les membres de l'équipe durant toute l'année. Nous étions tous passionnés par le domaine de l'intelligence artificielle et ce projet nous



a accordé une opportunité pour se familiariser avec cette branche informatique. Il faut avouer que le travail n'était pas simple et que nous avons rencontré plusieurs problèmes, tels que notre manque d'informations liées à la médecine et à la classification plus particulièrement, mais nous avons pu les surmonter. Cette expérience était constructive dans toutes les dimensions. En effet, nous avons développé nos compétences en matière de travail en équipe, nous avons appris à mieux s'organiser pour trouver un équilibre entre les cours enseignés à l'Ecole Centrale Casablanca et notre projet enjeu ainsi qu'à mieux gérer nos différences. C'était un plaisir de travailler ensemble et de développer nos connaissances informatiques et médicales.

## V. Références

- COHEN, Joseph Paul. GITHUB. “covid-chestxray-dataset/metadata.csv”. Disponible en ligne sur : <https://github.com/ieee8023/covid-chestxray-dataset/blob/master/metadata.csv>. Consulté le 20 juin 2020.
- ROSEBROCK, Adrian. “Detecting COVID-19 in X-ray images with Keras, TensorFlow, and Deep Learning”. 16 mars 2020. Disponible en ligne sur : <https://www.pyimagesearch.com/2020/03/16/detecting-covid-19-in-x-ray-images-with-keras-tensorflow-and-deep-learning/>. Consulté le 20 juin 2020.
- CHAKRABARTY, Navoneel. KAGGLE. “Brain MRI Images for Brain Tumor Detection”. Brain tumor Dataset.. Disponible en ligne sur : <https://www.kaggle.com/navoneel/brain-mri-images-for-brain-tumor-detection>. Consulté le 20 juin 2020.
- GITHUB. “Pretty & simple image classifier app template. Deploy your own trained model or pre-trained model (VGG, ResNet, Densenet) to a web app using Flask”. Disponible en ligne sur : <https://github.com/mtobeiyf/keras-flask-deploy-webapp>. Consulté le 20 juin 2020.

## VI. Annexes

- Les programmes finaux de détection du Covid19, le modèle de classification ainsi que le Dataset et le Testset : [https://centralecasablanca-my.sharepoint.com/:f:/g/personal/mohamed\\_chouhaidi\\_centrale-casablanca\\_ma/EtphnaxPC25BrgSUoLQSHclBodq3zur32xfjaJntOXting?e=CDoBow](https://centralecasablanca-my.sharepoint.com/:f:/g/personal/mohamed_chouhaidi_centrale-casablanca_ma/EtphnaxPC25BrgSUoLQSHclBodq3zur32xfjaJntOXting?e=CDoBow)
- Le programme final de détection du cancer de cerveau, le modèle de classification ainsi que le Dataset et le Testset : [https://centralecasablanca-my.sharepoint.com/:f:/g/personal/mohamed\\_chouhaidi\\_centrale-casablanca\\_ma/EhQfSfo1VBtli6ipCUWuJRwBEYBoARasjTZ9TyJxDYOBvQ?e=SabTSk](https://centralecasablanca-my.sharepoint.com/:f:/g/personal/mohamed_chouhaidi_centrale-casablanca_ma/EhQfSfo1VBtli6ipCUWuJRwBEYBoARasjTZ9TyJxDYOBvQ?e=SabTSk)
- Application web de détection du Covid-19 : [https://centralecasablanca-my.sharepoint.com/:f:/g/personal/mohamed\\_chouhaidi\\_centrale-casablanca\\_ma/Eno0owktb9hDjdO17Oh14ysBwM52iq3u1XQ-tnMMm5D7kA?e=n6J8LV](https://centralecasablanca-my.sharepoint.com/:f:/g/personal/mohamed_chouhaidi_centrale-casablanca_ma/Eno0owktb9hDjdO17Oh14ysBwM52iq3u1XQ-tnMMm5D7kA?e=n6J8LV)