

الجمهورية الجزائرية الديمقراطية الشعبية

**REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE**

وزارة التعليم العالي والبحث العلمي

**Ministère de l'Enseignement Supérieur et de la Recherche Scientifique**

جامعة أبي بكر بلقايد - تلمسان

Université Aboubakr Belkaïd – Tlemcen –

Faculté de TECHNOLOGIE



## **MEMOIRE**

Présenté pour l'obtention du **diplôme** de **MASTER**

**En** : Génie biomédical

**Spécialité** : Imagerie médicale

**Par** : Salhi Rawan

**Sujet**

**Détection et Classification de fumée dans les vidéos  
laparoscopiques**

Soutenu publiquement, le 20 / Juin / 2023 , devant le jury composé de :

|                    |     |                          |             |
|--------------------|-----|--------------------------|-------------|
| Mr. Bechar Hassane | MCA | Université de<br>Tlemcen | Président   |
| Mr. Gaouar Adil    | MAA | Université de<br>Tlemcen | Examinateur |
| Mme. Khemis Kamila | MCA | Université de<br>Tlemcen | Encadreur   |

Année universitaire : 2022/2023

## *Dédicace*

Je dédie ce travail :

À ma chère mère,

À mon cher père,

Qui ont toujours été là pour moi, me guidant avec amour, fournissant un soutien moral inébranlable et m'encourageant tout au long de mon parcours d'études, je dédie ce travail à vous, leur amour inconditionnel, leurs sacrifices, leurs précieux conseils ainsi que leur soutien financier ont été les fondations sur lesquelles j'ai pu construire ma réussite.

Chaque pas que j'ai fait, chaque défi que j'ai surmonté, je sais qu'ils étaient là, à mes côtés, m'encourageant à persévérer et à donner le meilleur de moi-même.

Que ce travail soit un témoignage de mon amour et de ma reconnaissance éternels envers eux. Je prie pour qu' *Allah* les préserve, leur accorde une santé, un bonheur et une longue vie.

Mes frères **Mohammed**, **Abderrahim** et **Firas** qui m'ont apporté un soutien moral précieux et des conseils tout au long de l'accomplissement de ce travail. Qu' *Allah* les protège et leur accorde la chance, le bonheur et surtout la réussite.

À tous ceux qui me sont chers et à tous ceux qui m'aiment,

## ***Remerciements***

Je commence par remercier, « **Allah** » tout puissant qui m'a béni ma santé, ma patience et ma volonté pour entreprendre et accomplir ce modeste travail, Cela représente le fruit des plusieurs années dévouement et de sacrifices.

J'exprime ma gratitude sincère à mon encadreur, **Mme Khemis Kamila**, dont les efforts inlassables ont contribué grandement à la réussite de ce travail. Je tiens à la remercier chaleureusement pour sa sympathie, sa patience, sa disponibilité, sa rigueur et son précieux soutien tout au long du processus.

Je tiens à remercier **Mr. Khouani Amin** pour leur aide pratique, leurs précieux conseils et leur encouragement constant.

J'aimerais également exprimer ma gratitude envers tous mes professeurs pour la qualité exceptionnelle de l'enseignement qu'ils m'ont dispensé tout au long de mes années passées au département de Génie biomédical de l'Université de Tlemcen.

Je tiens à adresser mes remerciements aux membres du jury pour l'honneur qui m'est accordé de soumettre ce travail à leur évaluation.

Enfin, un sincère remerciement est adressé à toutes les personnes qui ont joué un rôle, que ce soit directement ou indirectement, dans la réussite de ce travail.

---

## Résumé

---

La chirurgie laparoscopique présente des avantages évidents pour les patients, tels qu'une hospitalisation courte. Cependant, cette technique est complexe et peut entraîner facilement des complications. L'une des sources de ces complications est une vision altérée due à la contamination de l'optique de la caméra par la fumée, ce qui va introduire un laps de temps sans une vision claire. Il est donc crucial de rétablir rapidement une vision optimale pour améliorer le pronostic chirurgical. Une solution évidente serait de détecter automatiquement les images contaminées. Nous avons donc exploré la détection et la classification de la fumée dans les images laparoscopiques en utilisant l'approche deep learning avec les réseaux de neurones convolutionnels (CNN). Nous avons implémenté trois modèles, dont deux sont pré-entraînés (VGG16 et VGG19), tandis que le troisième a été développé spécifiquement pour cette étude. Nous avons comparé les trois modèles et noté l'influence des différents composants et paramètres dans le processus de classification, notamment la couche de convolution, la couche de pooling et la couche fully connected. Les résultats obtenus démontrent que le modèle VGG16 est le plus performant, avec une précision de **99,9 %** et une erreur de seulement **0,1 %**.

**Mots clés:**

Laparoscopie, fumée, deep learning, CNN, classification, VGG16, VGG19.

---

## Abstract

---

Laparoscopic surgery has obvious advantages for patients, such as short hospital stay. However, this technique is complex and can easily lead to complications. One of the sources of these complications is impaired vision due to smoke contamination of the camera optics, which will introduce a period of time without clear vision. It is therefore crucial to quickly restore optimal vision to improve the surgical prognosis. An obvious solution would be to automatically detect contaminated images. We therefore explored the detection and classification of smoke in laparoscopic images using the deep learning approach with convolutional neural networks (CNN). We implemented three models, two of which are pre-trained (VGG16 and VGG19), while the third was developed specifically for this study. We compared the three models and noted the influence of different components and parameters in the classification process, including the convolution layer, the pooling layer and the fully connected layer. The results obtained show that the VGG16 model is the most efficient, with an accuracy of 99.9% and an error of only 0.1%.

### **Keywords**

Laparoscopy, smoke, deep learning, CNN, classification, VGG16, VGG19.

## ملخص

للجراحة بالمنظار مزايا واضحة للمرضى ، مثل الإقامة القصيرة في المستشفى. ومع ذلك ، فإن هذه التقنية معقدة ويمكن أن تؤدي بسهولة إلى مضاعفات. ومن مصادر هذه المضاعفات ضعف الرؤية نتيجة تلوث دخان عدسات الكاميرا ، الأمر الذي سيدخل فترة من الزمن بدون رؤية واضحة. لذلك من الضروري استعادة الرؤية المثلى بسرعة لتحسين التشخيص الجراحي. سيكون الحل الواضح هو الكشف التلقائي عن الصور الملوثة. لذلك بحثنا في كيفية تصنيف الدخان في الصور بالمنظار باستخدام نهج التعلم العميق مع الشبكات العصبية التلافيفية (CNN). قمنا بتنفيذ ثلاثة نماذج ، اثنان منها تم تدريبهما مسبقاً (VGG16 و VGG19) ، بينما تم تطوير النموذج الثالث خصيصاً لهذه الدراسة. قارنا النماذج الثلاثة ولاحظنا تأثير المكونات والمعلومات المختلفة في عملية التصنيف ، بما في ذلك طبقة الالتفاف وطبقة التجميع والطبقة المتصلة بالكامل. أظهرت النتائج التي تم الحصول عليها أن نموذج VGG16 هو الأكثر كفاءة ، حيث بلغت الدقة 99.9٪ وخطأ 0.1٪ فقط.

### الكلمات الرئيسية

جراحة المنظار، الدخان، التعلم العميق، الشبكات العصبونية الالتفافية، التصنيف ، VGG16, VGG19.

# Table des matières

|   |           |
|---|-----------|
| Liste des figures.....  | 9         |
| Liste des tableaux.....                                       | 11        |
| Introduction générale.....                                    | 12        |
| La problématique.....   | 14        |
| <b>Chapitre I : Chirurgie laparoscopique.....</b>             | <b>15</b> |
| I.1 Introduction.....   | 16        |
| I.2 Définition de la chirurgie laparoscopique.....            | 16        |
| I.3 Historique de la laparoscopie.....                        | 17        |
| I.4 Les instruments de la laparoscopie.....                   | 17        |
| I.5 Les indications de laparoscopie.....                      | 22        |
| I.6 Étapes suivies en chirurgie laparoscopique.....           | 22        |
| I.7 Les avantages de la laparoscopie.....                     | 23        |
| I.8 Les inconvénients de la laparoscopie.....                 | 24        |
| I.9 Les limites de la laparoscopie.....                       | 24        |
| I.10 Les complications de la laparoscopie.....                | 25        |
| I.10.1 Les complications peropératoires.....                  | 25        |
| I.10.2 Les complications postopératoires.....                 | 25        |
| I.11 Conclusion.....  | 26        |
| <b>Chapitre II : Classification des images médicales.....</b> | <b>27</b> |
| II.1 Introduction.....  | 28        |
| II.2 Le traitement d'image.....                               | 28        |
| II.2.1 Le Prétraitement.....                                  | 28        |
| II.2.1.1 La normalisation.....                                | 28        |
| II.2.1.2 Le redimensionnement.....                            | 29        |

|  |           |
|--|-----------|
| II.2.2 L'Extraction des caractéristiques.....                            | 29        |
| II.2.2.1 Extraction manuelle des caractéristiques.....                   | 29        |
| II.2.2.2 Extraction automatique des caractéristiques.....                | 29        |
| II.2.3 La classification.....  | 30        |
| II.2.3.1 Définition.....   | 30        |
| II.2.3.2 Principe de classification.....                                 | 30        |
| II.2.3.3 Les phases de classification.....                               | 30        |
| II.2.3.3.1 Phase d'apprentissage.....                                    | 30        |
| II.2.3.3.2 Phase de test.....  | 32        |
| II.2.4 Classification avec apprentissage machine (machine learning)..... | 34        |
| II.3 Les réseaux de neurones convolutionnels.....                        | 35        |
| II.3.1 Définition.....   | 35        |
| II.3.2 Architecture de réseaux de neurones convolutionnels.....          | 36        |
| II.3.2.1 Couche de convolution(CONV).....                                | 36        |
| II.3.2.2 Couche de pooling (POOL).....                                   | 38        |
| II.3.2.3 Couches de correction (RELU).....                               | 39        |
| II.3.2.4 Couche entièrement connectée (FC).....                          | 39        |
| II.3.2.5 Couche de perte (LOSS).....                                     | 39        |
| II.3.3 Les modèles de réseaux de neurones.....                           | 40        |
| II.4 Application aux images médicales : Etat de l'art.....               | 41        |
| II.5 Conclusion.....   | 43        |
| <b>Chapitre III: Méthodes et résultats.....</b>                          | <b>44</b> |
| III.1 Introduction.....  | 45        |
| III.2 Configuration Matérielle.....                                      | 46        |
| III.3 Configuration Logicielle.....                                      | 46        |
| III.3.1 Environnement de programmation .....                             | 46        |

|  |    |
|--|----|
| III.3.1.1 Python.....                                    | 46 |
| III.3.1.2 Keras.....                                     | 46 |
| III.3.1.3 TensorFlow.....                                | 47 |
| III.3.1.4 Open Cv.....                                   | 47 |
| III.3.1.5 Scikit-learn.....                              | 48 |
| III.3.1.6 Numpy.....                                     | 48 |
| III.3.1.7 Matplotlib.....                                | 49 |
| III.3.2 Programmation en ligne.....                      | 49 |
| III.4 La base de données des images laparoscopiques..... | 51 |
| III.5 Architecture de notre réseau.....                  | 53 |
| III.5.1 Architecture du premier modèle.....              | 53 |
| III.5.2 Architecture du deuxième modèle.....             | 54 |
| III.5.3 Architecture du troisième modèle.....            | 55 |
| III.6 Résultats obtenus et discussion.....               | 56 |
| III.6.1 Résultats obtenus par notre modèle .....         | 57 |
| III.6.2 Résultats obtenus par le modèle 2 « VGG16 »..... | 62 |
| III.6.3 Résultats obtenus pour le modèle 3« VGG19 »..... | 68 |
| III.6.4 Comparaison des résultats.....                   | 73 |
| III.7 Conclusion.....                                    | 74 |
| Conclusion générale.....                                 | 75 |
| Références bibliographiques.....                         | 76 |

## Liste des figures

|   |    |
|---|----|
| <b>Figure I.1</b> Schéma de principe de la chirurgie laparoscopique.....  | 17 |
| <b>Figure I.2</b> Image d'une intervention laparoscopique mettant en évidence la vision indirecte des chirurgiens et l'utilisation des instruments laparoscopiques..... | 17 |
| <b>Figure I.3</b> Trocart de 5,5mm Vanne à ouverture manuelle et automatique Pointe pyramidale.....   | 19 |
| <b>Figure I.4</b> Pince à énergie bipolaire de 3 mm, composée de trois parties : la poignée, le tube et l'insert souhaité.....  | 19 |
| <b>Figure I.5</b> Pince à énergie monopolaire de 5 mm.....  | 20 |
| <b>Figure I.6</b> Ciseaux à crochet et ciseaux droits 5 mm.....   | 20 |
| <b>Figure I.7</b> Un porte-aiguille de 5 mm.....  | 20 |
| <b>Figure I.8</b> Electrodes (a : électrode monopolaire ; b : électrode bipolaire de 5 mm.....  | 21 |
| <b>Figure I.9</b> Laparoscope (a : le laparoscope coté prise en main ; b : extrémité distale comportant la caméra).....   | 21 |
| <b>Figure II.1</b> Architecture standard d'un réseau de neurone convolutionnel.....   | 36 |
| <b>Figure II.2</b> Max-Pooling avec un filtre 2x2 et un pas de 2.....   | 38 |
| <b>Figure III.1</b> Organigramme de notre étude .....   | 45 |
| <b>Figure III.2</b> Images issues de la base LVQ (a. image « no_smoke » , b. image « start_smoke » , c. image « smoke »).....   | 52 |
| <b>Figure III.3</b> Architecture du notre modèle .....  | 54 |
| <b>Figure III.4</b> Architecture du modèle 2 «VGG16».....   | 55 |
| <b>Figure III.5</b> Architecture du modèle 3 «VGG19».....   | 56 |
| <b>Figure III.6</b> Précision et Erreur pour notre modèle sur 893 images .....  | 57 |
| <b>Figure III.7</b> Matrice de confusion du notre modèle .....  | 57 |
| <b>Figure III.8</b> Nombre total des images mal et bien classé .....  | 58 |
| <b>Figure III.9</b> Taux d'erreur et de précision .....   | 58 |
| <b>Figure III.10</b> Nombre des images mal et bien classé pour notre modèle .....   | 58 |
| <b>Figure III.11</b> Précision et Erreur pour notre modèle .....  | 59 |

|   |    |
|---|----|
| <b>Figure III.12</b> Matrice de confusion du notre modèle .....                   | 60 |
| <b>Figure III.13</b> Nombre total des images mal et bien classé .....             | 60 |
| <b>Figure III.14</b> Taux d'erreur et de précision .....                          | 60 |
| <b>Figure III.15</b> Nombre des images mal et bien classé pour notre modèle ..... | 61 |
| <b>Figure III.16</b> Précision et Err.eur du VGG16 sur 893 images .....           | 62 |
| <b>Figure III.17</b> Matrice de confusion du VGG16 .....                          | 63 |
| <b>Figure III.18</b> Nombre total des images mal et bien classé .....             | 63 |
| <b>Figure III.19</b> Taux d'erreur et de précision .....                          | 63 |
| <b>Figure III.20</b> Nombre des images mal et bien classé par VGG16 .....         | 64 |
| <b>Figure III.21</b> Précision et Erreur du <b>VGG16</b> sur 9238 images.....     | 65 |
| <b>Figure III.22</b> Matrice de confusion du VGG16 .....                          | 66 |
| <b>Figure III.23</b> Nombre total des images mal et bien classé .....             | 66 |
| <b>Figure III.24</b> Taux d'erreur et de précision.....                           | 66 |
| <b>Figure III.25</b> Nombre des images mal et bien classé par VGG16 .....         | 65 |
| <b>Figure III.26</b> Précision et Erreur du VGG19 sur 893 images.....             | 68 |
| <b>Figure III.27</b> Matrice de confusion du VGG19 .....                          | 68 |
| <b>Figure III.28</b> Nombre total des images mal et bien classé.....              | 69 |
| <b>Figure III.29</b> Taux d'erreur et de précision.....                           | 69 |
| <b>Figure III.30</b> Nombre des images mal et bien classé par <b>VGG19</b> .....  | 69 |
| <b>Figure III.31</b> Précision et Erreur du VGG19 sur 9238 images.....            | 70 |
| <b>Figure III.32</b> Matrice de confusion du VGG19 .....                          | 71 |
| <b>Figure III.33</b> Nombre total des images mal et bien classé .....             | 71 |
| <b>Figure III.34</b> Taux d'erreur et de précision.....                           | 71 |
| <b>Figure III.35</b> Nombre des images mal et bien classé par VGG19 .....         | 72 |

## Liste des tableaux

|   |    |
|---|----|
| <b>Tableau II.1</b> Les définitions des grandeurs VP, VN, FP et FN.....             | 33 |
| <b>Tableau III.1</b> Comparaison des trois modèles (paramétrage et résultats) ..... | 73 |

## Introduction générale

Des avancées technologiques considérables dans le domaine de la santé ont permis, au cours des dernières décennies, à la chirurgie mini-invasive (CMI) de devenir une pratique courante dans la communauté clinique. Cette technique médicale permet de réduire au maximum les traumatismes physiques subis par les patients lors d'une opération chirurgicale classique. Aujourd'hui, certaines interventions chirurgicales sont presque exclusivement réalisées par des méthodes mini-invasives, comme c'est le cas pour la cholécystectomie, une procédure utilisée pour traiter les affections de la vésicule biliaire. Dans de telles situations ou dans des situations similaires, les médecins se fient à la vidéo-surveillance pour traiter l'anatomie interne des patients.

La vidéo-surveillance consiste à introduire une caméra haute définition ou un endoscope en plus d'une variété d'instruments par les orifices corporels.

Ce domaine médical est l'endoscopie. Il est subdivisé en diverses sous spécialités selon le lieu d'insertion du dispositif vidéo qui peut être une ouverture naturelle telle que le nez (rhinoscopie), l'oreille (otoscopie), l'anus (anoscopie), etc., ou une incision délibérément créée pour examiner les cavités intérieures des articulations (arthroscopie), du thorax (thoracoscopie) et de l'**abdomen** qui est la zone la plus fréquemment inspectée. Cette dernière est traitable par un grand nombre de procédures qui composent le domaine de la **laparoscopie**.

La vidéo laparoscopique est la source de données temps réel, sur laquelle s'appuie le chirurgien pendant son intervention. Elle doit être fiable pour garantir la continuité du geste chirurgical en toute sécurité. Cependant le geste chirurgical, en question, peut générer des facteurs perturbateurs pour l'optique laparoscopique comme c'est le cas de certaines projections de liquides corporels comme le sang qui rend l'optique opaque et donc inexploitable.

C'est également le cas de la fumée qui est générée en raison de l'utilisation de l'électrocautérisation. Tous ces événements, en plus du fait qu'ils perturbent la vision endoscopique, doivent être pris en charge le plus rapidement possible pour sécuriser le geste chirurgical (la vue c'est la vie), et pour optimiser le temps global de l'intervention qui est directement lié à la durée de rétablissement du patient.

La rapidité de prise en charge de ces événements est conditionnée par la rapidité de détection. Autrement dit, l'automatisation de détection de ces sources de bruits, est une phase clé pour assurer le bon déroulement de l'intervention laparoscopique.

Les techniques de détection et de classification proviennent de deux principales approches : la première est basée sur la segmentation, l'extraction d'attributs et la classification, tandis que la seconde repose sur l'intelligence artificielle, plus précisément l'apprentissage automatique «machine learning», qui combine l'extraction d'attributs et la classification en utilisant généralement des réseaux de neurones convolutionnels (CNN) pour la classification.

Dans ce projet de recherche nous nous intéressons à la détection automatique de la fumée dans les vidéos laparoscopiques issues de bases de données en appliquant des algorithmes de la catégorie CNN avec l'ajustement des hyperparamètres. Cette détection permet de classer les images en trois degrés de gravité du bruit. Cette classification permettra de commander le système pertinent tel que l'aspirateur de fumée.

Ce mémoire est organisé en trois chapitres. :

- ❖ Dans le premier chapitre on va consacrer à une introduction complète à la chirurgie laparoscopique, les principaux aspects de cette technique chirurgicale, en mettant l'accent sur les instruments utilisés, les étapes suivies en cette intervention, ainsi que les avantages, les limites et les complications per et post-opératoires.
- ❖ Dans le deuxième chapitre on va explorer le rôle essentiel de la classification et du deep learning dans le domaine de l'imagerie médicale et on examinera l'utilisation des réseaux de neurones, en mettant l'accent sur les réseaux de neurones convolutionnels, pour améliorer la classification des images médicales.
- ❖ Dans le troisième chapitre on va présenter la partie expérimentale de notre travail, où on exposera les résultats obtenus, suivis d'une discussion approfondie et en dernier lieu, on clôturera par une conclusion générale.

## La problématique

En laparoscopie la **vision** du chirurgien est indirecte et se fait par le biais d'une micro-caméra (interne) connectée à un moniteur de visualisation externe. La vision **doit être nette** sans bruits pour garantir la précision et la continuité du geste chirurgical. L'opération chirurgicale fait intervenir des outils mécaniques et parfois électriques. C'est le cas du bistouri électrique qui peut provoquer la formation de fumée dans la cavité abdominale. De manière générale n'importe quelle source de chaleur peut entraîner l'apparition de fumée. La fumée ainsi dégagée, sans tenir compte de son lien avec l'état des tissus cibles, est un bruit d'image qu'il faut supprimer.

Contrairement aux techniques de traitement d'image qui suppriment le bruit de manière logicielle, cette suppression doit être réelle avec l'usage de pompe d'aspiration... Plus vite la prise en charge de cet événement (dégagement de fumé) est faite meilleure sera le pronostique chirurgical. Par conséquent, il est essentiel de trouver des **solutions automatiques** pour **détecter et classier la fumée** dans les vidéos laparoscopiques afin d'améliorer la vision du chirurgien et réduire les risques de mauvaise manœuvre. De plus, la fumée peut contenir des particules nocives (virus...) pour la santé du personnel médical, ce qui ajoute une dimension supplémentaire à la problématique. Ainsi, la détection et la classification de fumée dans la chirurgie est un enjeu important pour la sécurité et la qualité des interventions chirurgicales laparoscopiques.

---

# Chapitre I

---

**Chirurgie laparoscopique**

## I.1 Introduction

En laparoscopie, contrairement à la chirurgie classique à ventre ouvert, la vision du chirurgien est restreinte au dispositif de visualisation du système laparoscopique c'est-à-dire à l'ensemble source de lumière, micro-caméra et moniteur de visualisation. Dans ce projet de recherche, nous nous intéressons à l'analyse des vidéos laparoscopiques et à leur classification en trois niveaux de gravité du bruit « fumée ». La fumée est considérée comme un bruit car elle perturbe la vision du site exploré et empêche la continuité du geste chirurgical. Pour bien aborder cette problématique, il est important de connaître l'environnement de travail à partir duquel sont issues ces vidéos.

Le domaine médical correspondant est nommé laparoscopie. C'est une technique chirurgicale mini-invasive qui utilise des instruments spécifiques et une caméra pour visualiser l'intérieur du corps du patient. Nous allons donc explorer les particularités de cette technique chirurgicale, ainsi que les paramètres qui peuvent influencer la qualité des images capturées lors d'une intervention laparoscopique.

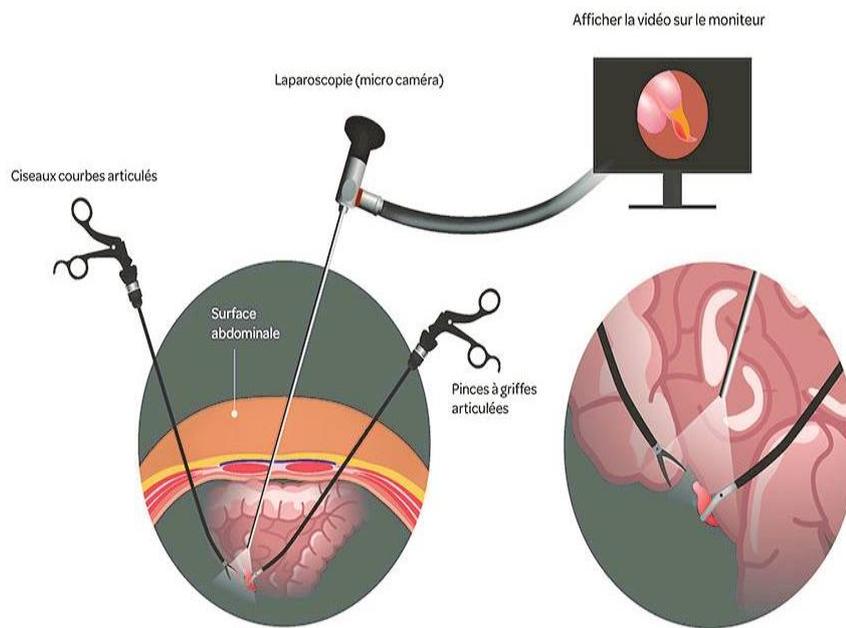
## I.2 Définition de la chirurgie laparoscopique

La laparoscopie, également connue sous le nom de « cœlioscopie », est une technique chirurgicale relativement récente qui a connu une forte expansion au cours des dernières années. Elle implique la réalisation de petites incisions dans la paroi abdominale afin de permettre l'accès du chirurgien à la cavité abdominale ou pelvienne et aux organes qu'elle renferme.

En pratique, la laparoscopie est réalisée sous anesthésie générale. Le chirurgien pratique une ou plusieurs petites incisions, mesurant quelques centimètres, dans la paroi abdominale. Ensuite, il introduit une aiguille ou un petit tube creux sous l'ombilic pour insuffler du gaz carbonique. Le gaz permet de soulever la paroi abdominale et de repousser les intestins, afin de permettre la visualisation de l'intérieur de la cavité abdominale. Différents instruments peuvent être ensuite insérés dans la cavité abdominale à travers ces incisions (**B.Estelle, 2018**) :

Différents instruments peuvent être ensuite insérés dans la cavité abdominale à travers ces incisions :

- Une caméra miniature mesurant quelques millimètres de diamètre (appelée laparoscope) connectée à un écran externe, permettant au chirurgien d'observer les organes et l'intérieur de la cavité abdominale (figure I.1).
- Des instruments chirurgicaux de petite taille (d'une longueur comprise entre 1 et 3 cm), tels que des pinces, des ciseaux (figure I.1 et I.2).
- Des instruments de coagulation ou de lavage, utilisés pour effectuer différentes tâches, telles que les biopsies, la section de tissus ou l'ablation de tumeurs.



**Figure. I.1** Schéma de principe de la chirurgie laparoscopique.



**Figure. I.2** Image d'une intervention laparoscopique mettant en évidence la vision indirecte des chirurgiens et l'utilisation des instruments laparoscopiques.

### I.3 Historique de la laparoscopie

L'histoire débute en 1806, lorsque le docteur italien Philippe Bozzini effectue la toute première endoscopie vaginale à l'aide d'un spéculum qu'il a lui-même inventé. Il a par la suite perfectionné cette technique en y ajoutant un dispositif de miroirs et de lentilles, ainsi qu'une source lumineuse électrique. Au fil du temps, on remplace le spéculum par deux tubes en argent, et les endoscopes sont munis d'ampoules plus sophistiquées ( Allodocteurs.fr, 2015).

Au début du XXe siècle, la pratique médicale évolue et l'on commence à utiliser des endoscopes éclairants introduits à travers de petites incisions pour explorer en profondeur l'abdomen et le thorax, plutôt que de se limiter aux orifices naturels (János Veress, 2022).

Au cours des années 1940, l'ajout de gaz carbonique est devenu courant pour faciliter l'exploration du corps, et c'est à ce moment-là que le terme "cœlioscopie" a été introduit. En 1955, un gynécologue de Paris nommé Raoul Palmer a tenté pour la première fois une laparoscopie à des fins diagnostiques, en effectuant une biopsie ovarienne.

En 1987, le chirurgien lyonnais Philippe Mouret a effectué la toute première cholécystectomie par cœlioscopie, qui consiste à enlever la vésicule biliaire. À partir de ce moment-là, la coelio-chirurgie est devenue une alternative viable à la chirurgie traditionnelle à ciel ouvert.

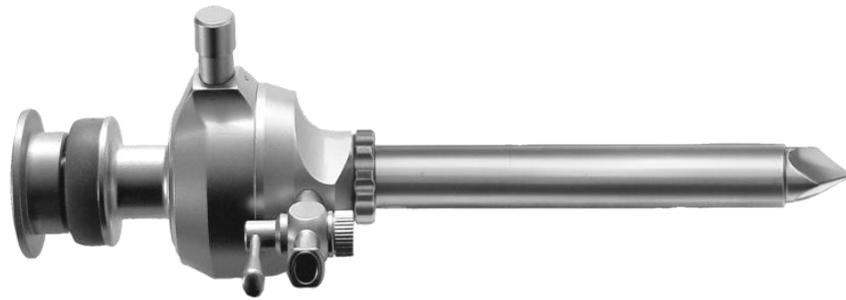
Actuellement, cette opération est pratiquée lorsqu'il y a une inflammation de la vésicule biliaire ou la présence de calculs biliaires. Les chirurgiens introduisent désormais un tube équipé d'un système optique à travers de petites incisions, pour réaliser l'intervention.

### I.4 Les instruments de la laparoscopie

Les instruments utilisés en laparoscopie sont spécialement conçus pour être utilisés à travers de petits incisions dans la paroi abdominale, afin de minimiser les traumatismes aux tissus environnants et de réduire les douleurs post-opératoires.

Certains des instruments les plus couramment utilisés en laparoscopie comprennent (Delmont Imaging) :

**Les trocarts** : sont des tubes métalliques creux insérés à travers les incisions pour permettre l'insertion des autres instruments chirurgicaux dans la cavité abdominale (Figure I. 3).



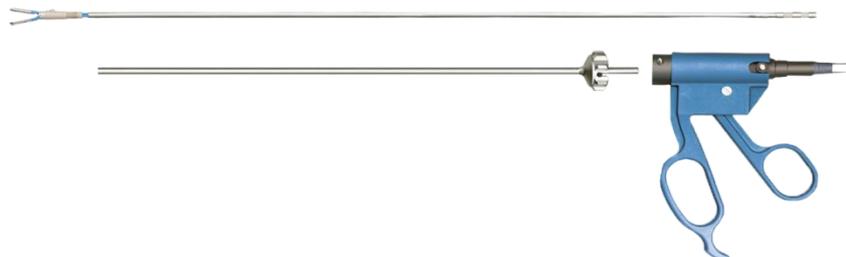
**Figure I .3** Trocart de 5,5mm Vanne à ouverture manuelle et automatique Pointe pyramidale.

**Les pinces** sont utilisées en laparoscopie, pour des fonctions spécifiques telles que la saisie, la dissection ou la ligature des vaisseaux sanguins, il existe de nombreuses pinces différentes :

○ **Une pince à énergie bipolaire** : est un instrument chirurgical utilisé pour couper et sceller les vaisseaux sanguins ou les tissus pendant une intervention chirurgicale (figure I. 4). En laparoscopie, la pince à énergie bipolaire est souvent utilisée pour des procédures chirurgicales minimales invasives, telles que la colectomie laparoscopique (ablation chirurgicale d'une partie du côlon). Les avantages de l'utilisation d'une pince à énergie bipolaire en laparoscopie sont qu'elle réduit le saignement, réduit les dommages aux tissus environnants et accélère la cicatrisation des plaies. En outre, l'utilisation de la pince à énergie bipolaire peut réduire la durée de l'intervention chirurgicale et le temps de récupération post-opératoire.

○ **Une pince à énergie monopolaire** : Une pince à énergie monopolaire est un instrument chirurgical qui utilise un seul point d'énergie pour couper et cautériser les tissus pendant une intervention chirurgicale. Ce type de pince utilise un point d'énergie à haute fréquence pour couper les tissus, puis utilise le même point d'énergie pour cautériser les vaisseaux sanguins afin de réduire les saignements (Figure I. 5).

En laparoscopie, la pince à énergie monopolaire peut être utilisée pour couper et cautériser les tissus à travers de petites incisions dans la paroi abdominale. Cependant, l'utilisation de la pince à énergie bipolaire est souvent préférée car elle présente des avantages supplémentaires.

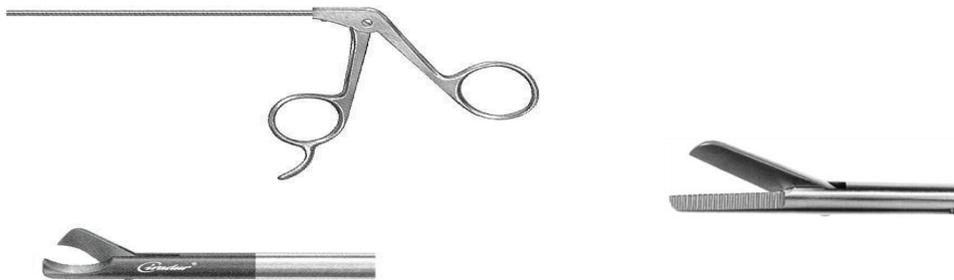


**Figure I. 4** Pince à énergie bipolaire de 3 mm, composée de trois parties : la poignée, le tube et l'insert souhaité.



**Figure I. 5** Pince à énergie monopolaire de 5 mm

**Les ciseaux** : il existe différents types de ciseaux utilisés en laparoscopie, tels que les ciseaux droits, courbes ou à crochet, qui sont utilisés pour couper et disséquer les tissus.



**Figure I. 6** Ciseaux à crochet et ciseaux droits 5 mm.

**Le porte-aiguille** : est un instrument chirurgical qui est utilisé pour tenir une aiguille pendant une intervention chirurgicale. Il est constitué de deux branches qui se croisent et se ferment en pince pour maintenir fermement l'aiguille (figure I.7).



**Figure I. 7** Un porte-aiguille de 5 mm.

**Les électrodes** : ce sont des instruments qui utilisent l'énergie électrique pour couper et cautériser les tissus.



**a.**



**b.**

**Figure I. 8** Electrodes (a : électrode monopolaire ; b : électrode bipolaire de 5 mm).

**Les laparoscopes** : est un instrument chirurgical utilisé pour visualiser l'intérieur de la cavité abdominale et réaliser des interventions chirurgicales minimales invasives. Il se compose d'un tube fin et allongé équipé d'une caméra vidéo et d'un système d'éclairage qui permet au chirurgien de voir l'intérieur de la cavité abdominale sur un écran vidéo (Société canadienne du cancer) (figure I.9).



**a.**

**b.**

**Figure I. 9** Laparoscope (a : le laparoscope coté prise en main ; b : extrémité distale comportant la caméra).

## I.5 Les indications de la laparoscopie

La laparoscopie est une technique chirurgicale qui permet d'explorer et d'intervenir sur les organes de la cavité abdominale. Elle se divise en deux types (**Estelle B. , 2018**) :

1. **La laparoscopie la laparoscopie exploratrice** est une procédure médicale utilisée pour examiner les organes internes de la cavité abdominale. Elle permet au chirurgien d'obtenir une vision directe des organes tels que l'estomac, les intestins, le foie et la vésicule biliaire. La laparoscopie exploratrice est utilisée à des fins diagnostiques ou pour des interventions chirurgicales mineures, offrant des avantages tels que des incisions plus petites, une récupération plus rapide et moins de douleur postopératoire.

2. **La laparoscopie thérapeutique ou opératoire** est une procédure chirurgicale utilisée pour traiter des affections spécifiques de la cavité abdominale telles que l'ablation de tumeurs, la réparation de hernies ou la réalisation de procédures gynécologiques. Cependant, sa faisabilité dépend de la complexité de l'affection et de l'expertise du chirurgien.

Par conséquent, la laparoscopie peut être utilisée pour (Chalret Du Rieu, 2022):

- En **gynécologie-obstétrique**, elle est utilisée pour examiner les voies génitales et vérifier leur perméabilité. Elle permet également de diagnostiquer des infections et des troubles tels que l'endométriose, de traiter les adhérences, d'effectuer une hystérectomie, d'enlever des kystes ovariens, ainsi que de réaliser des curetages utérins.

- En **chirurgie uro-digestive** (pratique chirurgicale pour les systèmes urogénital et digestif), cette méthode peut être utilisée pour réaliser diverses interventions, telles que la prostatectomie (enlèvement de la prostate), la splénectomie (ablation de la rate), la cholécystectomie (ablation de la vésicule biliaire), ainsi que pour traiter les hernies inguinales ou hiatales.

- En **chirurgie digestive**, cette méthode offre la possibilité de réaliser pratiquement toutes les interventions qui étaient auparavant effectuées en ouvrant la cavité abdominale, telles que l'ablation de la vésicule biliaire, du colon ou de la rate. Elle est également considérée comme la méthode de référence pour la réalisation de la chirurgie de l'obésité ou chirurgie bariatrique, incluant la sleeve gastrectomie, le bypass gastrique.

- En **médecine d'urgence**, cette méthode peut être utilisée pour traiter des pathologies telles qu'une grossesse extra-utérine, une appendicite ou encore une péritonite.

## I.6 Etapes suivie en chirurgie laparoscopique

La durée de la cœlioscopie dépend de la complexité de l'opération et peut s'étendre de 30 minutes à plusieurs heures, et elle implique plusieurs étapes (Ameli.fr, 2019) :

1. Après votre arrivée dans la chambre, vous prenez une douche avec un produit antiseptique qui permet d'éliminer les bactéries, champignons et virus ou d'en empêcher la reproduction, contribuant ainsi à la prévention et au traitement des infections (**URPS Infirmière PACA, 2019**). Par la suite, vous recevez un tranquillisant pour vous détendre avant l'intervention, et

vous enflevez une tenue stérile composée d'une blouse, d'un bonnet et de chaussons avant d'entrer dans la salle d'opération.

2. Vous êtes transféré dans une salle d'opération équipée d'appareils de surveillance de l'anesthésie, où l'anesthésie est administrée. Pendant que vous êtes sous anesthésie générale, un tube fin est inséré dans votre trachée (procédure appelée intubation) pour garantir une ventilation sûre. De plus, une sonde urinaire est introduite dans votre vessie afin de maintenir une vessie vide pendant la laparoscopie, facilitant ainsi l'observation de la cavité abdominopelvienne.

3. Le chirurgien effectue une incision d'environ un centimètre, généralement située dans la région du nombril. À travers cette petite ouverture, il introduit une aiguille fine et creuse dans la cavité abdominale. En insufflant du gaz carbonique à travers l'aiguille, la paroi abdominale est soulevée pour faciliter la visualisation des organes internes.

4. Une caméra miniature et une lampe de forte intensité sont intégrées dans un dispositif optique qui est inséré dans l'abdomen à travers un trocart (une tige creuse) utilisant le même orifice. Cette caméra permet d'obtenir un effet de zoom, améliorant ainsi la précision et la visibilité du geste chirurgical. Les images en temps réel de l'intérieur de la cavité abdominale sont transmises à un écran visible par le chirurgien et son équipe.

5. Le cas échéant, le chirurgien peut réaliser jusqu'à quatre petites incisions supplémentaires, qui mesurent entre 0,5 et 1,2 cm de longueur. Ces incisions lui permettent de faire passer des instruments spécifiques, fins et longs, tels que des ciseaux, des pinces, des crochets, des porte-aiguilles et du matériel de coagulation dans la cavité abdominale.

6. Une fois l'opération terminée, le gaz carbonique est évacué et les petites incisions cutanées sont refermées pour assurer une guérison optimale.

## **I.7 Les avantages de la laparoscopie**

### **-Avantage esthétique**

La laparoscopie présente un avantage esthétique car elle ne laisse généralement pas de cicatrices importantes sur l'abdomen. Les petites incisions utilisées lors de cette technique permettent d'avoir des cicatrices discrètes, voire invisibles une fois guéries. Cela peut contribuer à la satisfaction des patients en évitant des cicatrices visibles et en préservant leur apparence physique (**Estelle B. , 2018**).

### **-Précision du geste**

La précision du geste est un avantage de la laparoscopie, une technique chirurgicale qui utilise une caméra pour visualiser directement la zone opérée. Cette visualisation en temps réel permet aux chirurgiens d'avoir une meilleure précision lors de l'intervention. Ils peuvent observer avec précision les structures anatomiques et les manipuler avec plus de contrôle, ce qui peut conduire à des résultats chirurgicaux plus précis. Cette précision accrue peut être particulièrement bénéfique lors de procédures délicates ou complexes (**Laurence, 2021**).

### - Intervention moins douloureuse

L'intervention moins douloureuse de la cœlioscopie se traduit par des suites opératoires plus confortables pour les patients, avec moins de douleur et des hospitalisations plus courtes. Cela peut contribuer à une meilleure qualité de vie postopératoire et à un rétablissement plus rapide.

### - Risques infectieux plus limités

En préservant la paroi abdominale, la laparoscopie réduit les incisions, limitant ainsi la surface exposée aux germes, ce qui diminue le risque d'infection. De plus, des mesures supplémentaires comme l'utilisation d'antiseptiques pour nettoyer la peau du patient et les instruments chirurgicaux, ainsi que l'utilisation de draps stériles, assurent un environnement stérile pendant l'intervention chirurgicale. Ces précautions contribuent à minimiser l'entrée de germes dans la cavité abdominale, réduisant ainsi les risques infectieux (**Laurence, 2021**).

### - Moindre coût

La technique chirurgicale mini-invasive nécessite souvent plus de matériel, mais offre des avantages économiques significatifs. Elle réduit la durée d'hospitalisation des patients, ce qui réduit les coûts liés à l'hébergement et aux soins médicaux. De plus, la reprise plus rapide d'une activité professionnelle permet de diminuer les pertes de revenus, ce qui en fait une option moins coûteuse pour la société (**Estelle B. , 2018**).

## I.8 Les inconvénients de la laparoscopie

Il y a quelques inconvénients associés à la laparoscopie. Les principaux inconvénients sont (**Ducrottoy**):

- Il est impératif que l'opérateur possède une expertise avérée dans cette technique.
- Le matériel utilisé est très complexe et sophistiqué.
- Les complications opératoires peuvent parfois être graves, telles que des lésions de l'intestin, des voies urinaires ou des vaisseaux sanguins, notamment lors de l'insertion des tubes (trocarts) dans l'abdomen.

## I.9 Les limites de la laparoscopie

Il y a peu de contre-indications à la réalisation d'une laparoscopie. Les rares contre-indications sont liées à l'état de santé du patient, une maladie qui rendrait l'anesthésie générale dangereuse ou un jeune âge du patient (car l'innocuité de la laparoscopie n'a pas été démontrée pour les enfants de moins de 3 ans) (**Docteur Benchimol, 2021**).

Les contre-indications peuvent inclure l'insuffisance respiratoire, l'insuffisance cardiaque et l'obésité morbide (IMC > 40). Cependant, cette dernière était auparavant considérée comme une contre-indication, mais elle est maintenant considérée comme une indication favorable à la cœlioscopie (moins de risques infectieux), à condition que certaines précautions anesthésiques soient prises.

## **I.10 Les complications de la laparoscopie**

### **I.10.1 Les complications peropératoires**

La procédure de laparoscopie comporte le risque d'une insufflation d'air qui peut se produire à l'extérieur ou à l'intérieur de la cavité abdominale, ainsi que dans les vaisseaux sanguins (embolie gazeuse) ou sous la peau (emphysème) (**Laurence, 2021**).

L'embolie gazeuse se produit lorsqu'il y a une brèche vasculaire ou une intervention dans un organe très vascularisé, permettant ainsi au dioxyde de carbone (CO<sub>2</sub>) de passer dans la circulation sanguine.

Le mauvais positionnement de l'insufflateur ou une manipulation accidentelle peut causer un emphysème sous-cutané, qui se manifeste par des crépitations sous-cutanées neigeuses au niveau cervical. Bien que relativement fréquent, ce phénomène n'est généralement pas grave et se résorbe en quelques heures. Cependant, il peut entraîner des douleurs persistantes après l'intervention chirurgicale.

Une pression intra-abdominale excessive ou une intervention chirurgicale proche de la jonction entre le diaphragme et l'œsophage peut également entraîner un risque de pneumothorax ou de pneumomédiastin.

L'utilisation des instruments électriques pour coaguler ou couper les tissus, peut aussi causer des complications car ils génèrent de **la fumée**, qui contient des particules, des vapeurs et des substances chimiques potentiellement nocives. La fumée peut **obstruer la vision** des chirurgiens et rendre difficile la visualisation claire de la zone opérée.

Voici quelques complications liées à une vision altérée pendant une laparoscopie :

- Difficulté à visualiser les structures anatomiques : Lorsque la fumée ou le brouillard interfèrent avec la vision, les chirurgiens peuvent avoir du mal à identifier et à visualiser clairement les structures internes, ce qui peut augmenter le risque d'erreurs chirurgicales.
- Lésions des tissus adjacents : Une vision altérée peut entraîner des erreurs de manipulation des instruments chirurgicaux, ce qui peut entraîner des lésions des tissus environnants, tels que les vaisseaux sanguins, les organes ou les nerfs.
- Augmentation du temps opératoire : Lorsque la vision est altérée, les chirurgiens peuvent devoir prendre plus de temps pour réaliser l'intervention, car ils doivent procéder avec prudence et effectuer des ajustements supplémentaires pour compenser la perte de visibilité.

### **I.10.2 Les complications postopératoires**

Une complication courante lors d'une intervention chirurgicale est l'hémorragie, qui peut survenir en raison d'une hémostase insuffisante ou d'un relâchement des sutures.

La douleur après l'intervention chirurgicale est causée par trois éléments (**Institut Mutualiste Montsouris**):

- La douleur pariétale après la chirurgie est provoquée par une irritation ou une inflammation de la séreuse péritonéale causée par le dioxyde de carbone (CO<sub>2</sub>).

- La douleur viscérale après l'intervention chirurgicale est causée par des lésions tissulaires directes telles que la dissection tissulaire, l'électrocoagulation, la ligature et les points d'insertion des trocarts. Cette douleur peut être facilement maîtrisée par les antalgiques.

-La douleur ressentie au niveau de l'épaule est courante et peut être causée par la présence de gaz carbonique résiduel entre la coupole diaphragmatique droite et le dôme hépatique. Ce gaz peut provoquer une tension du foie sur les ligaments phréno-hépatiques, ce qui peut être douloureux. Pour éviter cette douleur postopératoire, il est donc essentiel de procéder à une exsufflation minutieuse du pneumopéritoine à la fin de l'intervention chirurgicale. Cette procédure doit être effectuée de manière systématique.

## **I.11 Conclusion**

En conclusion, la chirurgie laparoscopique est une avancée majeure dans le domaine médical, offrant de nombreux avantages par rapport à la chirurgie ouverte traditionnelle. Cette procédure utilise des instruments spéciaux et un laparoscope pour réaliser des interventions précises avec des incisions réduites. Les avantages de la chirurgie laparoscopique comprennent une récupération plus rapide, une diminution de la douleur postopératoire, de meilleures cicatrices et une hospitalisation plus courte.

Cependant, malgré ces avantages, il est important de reconnaître qu'il existe des inconvénients et des complications potentielles associées à la chirurgie laparoscopique, en particulier en ce qui concerne la formation de fumée.

En mettant en place des mesures de sécurité appropriées, il est possible de réduire les risques et d'assurer la sécurité du personnel médical et des patients lors de ces interventions avancées. La recherche continue et l'innovation dans ce domaine sont essentielles pour améliorer encore la sécurité et les résultats de la chirurgie laparoscopique.

---

# Chapitre II

---

**Classification des images médicales**

## **II.1 Introduction**

L'imagerie médicale est une discipline en constante évolution, avec des avancées technologiques rapides qui ont permis d'améliorer la précision et l'efficacité des diagnostics et des traitements. L'une de ces avancées majeures est l'introduction de l'intelligence artificielle et du **deep learning** dans l'analyse et la classification des images médicales. Ces technologies révolutionnaires ouvrent de nouvelles perspectives pour une prise de décision médicale plus précise et bien éclairée.

Dans ce chapitre et après une brève introduction au traitement d'image, nous explorerons le domaine de la classification par réseaux de neurones, en mettant l'accent sur les réseaux de neurones convolutionnels. Il s'agit de familles d'algorithmes appartenant au domaine **du deep learning** ou apprentissage profond et dont le succès ne cesse de croître dans tous les domaines de vision par ordinateur et en particulier en imagerie médicale.

## **II.2 Le traitement d'image**

Le traitement d'images désigne une discipline de l'informatique et des mathématiques appliquées qui étudie les images numériques et leurs transformations, dans le but d'améliorer leur qualité ou d'en extraire de l'information.

Dans le contexte de la vision artificielle, le traitement d'images se place après les étapes d'acquisition et de numérisation, assurant les transformations d'images et la partie de calcul permettant d'aller vers une interprétation des images traitées pour améliorer sa qualité, extraire des informations pertinentes ou la rendre compatible avec un modèle de classification spécifique (**Boukhlof, 2005**).

Le traitement d'images est généralement divisé en plusieurs étapes, qui peuvent varier selon l'application.

### **II.2.1 Le Prétraitement**

Le prétraitement est une étape importante dans le traitement d'images, qui vise à préparer les images pour la segmentation et l'extraction des caractéristiques. Le prétraitement comprend plusieurs étapes, notamment le filtrage, la normalisation et le redimensionnement.

#### **II.2.1.1 La normalisation**

La normalisation est une étape essentielle dans le traitement des données avant leur utilisation dans des techniques d'apprentissage machine, y compris dans le domaine du traitement d'images et de la classification. La normalisation vise à mettre les données dans une plage commune, en ajustant leurs valeurs pour qu'elles aient une distribution standardisée.

En imagerie médicale, les images sont souvent obtenues à partir de diverses sources, présentant des caractéristiques et des plages de valeurs différentes. La normalisation des images

consiste à ajuster les valeurs des pixels de manière à les ramener dans une plage spécifique ou à modifier leur distribution statistique. L'objectif de la normalisation est de réduire les différences entre les images et de s'assurer que les algorithmes de classification et d'analyse les traitent de manière cohérente.

### **II.2.1.2 Le redimensionnement**

Le redimensionnement en traitement d'images est une étape cruciale du prétraitement qui consiste à modifier la taille ou les dimensions d'une image. Cette technique est utilisée pour adapter les images à des formats spécifiques, à des exigences de mémoire ou à des paramètres de traitement, ou pour les aligner avec d'autres images dans un ensemble de données.

Le redimensionnement peut être effectué en augmentant ou en diminuant la résolution de l'image, et il est couramment utilisé dans divers domaines, y compris l'imagerie médicale.

## **II.2.2 L'Extraction des caractéristiques**

L'étape d'extraction des caractéristiques est l'une des étapes les plus importantes du traitement d'images. Elle consiste à extraire des informations significatives à partir des images afin de faciliter leur analyse et leur interprétation. Cette étape est souvent utilisée en combinaison avec la segmentation pour extraire les régions d'intérêt.

Il existe deux types d'extraction des caractéristiques : manuelle et automatique :

### **II.2.2.1 Extraction manuelle des caractéristiques**

L'extraction de caractéristiques implique la sélection manuelle des caractéristiques pertinentes par l'utilisateur. Cette méthode est souvent utilisée lorsque les caractéristiques d'intérêt sont connues à l'avance, par exemple dans le cas d'une tumeur solide où la forme, la taille et la texture sont des caractéristiques importantes. L'extraction manuelle des caractéristiques peut nécessiter une segmentation de l'image car elle peut faciliter l'extraction des caractéristiques dans ces régions spécifiques et peut améliorer la précision de l'extraction des caractéristiques en réduisant la complexité. Cela peut nécessiter une expérience considérable de la part de l'utilisateur (Zohra, 2015).

### **II.2.2.2 Extraction automatique des caractéristiques**

L'extraction automatique des caractéristiques est une étape du processus de traitement d'image qui consiste à extraire automatiquement des informations significatives, sans nécessiter une intervention humaine directe (Miguel Molina-Moreno, ACME, 2022).

Cette étape utilise des algorithmes de traitement d'image, notamment des réseaux de neurones convolutifs (CNN) (Li & Zhu, 2021) dans le cadre de l'apprentissage profond (Deep Learning), pour extraire les caractéristiques pertinentes de l'image, ces caractéristiques extraites peuvent être des contours, des textures, des motifs, des couleurs ou des formes, qui sont ensuite utilisées pour effectuer des tâches de classification, de détection d'objets, de reconnaissance de formes.

L'extraction automatique des caractéristiques permet de réduire le temps et l'effort nécessaires pour traiter les images, tout en augmentant la précision et la robustesse des résultats obtenus.

En conclusion, l'étape d'extraction des caractéristiques est une étape importante du traitement d'images médicales. Elle permet d'extraire des informations significatives à partir des images pour faciliter leur analyse et leur interprétation. L'extraction manuelle et automatique des caractéristiques sont des approches complémentaires qui peuvent être utilisées en fonction du contexte et des caractéristiques d'intérêt.

### **II.2.3 La classification**

#### **II.2.3.1 Définition**

La classification d'image peut être définie comme le processus consistant à affecter des étiquettes préétablies aux différents éléments d'une image. Cela implique l'utilisation d'un algorithme capable d'attribuer des étiquettes à des groupes de pixels ou à des vecteurs, en fonction des règles définies au préalable. La classification d'images est une tâche essentielle dans la fouille de données pour extraire des connaissances à partir de bases de données.

#### **II.2.3.2 Principe de classification**

La classification implique l'attribution d'objets, qui sont caractérisés par des variables descriptives, à des classes prédéfinies. Pour accomplir cette tâche, il est nécessaire d'utiliser un classifieur, qui peut être un algorithme ou une procédure chargée de réaliser cette attribution. La précision de la classification dépend largement de la pertinence de la représentation des objets. Par conséquent, l'étape d'extraction des descripteurs, qui consiste à préparer la base de données d'exemples, est d'une importance cruciale. La qualité de cette base de données a un impact direct sur l'apprentissage et les performances des classifieurs.

#### **II.2.3.3 Les phases de classification**

La classification d'une classe peut reposer soit sur une description purement descriptive, soit sur une interprétation. Le processus de classification comprend l'utilisation d'un classifieur qui se déroule en deux phases distinctes : la phase d'apprentissage et la phase de test (**Zohra, 2015**).

##### **II.2.3.3.1 Phase d'apprentissage**

L'objectif de l'apprentissage consiste à identifier les règles qui gouvernent les formes à classifier, en utilisant un processus computationnel. Ce processus vise à générer des prédictions précises et à généraliser ces prédictions à de nouveaux exemples.

On distingue trois types d'apprentissage courants : l'apprentissage supervisé, l'apprentissage non-supervisé et l'apprentissage par renforcement.

## **a. Apprentissage supervisé**

On utilise couramment l'apprentissage supervisé pour classifier les données. Le principe de cet apprentissage est simple : le classifieur est alimenté avec un grand nombre d'exemples pour lesquels la sortie souhaitée est connue. Les paramètres d'apprentissage sont ensuite ajustés pour minimiser l'erreur du classifieur, c'est-à-dire la différence entre la sortie désirée et la sortie réelle du classifieur pour chaque entrée. Le but est de faire correspondre au mieux les exemples et de trouver une représentation mathématique qui permette de généraliser les exemples pour traiter de nouvelles situations non rencontrées auparavant.

En entrée du classifieur, on fournit directement soit des couples d'entrée et de sortie désirée, soit des paramètres descripteurs et leurs classes d'appartenance.

Exemple de classification supervisée : les Support Vecteur Machine (SVM), les réseaux de neurones.

## **b. Apprentissage non-supervisé et auto-organisation**

L'apprentissage non-supervisé consiste à fournir uniquement une base de données en entrée au classifieur. Ce dernier doit alors déterminer ses sorties en fonction des similitudes détectées entre les différentes entrées, à l'aide d'une règle d'auto-organisation. Il n'y a pas de "professeur" dans ce mode d'apprentissage, car le classifieur doit lui-même découvrir les ressemblances entre les éléments de la base de données. Cette méthode est également appelée apprentissage sans supervision.

Des exemples d'algorithmes d'apprentissage proposés en littérature incluent les cartes auto-organisatrices de Kohonen (Zohra, 2015).

## **c. Apprentissage par renforcement**

L'apprentissage par renforcement implique d'apprendre comment agir en associant des actions à des situations spécifiques afin de maximiser une récompense quantitative. Contrairement à d'autres approches, l'apprenant n'est pas directement instruit sur quelle action prendre, mais il doit plutôt découvrir par lui-même quelles actions génèrent le plus de récompenses en les essayant. Dans les cas les plus intéressants, les actions peuvent influencer non seulement les récompenses immédiates, mais aussi les situations futures, et par conséquent, les récompenses à plus long terme. Ces deux caractéristiques, à savoir la recherche par essai-erreur et les récompenses à long terme, sont les aspects les plus importants de l'apprentissage par renforcement.

Contrairement à l'apprentissage supervisé qui requiert un superviseur qui indique au classifieur quelle action est correcte dans une situation donnée, l'apprentissage par renforcement se distingue par l'absence de superviseur pour le réseau. Ce dernier interagit directement avec l'environnement et reçoit un retour quantitatif sur la valeur de ses actions, sans intervention extérieure.

## ***Les modes d'apprentissage : « en ligne » et « hors-ligne »***

Il existe deux principaux modes d'apprentissage : l'apprentissage en ligne et l'apprentissage hors ligne.

L'apprentissage en ligne, également appelé "on-line", consiste à ajuster les paramètres d'apprentissage immédiatement après la présentation d'un objet ou d'un exemple. Cela est généralement le cas dans l'apprentissage par renforcement, où une réponse de l'environnement est nécessaire pour guider l'apprentissage. Dans ce mode, les modifications des paramètres sont effectuées au fur et à mesure de l'interaction avec les données.

D'un autre côté, l'apprentissage hors ligne, également appelé "off-line", implique que le classifieur soit entraîné à partir d'une base d'apprentissage, généralement une collection d'exemples. Les paramètres du classifieur sont adaptés une fois que tous les objets d'apprentissage ont été présentés. Ce mode permet une meilleure acquisition des connaissances car il est basé sur l'ensemble des objets plutôt que sur un seul exemple.

L'apprentissage en ligne peut sembler offrir une progression plus rapide, car chaque forme ou exemple est représenté une seule fois. Cependant, il nécessite généralement plus de données d'apprentissage pour obtenir des résultats significatifs. En revanche, l'apprentissage hors ligne peut fournir des résultats plus solides en utilisant un ensemble complet d'objets d'apprentissage. (Zohra, 2015).

### **II.2.3.3.2 Phase de test**

L'objectif de cette phase est d'attribuer un nouvel objet à l'une des classes en utilisant une règle de décision qui intègre les résultats de la phase d'apprentissage. L'objectif est d'obtenir une estimation aussi précise que possible du comportement du classifieur dans des conditions réelles d'utilisation. Pour cela, des critères classiques tels que les taux de classification et les taux d'erreur sont généralement utilisés. Cependant, d'autres critères tels que la spécificité et la sensibilité fournissent également des informations utiles.

#### **a. Taux de classification**

Les taux de classification et d'erreurs sont utilisés pour évaluer la qualité du classifieur par rapport au problème pour lequel il a été conçu. Ces taux sont évalués en utilisant une base de test qui contient des exemples étiquetés avec leur classe réelle, similaire à ceux utilisés pour l'apprentissage, afin de vérifier les réponses du classifieur.

Afin d'obtenir une estimation fiable du taux de reconnaissance, il est important que le classifieur n'ait jamais utilisé les échantillons de cette base de test pour son apprentissage. De plus, cette base de test doit être suffisamment représentative du problème de classification.

En général, lorsque suffisamment d'échantillons étiquetés sont disponibles, ils sont séparés en deux parties distinctes en respectant les proportions par classe de la base initiale. Une partie est utilisée pour former la base d'apprentissage, tandis que l'autre constitue la base de test. Les performances en termes de taux de classification sont ensuite déterminées en présentant à chaque

exemple de la base de test au classifieur, puis en comparant la classe prédite avec la vraie classe de l'exemple.

On définit le taux de classification correcte comme :

$$Tc = \frac{Vp(i)+Vn(i)}{Vp(i)+Vn(i)+fp(i)+fn(i)} \quad (II.1)$$

**Vp** : Vrai positif. **Vn** : Vrai négatif.

**fp** : Faux positif. **fn**: Faux négatif.

### b. Sensibilité et spécificité

On peut évaluer les performances d'un classifieur en utilisant deux mesures statistiques importantes, à savoir la sensibilité et la spécificité.

La sensibilité est une mesure de la capacité d'un classifieur à identifier correctement les membres d'une classe donnée. Elle correspond au nombre de vrais positifs divisé par la somme des vrais positifs et des faux négatifs :

$$Se(i) = \frac{Vp(i)}{Vp(i)+fn(i)} \quad (II.2)$$

La spécificité est une mesure de la capacité d'un classifieur à identifier correctement les non-membres d'une classe donnée. Elle correspond au nombre de vrais négatifs divisé par la somme des vrais négatifs et des faux positifs :

$$Sp(i) = \frac{Vn(i)}{Vn(i)+fp(i)} \quad (II.3)$$

Les grandeurs VP(i), FN(i), VN(i), FP(i) sont définies dans le tableau II.1:

|                         | Présence d'événement<br>de classe i | Absence d'événement<br>de classe i |
|-------------------------|-------------------------------------|------------------------------------|
| Classification Positive | Vrai Positif VP(i)                  | Faux Positif FP(i)                 |
| Classification Négative | Faux Négatif FN(i)                  | Vrai Négatif VN(i)                 |

**Tableau II.1 les définitions des grandeurs VP, VN, FP et FN**

## II.2.4 Classification avec apprentissage machine (machine learning)

Avant l'avènement de l'apprentissage automatique, les systèmes de reconnaissance d'images se composaient généralement de deux parties distinctes : un extracteur de caractéristiques et un classifieur. L'extracteur de caractéristiques était conçu par des programmeurs (méthode manuelle) et transformait une image en un vecteur de caractéristiques, chaque élément représentant la présence ou l'absence d'un motif simple dans l'image. Ce vecteur était ensuite envoyé au classifieur, qui effectuait une somme pondérée des caractéristiques en utilisant des poids spécifiques à chaque catégorie de classification. Si la somme dépassait un certain seuil, la classe était reconnue. Les poids étaient ajustés pendant l'apprentissage pour améliorer la précision de la classification.

Ces méthodes classiques de classification linéaire, telles que le perceptron ou la régression logistique, ont été développées dans les années 1950 et sont encore utilisées aujourd'hui (**Yann LeCun, 1998**). Cependant ces méthodes ont montré leur difficulté à être appliquées, même pour des tâches en apparence simples comme la classification d'images, la reconnaissance d'objets dans les images ou la reconnaissance vocale. En effet, les données provenant du monde réel telles que les échantillons d'un son ou les pixels d'une image sont complexes, variables et souvent bruitées, rendant ainsi leur traitement manuel très ardu.

Pour rendre ce traitement automatique, il faut que la machine puisse transformer une matrice de nombres représentant les pixels d'une image en une information de haut niveau tel que la présence, dans cette image, d'une tumeur, d'un organe comme le foie ou le cœur et ceci malgré la grande variabilité de l'apparence de ces objets et de leur environnement et en ignorant les bruits de fond.

Il est quasiment impossible d'écrire un programme qui soit capable de fonctionner de manière fiable dans toutes les situations. C'est pour cela que l'apprentissage machine, également appelé apprentissage automatique, est devenu crucial. Les grandes entreprises d'Internet l'utilisent depuis longtemps pour effectuer des tâches telles que la détection de contenus indésirables, la classification de réponses de recherche, la recommandation de produits ou de services et la sélection de contenus pertinents pour chaque utilisateur.

Un système d'apprentissage automatique peut être comparé à une boîte noire qui reçoit un signal en entrée, comme une image scanner et fournit en sortie une catégorie prédite, telle que présence de tumeur. Il s'agit de systèmes de classification ou de reconnaissance de formes.

Dans sa forme la plus utilisée, l'apprentissage machine est supervisé, la machine est entraînée en lui montrant des exemples étiquetés avec leur catégorie, la machine ajuste alors ses paramètres internes pour se rapprocher de la sortie désirée après chaque exemple présenté. Après avoir vu des milliers ou des millions d'exemples étiquetés, la machine peut classifier correctement la plupart des nouveaux exemples. Ce qui est encore plus fascinant, c'est que la machine peut également classifier correctement des images qu'elle n'a jamais vues pendant l'apprentissage (capacité de la machine à généraliser) (**Zakaria, 2017**).

L'apprentissage profond, ou **Deep Learning**, branche de l'intelligence artificielle, est une catégorie de l'apprentissage machine qui permet de tels performances. Il utilise des réseaux de neurones artificiels pour apprendre à partir de grandes quantités de données (big Data) non

structurées et de découvrir des relations complexes et non linéaires entre les différentes caractéristiques des données.

Les réseaux de neurones du deep learning, sont inspirés du fonctionnement du cerveau humain. Ils sont constitués d'un grand nombre de neurones interconnectés, organisés en couches, et qui sont capables de traiter de grandes quantités de données simultanément. Les couches sont organisées en divers modules représentant chacun, une étape de traitement.

Chaque module est doté de paramètres ajustables similaires aux poids des classifieurs linéaires, ce qui permet à l'ensemble du système d'être entraînable. Le système est entraîné de bout en bout, c'est-à-dire que tous les paramètres de tous les modules sont ajustés à chaque exemple pour rapprocher la sortie produite par le système de la sortie désirée. Les modules sont disposés en couches successives, ce qui justifie l'utilisation du terme "profond".

Afin de réaliser l'entraînement du système, il est nécessaire de déterminer comment ajuster chaque paramètre de chaque module de manière à minimiser l'erreur de sortie. Cette opération nécessite le calcul d'un gradient, qui indique la variation de l'erreur de sortie en fonction de la modification d'un paramètre. Le calcul de ce gradient est effectué par une méthode appelée rétropropagation, une technique de calcul dérivé qui a été introduite dans les années 1980 (Zakaria, 2017).

Il existe de nombreux algorithmes de deep learning, chacun ayant ses propres avantages et inconvénients en fonction du type de données à traiter et de l'objectif de l'application. Nous pouvons citer :

- **Le perceptron multicouche (MLP)** : il s'agit d'un réseau de neurones composé d'au moins trois couches : une couche d'entrée, une ou plusieurs couches cachées et une couche de sortie. Chaque neurone est connecté à tous les neurones de la couche suivante.
- **Le réseau de neurones convolutifs (CNN)** : il est principalement utilisé pour la reconnaissance d'images et est composé de plusieurs couches de convolution et de pooling. Les couches de convolution extraient les caractéristiques importantes de l'image, tandis que les couches de pooling réduisent la taille des données en les échantillonnant.
- **Les réseaux de neurones profonds (Deep Neural Networks)** : sont des réseaux similaires aux réseaux MLP mais qui comportent un nombre plus important de couches cachées. Cette augmentation du nombre de couches permet au réseau de détecter des variations subtiles du modèle d'apprentissage, mais peut également favoriser le sur-apprentissage ou le sur-ajustement (« overfitting »).

Nous allons détailler les CNN en raison de leurs robustesses pour la classification et la segmentation d'images médicales (voir état de l'art).

## **II.3 Les réseaux de neurones convolutionnels**

### **II.3.1 Définition**

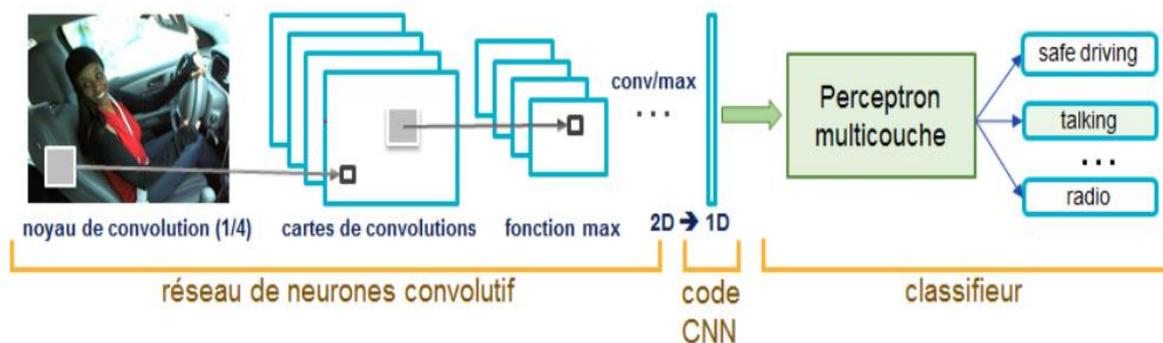
Les réseaux de neurones convolutionnels, également appelés CNN, sont des architectures de réseau de neurones profonds spécialement conçues pour la classification d'images et la

reconnaissance de formes. Ils sont appelés "convolutionnels" car ils utilisent des opérations de convolution pour extraire des caractéristiques clés des données.

### II.3.2 Architecture de réseaux de neurones convolutionnels

Une architecture CNN est formée par un empilement de couches de traitement indépendantes :

- La couche de convolution (CONV) qui traite les données d'un champ récepteur.
- La couche de pooling (POOL), qui permet de compresser l'information en réduisant la taille de l'image intermédiaire (souvent par sous-échantillonnage).
- La couche de correction (ReLU), souvent appelée par abus 'ReLU' en référence à la fonction d'activation (Unité de rectification linéaire).
- La couche "entièrement connectée" (FC), qui est une couche de type perceptron.
- La couche de perte (LOSS).



**Figure II.1** Architecture standard d'un réseau de neurone convolutionnel

#### II.3.2.1 Couche de convolution(CONV) :

La couche de convolution est effectivement un élément clé des réseaux de neurones convolutionnels (CNN). Elle permet de détecter des motifs dans les images en effectuant des opérations de convolution entre des filtres et l'image en entrée (**Belhadjer H, 2018**).

Il existe plusieurs paramètres qui permettent de dimensionner la couche de convolution, notamment :

- **La profondeur** : La profondeur de la couche de convolution d'un CNN correspond au nombre de noyaux de convolution, également appelés filtres, utilisés dans cette couche. Chaque filtre permet de détecter un motif ou une caractéristique spécifique dans l'image d'entrée. Le nombre de filtres détermine donc la diversité des motifs que la couche de convolution peut

détecter. Par exemple, si une couche de convolution utilise 64 filtres, elle sera capable de détecter 64 motifs différents dans l'image d'entrée.

- **Le pas**: qui correspond au nombre de pixels de décalage entre chaque application du filtre sur l'image en entrée. (Contrôle le chevauchement des champs récepteurs). Un pas plus grand réduit la taille de l'image en sortie mais peut potentiellement perdre des informations.

- **La marge**: également appelée "zero padding" en anglais, correspond à l'ajout de zéros autour de l'image en entrée avant d'appliquer la convolution. Cette technique permet de conserver la même taille d'image en sortie, en évitant une réduction de la taille due à l'application du filtre. La taille de la marge est un hyper paramètre qui peut être ajusté pour contrôler la dimension spatiale du volume de sortie.

Lorsque l'on applique une couche de convolution à une image, les paramètres de pas et de marge permettent de contrôler la zone de traitement de l'image en entrée, ce qui affecte la taille du volume de sortie. En outre, la profondeur de la couche correspond au nombre de filtres appliqués à l'image en entrée pour détecter différents motifs. La sortie de chaque filtre est une carte d'activation en 2D. Si l'on prend une image couleur avec trois canaux RGB, chaque canal étant une image en niveaux de gris, on peut considérer que l'image a une profondeur de 3. De la même manière, une couche de convolution prend en entrée un volume avec une certaine profondeur et renvoie en sortie un volume avec une profondeur différente, déterminée par le nombre de filtres appliqués. C'est pourquoi on parle de volume d'entrée et de volume de sortie. L'entrée d'une couche de convolution peut être soit une image soit la sortie d'une autre couche de convolution (**Belhadjer H, 2018**).

Lorsque l'on applique une couche de convolution à un volume d'entrée, la taille spatiale du volume de sortie peut être calculée en fonction de plusieurs paramètres tels que la taille du volume d'entrée ( $W_i$ ), le nombre de champs récepteurs ( $K$ ), le pas ( $S$ ) avec lequel ils sont appliqués et la taille de la marge ( $P$ ).

La formule pour calculer le nombre de neurones du volume de sortie est :

$$W_0 = \frac{W_i - K + 2P}{S} + 1 \quad (\text{II.4})$$

Cette formule nous donne la taille spatiale du volume de sortie en fonction des paramètres de la couche de convolution.

Si le résultat de  $W_0$  n'est pas un nombre entier, cela signifie que les neurones périphériques n'auront pas autant d'entrées que les autres neurones. Dans ce cas, il peut être nécessaire d'augmenter la taille de la marge pour recréer des entrées virtuelles et assurer une couverture uniforme des entrées.

Dans de nombreux cas, le pas  $S$  est défini à 1, ce qui signifie que chaque champ récepteur est appliqué à des positions adjacentes sur le volume d'entrée. Dans ce cas, la marge peut être calculée comme  $P = \frac{K-1}{2}$  pour obtenir un volume de sortie de même taille que le volume d'entrée. Lorsque la marge est ainsi définie, la couche est dite "connectée localement" (**Zakaria, 2017**).

### II.3.2.2 Couche de pooling (POOL) :

Un autre élément clé des CNN est le pooling, qui consiste à sous-échantillonner l'image. Pour cela, l'image d'entrée est découpée en une série de rectangles non chevauchants de  $n$  pixels de côté (pooling). Chaque rectangle peut être considéré comme une tuile et le signal en sortie de chaque tuile est déterminé par les valeurs des pixels qu'elle contient (Belhadjer H, 2018).

Le pooling permet de diminuer la taille d'une image intermédiaire, ce qui réduit le nombre de paramètres et le coût de calcul dans le réseau. Pour cette raison, on insère souvent une couche de pooling entre deux couches convolutives successives d'une architecture CNN pour éviter l'overfitting (sur-apprentissage). En plus de cela, le pooling crée une certaine invariance par translation, ce qui signifie que les motifs détectés par les couches précédentes restent détectés même s'ils se déplacent légèrement dans l'image.

Il existe deux types de pooling :

- **Le maximum pooling :**

Le max pooling est une opération de pooling qui permet de réduire la taille spatiale d'une image en prenant le maximum des carrés de pixels de taille donnée (généralement  $2 \times 2$ ). Cette opération consiste à diviser la taille de l'entrée par 2 en prenant le maximum des valeurs de chaque carré. Le max pooling permet de réduire la quantité de paramètres et de calcul dans le réseau, tout en créant une forme d'invariance par translation (Zakaria, 2017).

- **L'average pooling :**

L'average pooling est une opération de sous-échantillonnage qui réduit la taille spatiale de l'entrée en calculant la valeur moyenne des pixels dans chaque région d'une certaine taille (par exemple,  $2 \times 2$ ). Cette opération permet de réduire le nombre de paramètres et de calculs dans le réseau, tout en conservant une représentation globale de l'image. Contrairement au max pooling, l'average pooling ne conserve pas les valeurs les plus fortes de l'image, mais plutôt une moyenne des valeurs, ce qui peut réduire l'impact des valeurs extrêmes et rendre le modèle plus robuste aux variations de l'image.

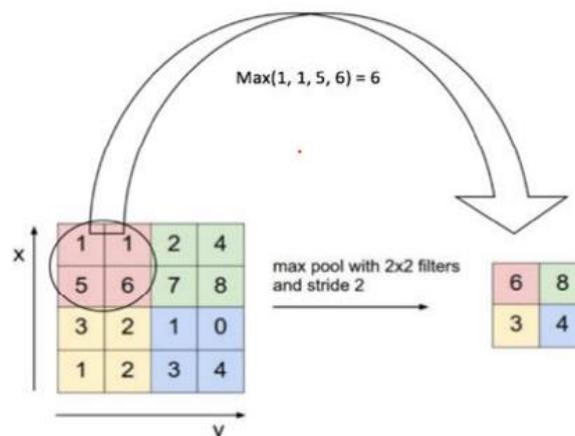


Figure II.2 Max-Pooling avec un filtre  $2 \times 2$  et un pas de 2.

### II.3.2.3 Couches de correction (RELU) :

Les couches de correction, également appelées couches RELU (Rectified Linear Unit), sont des couches non linéaires couramment utilisées dans les architectures de CNN. Leur objectif est de modifier la sortie de la couche convolutive précédente en appliquant une fonction non linéaire simple qui introduit de la non-linéarité dans le modèle.

La fonction RELU applique une transformation sur chaque élément de la carte de caractéristiques de la sortie de la couche convolutive, en remplaçant les valeurs négatives par zéro et en laissant les valeurs positives intactes. Cela permet d'augmenter l'efficacité de l'apprentissage, d'éviter la saturation des fonctions d'activation et de réduire le temps nécessaire pour entraîner le modèle (Zakaria, 2017).

En résumé, les couches RELU sont utilisées pour introduire de la non-linéarité dans les modèles CNN et augmenter l'efficacité de l'apprentissage (Belhadjer H, 2018).

La fonction ReLU (abréviation de Unités Rectifié linéaires) :  $F(x)=\max(0, x)$

### II.3.2.4 Couche entièrement connectée (FC) :

Les couches entièrement connectées, également appelées couches de sortie, sont situées à la fin de l'architecture d'un réseau de neurones convolutif (CNN). Contrairement aux couches de convolution et de pooling, ces couches sont entièrement connectées, c'est-à-dire que chaque neurone de la couche est connecté à tous les neurones de la couche précédente.

Ces couches sont utilisées pour la classification ou la régression et produisent des scores de classe ou des valeurs numériques. Dans le cas de la classification, chaque neurone de la couche de sortie représente une classe et fournit un score de probabilité pour cette classe. La classe avec le score le plus élevé est alors choisie comme prédiction finale (Belhadjer H, 2018).

### II.3.2.5 Couche de perte (LOSS) :

La couche de perte est une couche qui spécifie comment le réseau de neurones est entraîné en mesurant l'écart entre la sortie prédite par le modèle et la valeur de sortie réelle. Cela permet d'optimiser les poids du modèle afin qu'il puisse faire des prédictions plus précises. Elle est généralement placée à la fin du réseau.

Différentes fonctions de perte peuvent être utilisées dans la couche de perte en fonction de la tâche à accomplir.

La fonction « Softmax » est une fonction d'activation qui est souvent utilisée en conjonction avec la couche de perte pour calculer la distribution de probabilités sur les différentes classes de sortie. Elle transforme les scores de sortie du modèle en probabilités qui s'additionnent à 1. Cette fonction est souvent utilisée pour la classification multiclassées.

### II.3.3 Les modèles de réseaux de neurones

Effectivement, il existe différents modèles de réseaux de neurones qui peuvent être non pré-entraînés ou pré-entraînés :

- a. **Réseaux de neurones non pré-entraînés** : ce sont des modèles qui sont initialisés avec des poids aléatoires ou spécifiques définis par l'utilisateur. Ces modèles doivent être entraînés à partir de zéro sur des données spécifiques à la tâche. L'entraînement implique l'ajustement des poids du réseau à partir des exemples d'entraînement pour apprendre à effectuer la tâche spécifique pour laquelle le modèle est conçu. Les modèles non pré-entraînés nécessitent généralement plus de données et d'itérations d'entraînement pour obtenir des performances similaires aux modèles pré-entraînés.
  
- b. **Réseaux de neurones pré-entraînés** : ce sont des modèles qui ont été entraînés sur de grandes quantités de données dans le but d'apprendre des caractéristiques générales à partir de ces données. Ces modèles sont souvent entraînés sur des ensembles de données massifs tels que ImageNet, qui contient des millions d'images appartenant à différentes classes. En pré-entraînant un modèle sur de telles données, il apprend des représentations utiles des images, ce qui peut être bénéfique lorsqu'on l'applique à des tâches spécifiques (transfert d'apprentissage).

En résumé, les modèles de réseaux de neurones pré-entraînés sont des modèles déjà entraînés sur de grandes quantités de données, tandis que les modèles non pré-entraînés sont initialisés à partir de zéro et nécessitent un entraînement spécifique à la tâche. L'utilisation de modèles pré-entraînés peut être avantageuse, car ils capturent des connaissances générales à partir de données massives, ce qui permet une meilleure généralisation sur des tâches spécifiques et nécessite moins de données et d'itérations d'entraînement.

Voici quelques exemples de différents modèles de réseaux de neurones pré-entraînés :

- **VGG16 (Visual Geometry Group 16)** : Un réseau de neurones convolutionnels profond comprenant 16 couches, caractérisé par des convolutions de taille 3x3 et des opérations de max pooling de taille 2x2. Il a été pré-entraîné sur l'ensemble de données ImageNet et est souvent utilisé pour des tâches de classification d'images.
- **VGG19 (Visual Geometry Group 19)** : est une architecture de réseau de neurones convolutionnels (CNN) qui fait partie de la famille VGG. Elle est similaire à VGG16, mais se distingue par sa profondeur accrue avec un total de 19 couches, comprenant des couches de convolution et des couches entièrement connectées. VGG19 utilise des convolutions de taille 3x3 et des opérations de max pooling de taille 2x2. Il a été pré-entraîné sur le vaste ensemble de données ImageNet et est couramment utilisé dans des tâches de classification d'images où des caractéristiques complexes doivent être extraites et analysées.
- **ResNet-50 (Residual Network-50)** : Une architecture de réseau de neurones convolutifs profonds qui utilise des blocs résiduels pour faciliter l'apprentissage de réseaux plus profonds. Il comprend 50 couches et a été pré-entraîné sur ImageNet. Il

est couramment utilisé pour la classification d'images, la détection d'objets et d'autres tâches de vision par ordinateur.

## II.4 Application aux images médicales : Etat de l'art

Les CNNs sont devenus presque incontournables pour l'aide au diagnostic comme le prouvent plusieurs études récentes :

- Dans l'étude intitulée "**End-To-End Discriminative Deep Network For Liver Lesion Classification**" (Perdigón Romero, 2019), une approche d'apprentissage en profondeur de bout en bout utilisant **un réseau neuronal convolutif (CNN)** a été utilisée pour la discrimination entre les métastases hépatiques du cancer colorectal et les kystes bénins dans les images CT abdominales du foie. L'objectif était de faciliter cette discrimination, qui est critique pour le diagnostic et la stratégie thérapeutique, et qui nécessite une expertise, une expérience et un temps considérables.

L'approche proposée utilisait l'extraction efficace des caractéristiques à l'aide d'InceptionV3 combinée avec des connexions résiduelles et des poids pré-entraînés provenant d'ImageNet. L'architecture comprenait également des couches de classification entièrement connectées pour générer une sortie probabiliste du type de lésion. L'étude a utilisé une banque clinique interne comprenant 230 lésions hépatiques provenant de 63 patients.

Avec une précision de **0,96** et un score F1 de 0,92, les résultats obtenus avec cette approche dépassent les méthodes de pointe. Ce travail fournit une base pour l'intégration d'outils **d'apprentissage automatique**, tels que les réseaux neuronaux convolutifs, dans les logiciels de radiologie spécialisés afin d'aider les médecins dans la détection précoce et le traitement des lésions hépatiques.

- Dans une autre étude intitulée "**Deep Learning for Brain Tumor Classification from MRI Images**" (Arora, 2022), l'objectif était de détecter avec précision les tumeurs cérébrales à partir d'images IRM et de les classer en utilisant différentes techniques de transfert d'apprentissage basées sur **des réseaux neuronaux convolutifs (CNN)**.

Pour atteindre cet objectif, plusieurs méthodes de classification ont été utilisées. Tout d'abord, différents modèles de transfert d'apprentissage basés sur des **CNN**, tels que **VGG16**, ResNet50, InceptionV3 et DenseNet, ont été employés. Ces modèles ont été pré-entraînés sur de grandes bases de données d'images et ont montré une excellente capacité à extraire des caractéristiques significatives.

Les images IRM ont été prétraitées et les caractéristiques pertinentes ont été extraites en utilisant les couches pré-entraînées des modèles **CNN**. Ensuite, des couches de classification entièrement connectées ont été ajoutées pour générer une sortie probabiliste indiquant la présence ou l'absence de tumeur cérébrale.

Les résultats ont été évalués en mesurant la précision d'entraînement et la précision de test des modèles. Une comparaison des performances des différents modèles **CNN** a été réalisée pour déterminer lequel est le plus efficace pour la classification des tumeurs cérébrales dans les images IRM.

Les résultats obtenus ont démontré une précision de classification élevée pour la détection des tumeurs cérébrales à partir des images IRM. Les performances des modèles CNNs ont surpassé celles des méthodes existantes, mettant en évidence **l'efficacité des techniques d'apprentissage en profondeur** pour la classification des tumeurs cérébrales.

Cette étude fournit ainsi des informations précieuses pour le développement de systèmes d'aide au diagnostic de tumeurs cérébrales basées sur l'apprentissage automatique, ouvrant la voie à des outils plus précis et efficaces pour la détection précoce et le traitement des tumeurs cérébrales.

### ➤ **En Endoscopie**

**La détection de fumée dans les vidéos de chirurgie laparoscopique** est une tâche importante pour améliorer la sécurité et la précision des interventions chirurgicales.

- Dans l'article intitulé "**Improving endoscopic smoke detection with semi-supervised noisy student models**" (Reiter, 2020) , les auteurs se sont penchés sur l'amélioration de la détection de fumée endoscopique en utilisant des modèles semi-supervisés d'étudiants bruyants.

Les méthodes utilisées dans l'étude comprenaient la comparaison de deux techniques performantes : l'analyse des pics de saturation (SPA) et la classification **avec des réseaux de neurones convolutifs (CNN)**. Ces deux approches sont considérées comme des techniques de pointe, mais utilisent des modèles d'apprentissage flous différents.

Les résultats de l'étude ont montré que des améliorations supplémentaires pouvaient être obtenues en utilisant des quantités croissantes de données faussement étiquetées. Bien que la prédiction des faux positifs ait démontré la robustesse du classifieur, il subsistait encore un nombre significatif de faux négatifs dans les prédictions. Cela suggère qu'il est nécessaire d'inclure, d'avantage de données d'entraînement avec des cas extrêmes où la fumée apparaît en petites quantités ou lors de son évacuation pour améliorer les performances dans de telles situations difficiles.

Cette recherche met en évidence l'importance de l'utilisation de modèles semi-supervisés et de l'exploration de nouvelles approches pour améliorer la détection de fumée dans les vidéos laparoscopiques. Les résultats obtenus soulignent également les opportunités d'amélioration continue dans ce domaine et la nécessité de développer des méthodes plus robustes pour une détection plus précise de la fumée dans les procédures chirurgicales laparoscopiques.

- Dans l'étude intitulée "**Image-Based Smoke Detection in Laparoscopic Videos**" (Andreas Leibetseder, 2017), les auteurs se sont penchés sur la détection de la fumée dans les vidéos laparoscopiques en utilisant une approche basée sur l'image.

Les auteurs ont entraîné des modèles de **réseaux de neurones convolutionnels (CNN)** en utilisant un ensemble de données personnalisé comprenant plus de 30 000 images

laparoscopiques de fumée et de non-fumée. Ils ont également développé un classificateur non-CNN basé sur l'observation du canal de saturation d'un échantillon d'image dans l'espace colorimétrique HSV pour pouvoir comparer les performances.

Les résultats obtenus ont montré que les approches d'apprentissage en profondeur, en particulier l'utilisation de CNN, ont donné d'excellents résultats avec des courbes de caractéristique de fonctionnement du récepteur (ROC) présentant des zones de **plus de 0,98**. Cela démontre l'efficacité de l'approche CNN dans la détection de la fumée dans les vidéos laparoscopiques.

Ces résultats prometteurs soulignent **le potentiel** des techniques d'apprentissage en profondeur, en particulier des CNN, pour améliorer la détection de la fumée dans les vidéos laparoscopiques, ce qui pourrait contribuer à renforcer **la sécurité et l'efficacité** des interventions chirurgicales laparoscopiques.

- Dans l'étude intitulée "**Classifying smoke in laparoscopic videos using SVM**" (Nour Aldeen Jalal , 2017), les auteurs se sont intéressés à la détection de fumée dans les vidéos laparoscopiques, qui apparaît généralement lors de l'utilisation de l'électrocautérisation lors de la découpe ou de la coagulation des tissus. La méthode proposée repose sur l'extraction de diverses caractéristiques visuelles à partir des images laparoscopiques, puis les fournir à un classifieur de **machine à vecteurs de support (SVM)**. Les caractéristiques sont basées sur le mouvement, la couleur et les motifs de texture de la fumée. L'algorithme a été validé en utilisant une évaluation expérimentale sur quatre vidéos de cholécystectomie laparoscopique. Ces vidéos ont été annotées manuellement en définissant chaque image comme étant une image de fumée ou une image sans fumée. Les résultats expérimentaux montrent que la combinaison de toutes les caractéristiques proposées donne la meilleure performance de classification. La précision globale est d'environ **84%**, avec une sensibilité de 89% et une spécificité de 80%.

## **II.5 Conclusion**

L'apprentissage profond a révolutionné la classification en offrant une extraction automatique des caractéristiques, une amélioration des performances de classification et l'émergence de nouveaux modèles de réseaux de neurones. Ces avancées ont ouvert de nouvelles perspectives dans de nombreux domaines et continuent de jouer un rôle crucial dans les développements futurs de la classification et de l'analyse des données. À la lumière de ces recherches, nous allons appliquer les réseaux de neurones convolutionnels pour la détection et la classification de fumée dans les vidéos laparoscopiques.

---

# Chapitre III

---

**Méthodes et résultats**

### III.1 Introduction

Dans ce chapitre, nous allons présenter l'environnement de travail (logiciel et matériel) et la méthodologie adoptée pour résoudre le problème de détection de fumées dans les vidéos laparoscopiques. Nous exposerons à cet effet, trois modèles de classification de type CNN que nous avons choisi, paramétré et testé sous python. Les résultats obtenus seront discutés en fin de chapitre ce qui ouvrira des perspectives de recherche.

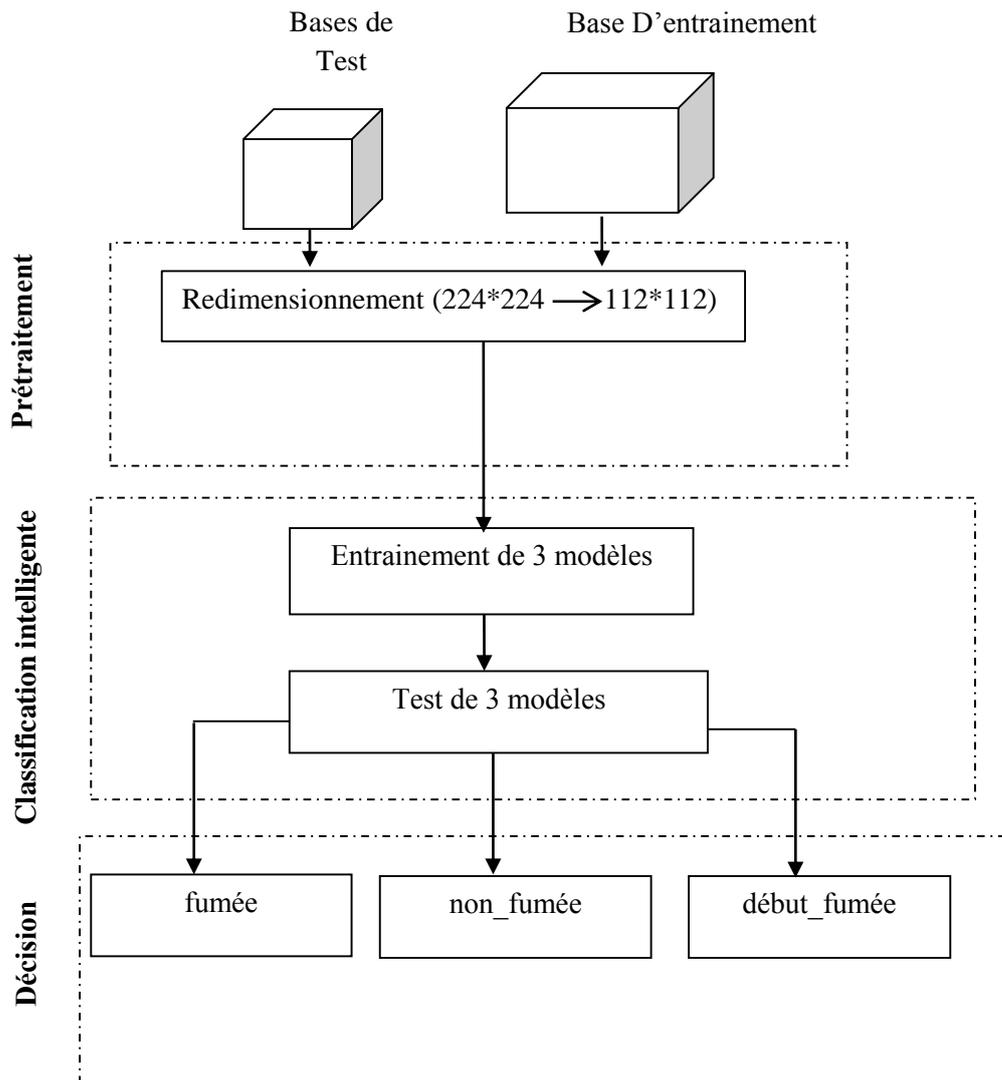


Figure III.1 Schéma de notre étude

## **III.2 Configuration Matérielle**

Notre environnement matériel est ordinateur personnel avec les spécifications suivantes :

- Processeur i5 CPU 3.9 GHz.
- Carte graphique Nvidia GeForce MX130.
- RAM de taille 4 GO.
- Disque dur de taille 1000 GB HDD.
- Système d'exploitation Windows 10 version 21H2.

## **III.3 Configuration Logicielle**

### **III.3.1 Environnement de programmation**

La classification basée sur l'approche du deep learning est principalement résolue par les communautés scientifiques en utilisant le langage de programmation Python. Ces communautés ont considérablement enrichi les bibliothèques Python, ce qui permet de traiter efficacement les divers problèmes rencontrés dans ce domaine. Ainsi, dans notre projet, nous avons également choisi Python comme langage de programmation pour bénéficier de sa flexibilité et de ses bibliothèques puissantes, en particulier les fonctionnalités avancées de deep learning. Ceci nous permettra de résoudre notre problème de classification avec précision et efficacité. Nous allons exposer quelques informations sur Python et certaines de ses bibliothèques.

#### **III.3.1.1 Python**

Python est un langage de programmation de haut niveau open source, interprété et polyvalent. Il a été créé dans les années 1990 par Guido van Rossum et se distingue par sa simplicité et sa lisibilité, ce qui en fait un choix populaire parmi les programmeurs débutants et expérimentés.

Python offre une vaste bibliothèque standard qui couvre de nombreux domaines d'application, ce qui facilite le développement de divers types de projets. De plus, Python est également réputé pour sa communauté active et engagée, qui développe et maintient de nombreuses bibliothèques tierces spécialisées, telles que TensorFlow, Keras, NumPy, Pandas, etc. Ces bibliothèques étendent les fonctionnalités de base de Python et permettent de travailler sur des domaines spécifiques tels que le machine learning, le traitement d'images, la science des données, etc.

#### **III.3.1.2 Keras**

La bibliothèque Keras est une bibliothèque open source pour le développement et l'entraînement de modèles de réseaux neuronaux. Elle a été initialement développée par François Chollet et est désormais intégrée dans le framework TensorFlow.

Keras offre une interface conviviale et simple à utiliser pour la création de modèles de réseaux neuronaux. Elle permet aux développeurs de définir, de configurer et de former des réseaux neuronaux avec un minimum de code. Keras prend en charge la création de modèles séquentiels, où les couches sont empilées linéairement les unes sur les autres, ainsi que la

création de modèles fonctionnels plus complexes, où les couches peuvent avoir des connexions multiples.

La bibliothèque Keras fournit également un large éventail de couches prédéfinies, telles que les couches denses (fully connected), les couches de convolution, les couches de pooling, les couches récurrentes, etc. Ces couches peuvent être facilement ajoutées à un modèle à l'aide de simples appels de fonction.

Keras permet également de configurer et d'optimiser les paramètres d'entraînement du modèle, tels que la fonction de perte (loss function), l'optimiseur et les métriques d'évaluation. Il est également possible de régulariser le modèle avec des techniques telles que la régularisation L1/L2 et le dropout.

L'une des principales caractéristiques de Keras est sa portabilité, car il peut être exécuté sur plusieurs backends, y compris TensorFlow, Theano et CNTK. Cela offre aux développeurs la flexibilité de choisir le backend qui convient le mieux à leurs besoins.

### **III.3.1.3 TensorFlow**

TensorFlow est un framework de programmation open source pour le calcul numérique, développé par Google et publié en novembre 2015. Depuis sa sortie, TensorFlow est devenu extrêmement populaire dans le domaine du deep learning et des réseaux neuronaux. Son succès est dû à sa flexibilité, sa performance et sa large communauté de développeurs.

TensorFlow est utilisé par de nombreuses entreprises et par beaucoup de chercheurs du monde entier pour des tâches de deep learning, telles que la vision par ordinateur, la traduction automatique, la génération de texte, la reconnaissance d'images, etc. Il offre une grande flexibilité pour la conception et l'entraînement de différents types de réseaux neuronaux, en fournissant des outils et des bibliothèques puissants pour les développeurs.

### **III.3.1.4 Open Cv**

La bibliothèque cv2, également connue sous le nom d'OpenCV (Open Source Computer Vision Library), est une bibliothèque open source largement utilisée pour le traitement d'images et la vision par ordinateur. Elle fournit des outils et des algorithmes avancés pour effectuer des opérations sur des images et des vidéos.

Voici quelques fonctionnalités de la bibliothèque cv2 :

1. Lecture et écriture d'images et de vidéos : cv2 permet de lire des images et des vidéos à partir de fichiers ou de flux en direct, et de les enregistrer dans différents formats.
2. Traitement d'images : cv2 offre un large éventail de fonctions pour effectuer des opérations de base sur les images, telles que le redimensionnement, la conversion de couleurs, le rognage, la mise à l'échelle, la rotation, le flou, l'amélioration des contrastes, etc.

3. Détection d'objets et de contours : cv2 fournit des algorithmes de détection d'objets, tels que la détection de visages, de contours, de lignes, de coins, etc. Ces fonctionnalités sont souvent utilisées pour l'analyse d'images et la reconnaissance d'objets.

4. Transformation d'images : cv2 permet d'effectuer diverses transformations sur les images, notamment la transformation de perspective, la transformation affine, la dilatation, l'érosion, la mise en correspondance des formes, etc.

5. Traitement vidéo : cv2 offre des fonctionnalités avancées pour le traitement des vidéos, telles que l'extraction de cadres, la capture d'écran, la segmentation d'objets, le suivi d'objets, etc.

6. Vision stéréo : cv2 propose des fonctionnalités de vision stéréo permettant de reconstruire la profondeur à partir de paires d'images stéréo.

Pour cela, OpenCV est largement utilisé dans divers domaines tels que la reconnaissance faciale, la détection d'objets, la vision industrielle, la réalité augmentée, la surveillance vidéo, la robotique, etc. Sa grande communauté de développeurs contribue constamment à son amélioration et à l'ajout de nouvelles fonctionnalités.

### **III.3.1.5 Scikit-learn**

Scikit-learn est une bibliothèque en Python dédiée à l'apprentissage automatique (machine learning). Elle fournit des outils et des fonctionnalités pour effectuer diverses tâches liées à l'apprentissage automatique, telles que la classification, la régression, le regroupement, la sélection de modèles, la préparation des données et l'évaluation des performances.

Scikit-learn propose une grande variété d'algorithmes d'apprentissage automatique, tels que les machines à vecteurs de support (SVM), les forêts aléatoires, les réseaux de neurones, les méthodes ensemblistes, la régression linéaire, les k-plus proches voisins (KNN), etc. Il offre également des outils pour la préparation des données, tels que le prétraitement, la normalisation et la sélection de fonctionnalités. Elle est conçue pour être compatible et pour s'harmoniser avec d'autres bibliothèques Python populaires, notamment NumPy et SciPy.

Scikit-learn est apprécié pour sa simplicité d'utilisation, sa documentation complète et ses bonnes pratiques de mise en œuvre des modèles d'apprentissage automatique. Il est largement utilisé dans la communauté de l'apprentissage automatique et est considéré comme l'un des principaux outils pour développer et déployer des modèles d'apprentissage automatique en Python.

### **III.3.1.6 Numpy**

La bibliothèque NumPy (Numerical Python) est une bibliothèque logicielle puissante et open source utilisée pour le calcul numérique en Python. Elle fournit des structures de données de tableau multidimensionnel (ndarray) efficaces et des fonctions mathématiques avancées pour effectuer des opérations rapides et vectorisées sur ces tableaux.

### III.3.1.7 Matplotlib

Matplotlib est une bibliothèque de visualisation de données en Python. Elle permet de créer une grande variété de graphiques avec une grande flexibilité pour leur personnalisation. Elle est largement utilisée dans le domaine de la science des données, de la visualisation de données, de l'analyse exploratoire et de la présentation de résultats.

D'autres bibliothèques utilisées :

- **imghdr** est une bibliothèque intégrée à Python qui permet de déterminer le type de fichier d'une image en analysant son en-tête.

- **pickle** est une bibliothèque intégrée à Python qui permet de sérialiser et désérialiser des objets Python, ce qui permet de les enregistrer dans des fichiers ou de les transférer entre des programmes.

- **datetime** est une bibliothèque intégrée à Python qui permet de manipuler et formater des dates et des heures.

- **PIL** (Python Imaging Library) est une bibliothèque externe populaire pour le traitement d'images en Python. Elle offre des fonctionnalités avancées pour ouvrir, manipuler et enregistrer des images dans différents formats.

### III.3.2 Programmation en ligne

Dans le cadre de notre projet, nous avons initialement utilisé Python sous l'IDE PyCharm. Cependant, nous avons rencontré des problèmes lors de l'utilisation et de l'installation de certaines bibliothèques essentielles telles que Keras et CV2. Ces difficultés ont entravé notre progression et nous ont amenés à chercher une solution alternative.

Après avoir exploré différentes options, nous nous sommes tournés vers la programmation en ligne avec Google Colab. Cette plateforme offre un environnement de développement Python basé sur le cloud, qui présente plusieurs avantages significatifs par rapport aux **IDE** locaux.

#### Google Colab

Google Colab (ou Colaboratory) est un service gratuit de Google qui permet d'exécuter du code Python dans un environnement Jupyter Notebook en ligne, sans avoir besoin de configurer un environnement Python localement sur sa machine. Google Colab permet également d'utiliser des ressources puissantes de calcul, notamment **des processeurs graphiques (GPU) et des unités de traitement tensoriel (TPU)**, pour accélérer l'exécution de certaines tâches de calcul intensif telles que l'entraînement de modèles d'apprentissage en profondeur. Il permet également de partager facilement des notebooks avec d'autres personnes, de travailler en équipe sur des projets de manière collaborative et de stocker les notebooks sur Google Drive.

## Le GPU

**GPU (Graphics Processing Unit)** est un type de processeur spécialisé conçu pour effectuer des calculs intensifs liés aux graphiques, tels que le rendu d'images et la manipulation de modèles 3D. Cependant, en raison de leur architecture parallèle et de leur capacité à effectuer de nombreuses opérations simultanément, les GPU sont également devenus extrêmement utiles pour accélérer les calculs dans d'autres domaines, notamment l'apprentissage automatique et le calcul scientifique. Les GPU sont capables de traiter de grandes quantités de données et d'effectuer des opérations vectorielles et matricielles de manière rapide et efficace.

Dans Google Colab, le GPU fourni est généralement un NVIDIA Tesla K80, un NVIDIA Tesla T4, un NVIDIA Tesla P4 ou un NVIDIA Tesla P100, selon la disponibilité. Les utilisateurs peuvent accéder gratuitement à ces ressources, mais avec des limites d'utilisation par session.

## Le TPU

**TPU (Tensor Processing Unit)** est un circuit intégré conçu spécialement pour accélérer les calculs d'apprentissage automatique en parallèle. Il est principalement utilisé pour améliorer la vitesse et l'efficacité de l'entraînement de réseaux de neurones profonds. Les TPU sont développés et utilisés par Google pour leur infrastructure de *cloud computing*.

En utilisant Google Colab, nous avons pu surmonter les problèmes d'installation des bibliothèques et bénéficier des avantages suivants :

- **Facilité d'utilisation et d'accès** : Google Colab offre un environnement prêt à l'emploi, sans nécessité d'installation ni de configuration supplémentaire. Nous avons pu commencer à coder rapidement et facilement, en utilisant simplement un navigateur web.
- **Gestion des bibliothèques simplifiée** : Google Colab prend en charge l'installation automatique des bibliothèques et des dépendances requises. Cette fonctionnalité a grandement simplifié le processus de configuration de notre environnement de développement, nous permettant de nous concentrer sur la mise en œuvre de nos modèles plutôt que sur les problèmes d'installation.
- **Accès aux ressources de calcul** : Google Colab offre des ressources de calcul puissantes, y compris des processeurs et des GPU. Cela nous a permis de bénéficier de la puissance de calcul distribuée, en particulier pour l'entraînement de modèles de réseaux de neurones profonds, sans nécessiter un matériel informatique coûteux.
- **Partage et collaboration** : Google Colab facilite le partage de notebooks avec d'autres utilisateurs. Nous avons pu partager notre code et nos résultats avec nos collègues et nos superviseurs, ce qui a favorisé la collaboration et les échanges d'idées.

## Importance de l'activation du GPU

Il faut savoir que le GPU n'est pas activé automatiquement. Ne connaissant pas cette information, nous l'avons utilisé Google sans activer l'unité de traitement graphique (GPU). Nous avons constaté une lenteur d'exécution inattendue. En effet, nous avons travaillé sur des tâches d'entraînement de modèles d'apprentissage automatique complexes et sur le traitement

d'images volumineuses. Ces opérations, qui impliquaient des boucles itératives et des calculs matriciels, prenaient normalement **de 4 à 7 heures** pour se terminer.

Cependant, **après avoir activé le GPU dans Google Colab, nous avons remarqué une différence spectaculaire dans les performances.** Les mêmes opérations qui prenaient **des heures** ont été exécutées en seulement **quelques minutes**. Cette accélération significative est due à la nature parallèle du plusieurs GPU, qui peut effectuer calculs simultanément, optimisant ainsi le traitement des données.

En activant le GPU, nous avons pu profiter de sa puissance de calcul élevée pour accélérer les tâches intensives en calcul. Ce qui a considérablement réduit le temps nécessaire pour effectuer les tâches.

L'utilisation du GPU dans Google Colab nous a également permis de tirer parti des bibliothèques d'apprentissage automatique (TensorFlow), qui sont spécifiquement optimisées pour utiliser les ressources de calcul du GPU. Cela a accéléré le processus d'entraînement des modèles, réduisant le temps nécessaire pour ajuster les poids et les paramètres du modèle.

De plus, l'activation du GPU a également accéléré les tâches de traitement d'images ou de vidéos, telles que la détection d'objets, la segmentation d'images ou le suivi de mouvement. Ces tâches impliquent souvent des calculs complexes sur des données volumineuses, et le GPU a joué un rôle clé dans l'accélération de ces opérations.

### **III.4 La base de données des images laparoscopiques**

La base de données Laparoscopic Video Quality (LVQ) (**Z. A. Khan, 2020**) (**Beghdadi, 2020**) est une collection de vidéos laparoscopiques utilisées dans la recherche sur l'évaluation et l'amélioration de la qualité des vidéos laparoscopiques. La base de données contient 10 vidéos de référence d'une durée de 10 secondes chacune, qui sont déformées avec 5 types de distorsions différents à 4 niveaux différents. Les distorsions incluent certaines des distorsions couramment rencontrées lors de la chirurgie laparoscopique, notamment le flou de mise au point, le flou de mouvement, l'éclairage inégal, la fumée et le bruit.

Les scores subjectifs dans la base de données ont été obtenus à la fois auprès d'observateurs non médicaux (29 au total) et d'observateurs médicaux (9 au total). La base de données a été créée et supervisée par le chercheur principal Azeddine Beghdadi, avec la contribution d'autres chercheurs tels que Z.A. Khan, F.A. Cheikh, M. Kaaniche, E. Pelanis, R. Palomar, Å.A. Fretland, B. Edwin et O.J. Elle.

Cette base de données est destinée uniquement à un usage strictement orienté vers la recherche.

#### **Préparation des images pour l'apprentissage**

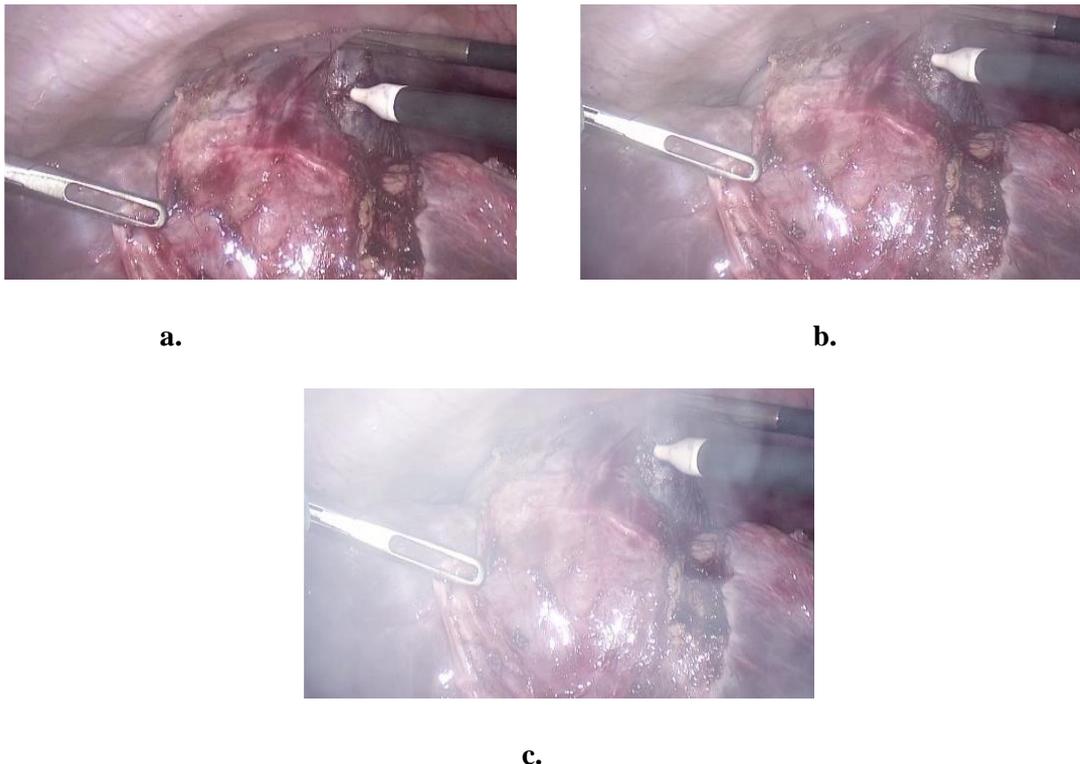
La base de données LVQ comprend des vidéos laparoscopiques avec différentes distorsions dont la fumée. Pour ce projet de recherche, nous avons choisi uniquement le type de distorsion "**Smoke**".

Comme nous venons de dire, LVQ contient des vidéos, donc nous avons entrepris de construire une base d'images à partir de ces vidéos (Figure III.2). Au total, nous avons travaillé sur 36 vidéos issues de la base de données.

Pour faciliter l'apprentissage, nous avons divisé ces vidéos en trois classes distinctes : "smoke" (fumée), "no\_smoke" (pas de fumée) et "start\_smoke" (début de fumée). Cela nous a permis d'organiser les données en fonction de la présence ou de l'absence de fumée dans les vidéos.

Au début de notre processus, la division des images a abouti à un total de 893 images. Cependant, nous avons apporté une modification en réduisant l'intervalle de temps entre les captures, nous avons augmenté la fréquence des captures, ce qui s'est traduit par une augmentation significative du nombre total d'images obtenues qui passe ainsi de **893** à **9238** images.

Plus précisément, nous avons obtenu 2307 images pour la classe "no\_smoke", 2307 images pour la classe "smoke", et 4614 images pour la classe "start\_smoke". Ces images extraites serviront de données d'entraînement pour le classifieur.



**Figure III.2** Images issues de la base LVQ (a. image « no\_smoke », b. image « start\_smoke », c. image « smoke »).

### **III.5 Architecture de notre réseau**

Au cours de nos expérimentations, nous avons utilisé trois modèles : VGG16, VGG19 et un troisième modèle que nous avons créé. L'objectif est d'évaluer l'influence des différentes architectures et paramètres sur la résolution de notre problème (détection et la classification de fumée), et d'obtenir des résultats dans le cadre de notre étude.

#### **Adaptation de l'architecture des modèles à nos images**

Pour adapter l'architecture des trois modèles à nos images de base de données de taille  $224 \times 224 \times 3$ , nous avons suivi plusieurs étapes. Tout d'abord, nous avons réduit la taille de nos images, passant de  $224 \times 224$  à  $112 \times 112$ , en utilisant une méthode de redimensionnement appropriée, à l'aide de bibliothèque PIL (Python Imaging Library) et OpenCV. Ensuite, nous avons modifié la taille d'entrée de la première couche Conv2D pour qu'elle corresponde à la nouvelle taille des images. De plus, nous avons adapté les dimensions des couches Flatten et des couches Fully Connected (FC) pour assurer la compatibilité dimensionnelle avec les nouvelles dimensions des images. Ces ajustements sont essentiels pour garantir que les modèles puissent traiter efficacement les images de taille  $112 \times 112 \times 3$ .

Dans ce qui suit on présente l'architecture des trois modèles :

#### **III.5.1 Architecture du premier modèle**

Le modèle présenté dans la **figure III.3** comprend cinq couches de convolution, deux couches de maxpooling et trois couches fully connected. L'image d'entrée a une taille de  $112 \times 112$ .

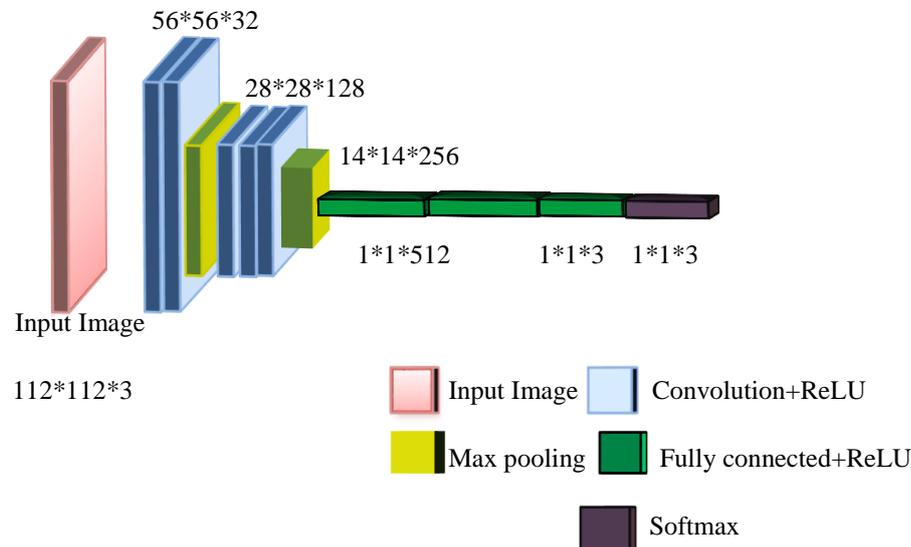
La première étape consiste à traiter l'image d'entrée de dimensions  $112 \times 112$  à travers la première couche de convolution. Cette couche est formée de 32 filtres de taille  $3 \times 3$ . Chaque couche de convolution est suivie d'une fonction d'activation appelée ReLU, qui contraint les neurones à produire des valeurs positives. Après cette convolution, 32 cartes de caractéristiques de dimensions  $112 \times 112$  sont générées.

Les 32 cartes de caractéristiques obtenues précédemment sont utilisées en entrée pour la deuxième couche de convolution, qui est également composée de 32 filtres. Une fonction d'activation ReLU est appliquée sur cette couche de convolution. Ensuite, nous utilisons la technique de Maxpooling pour réduire la taille de l'image ainsi que la quantité de paramètres et de calcul nécessaires. À la sortie de cette couche, nous obtenons 32 cartes de caractéristiques de dimensions  $56 \times 56$ .

Le processus est répété pour les couches de convolution trois, quatre et cinq. Chacune de ces couches est composée de 128 filtres, et la fonction d'activation ReLU est appliquée après chaque convolution. Après la cinquième couche de convolution, une couche de Maxpooling est appliquée. À la sortie de cette couche, nous obtenons 256 cartes de caractéristiques de dimensions  $14 \times 14$ . Le vecteur de caractéristiques résultant des convolutions a une dimension de 1024.

Une fois ces cinq couches de convolution terminées, nous utilisons un réseau de neurones composé de trois couches fully connected. Les deux premières couches comportent chacun 512

neurones, utilisant la fonction d'activation ReLU. La troisième couche est une couche softmax qui calcule la distribution de probabilité des 3 classes possibles.



**Figure III.3** Architecture du notre modèle

### III.5.2 Architecture du deuxième modèle

Le deuxième modèle « **VGG16** » que nous présentons dans la **figure III.4** comprend cinq blocs de convolution, cinq couches de maxpooling et trois couches fully connected.

L'image d'entrée a une taille de 112x112 pixels, l'image passe à travers chaque couche de convolution, suivi d'une fonction d'activation ReLU, et ensuite à travers chaque couche de maxpooling.

La première couche de convolution utilise 64 filtres de taille 3x3. Après la convolution, la taille de la carte de caractéristiques reste la même, c'est-à-dire 112x112.

Ensuite, la deuxième couche de convolution utilise également des filtres de taille 3x3. Après la convolution, la taille de la carte de caractéristiques reste à nouveau la même, soit 112x112.

Après chaque couche de convolution, une couche de maxpooling est appliquée. Le maxpooling réduit la taille de la carte de caractéristiques de moitié. Ainsi, après la première couche de maxpooling, la taille devient 56x56. Après la deuxième couche de maxpooling, la taille devient 28x28.

Le processus de convolution et de maxpooling se poursuit avec les couches suivantes. Les filtres de taille 3x3 sont appliqués à chaque couche de convolution, conservant la taille de la carte de caractéristiques.

Après le quatrième bloc de convolution, la taille de la carte de caractéristiques devient 7x7. Ensuite, le dernier bloc de convolution réduit la taille à 3x3.

Après les cinq blocs de convolution, la carte de caractéristiques est aplatie en un vecteur unidimensionnel à l'aide de la couche "flatten". Ce vecteur a une taille de 4608.

Ensuite, les trois couches fully connected (FC) sont utilisées pour la classification finale de l'image. Chaque couche fully connected contient 4096 neurones. La fonction d'activation utilisée dans les deux premières couches fully connected est ReLU. Enfin, la couche de sortie utilise la fonction softmax pour calculer la distribution de probabilité des 3 classes (nombre de classe dans notre base d'image LVQ).

En résumé, l'image d'entrée de taille 112x112 passe à travers les couches de convolution et de maxpooling, conservant la taille de la carte de caractéristiques à chaque étape, jusqu'à ce qu'elle atteigne la taille finale de 1x1 avant d'être aplatie en un vecteur de taille 4608. Ensuite, les couches fully connected sont utilisées pour la classification.

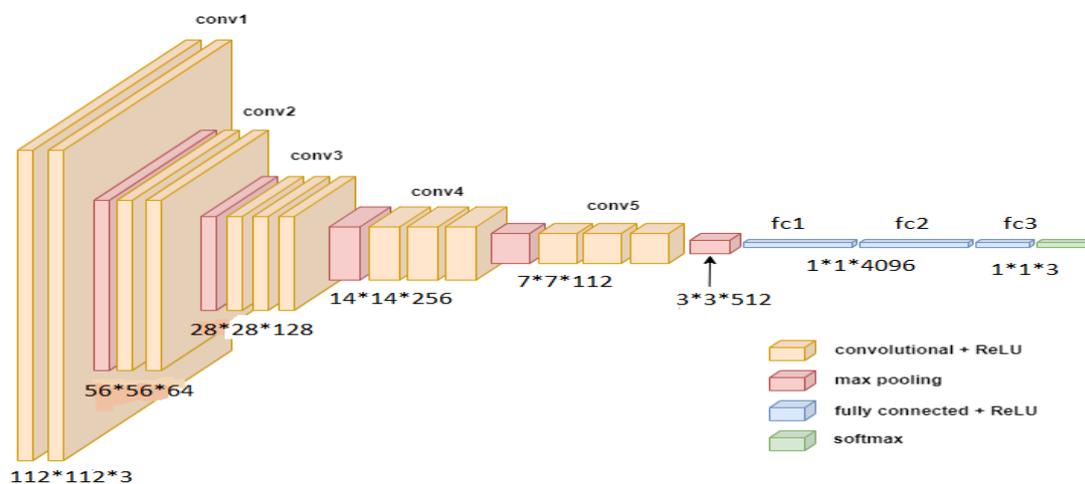


Figure III.4 Architecture du modèle 2 «VGG16 »

### III.5.3 Architecture du troisième modèle

Le troisième modèle « **VGG19** » que nous présentons dans la **figure III.5** se compose de 16 couches de convolution, 5 couches de maxpooling et suivies de 3 couches entièrement connectées. Chaque couche de convolution est généralement associée à une couche de ReLU (Rectified Linear Unit) qui applique une fonction d'activation non linéaire.

L'image d'entrée a une taille de 112x112 pixels, l'image passe à travers chaque couche de convolution, suivi d'une fonction d'activation ReLU, et ensuite à travers chaque couche de maxpooling.

La première paire de couches de convolution comporte 64 filtres, suivie d'une couche de max pooling avec une fenêtre de 2x2 et un pas de 2. La deuxième paire de couches de convolution suit le même schéma.

Ensuite, le processus de convolution et de max pooling est répété avec des filtres de taille 3x3. La troisième paire de couches de convolution comprend 128 filtres, suivie d'une couche de max pooling. La quatrième paire comprend 256 filtres, et la cinquième paire comprend 512 filtres. Après chaque paire de couches de convolution, une couche de max pooling est appliquée pour réduire la dimensionnalité.

Une fois que les caractéristiques ont été extraites par les couches de convolution et de pooling, l'image est aplatie pour être utilisée comme entrée pour les couches entièrement connectées. Le VGG19 comporte trois couches entièrement connectées : une couche avec neurones, suivie d'une autre couche avec 4096 neurones, et enfin une couche avec 3 neurones correspondant aux classes de sortie.

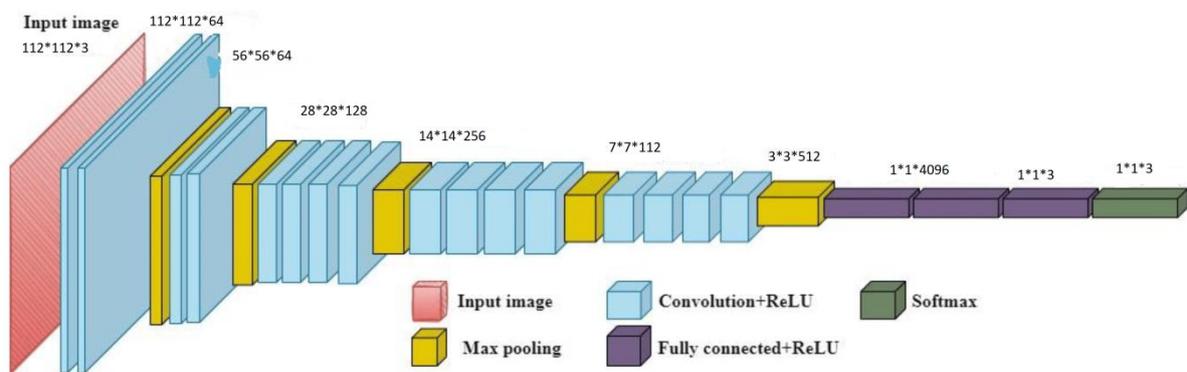


Figure III.5 Architecture du modèle 3 «VGG19 »

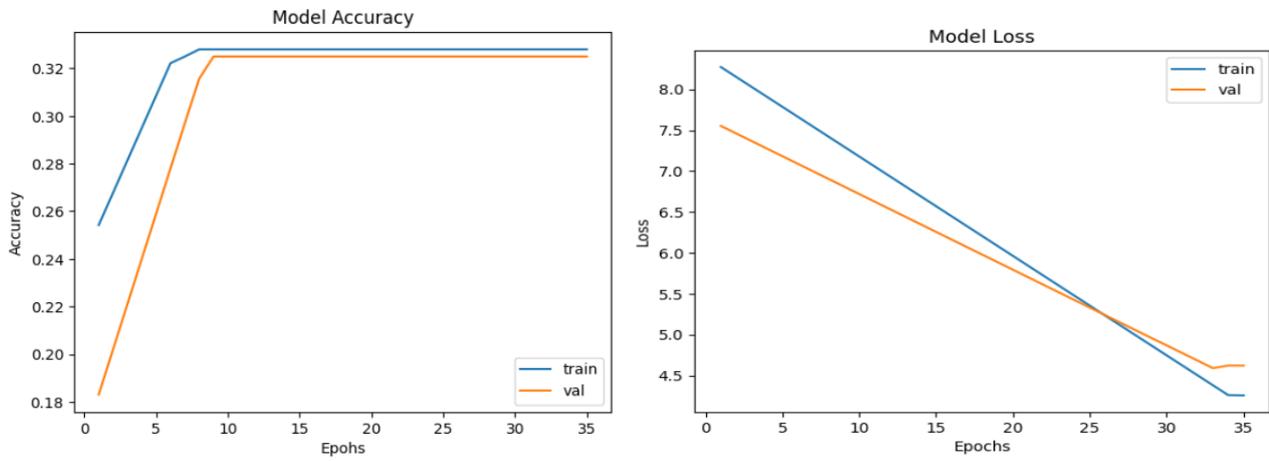
### III.6 Résultats obtenus et discussion

Dans cette section, nous examinerons les résultats des trois modèles évalués lorsque la taille de la base était de 893 images, ainsi que lorsqu'elle a été augmentée à 9238 images. Pour chaque cas, nous avons utilisé une répartition de **80%** des images pour l'apprentissage et **20%** des images pour le test.

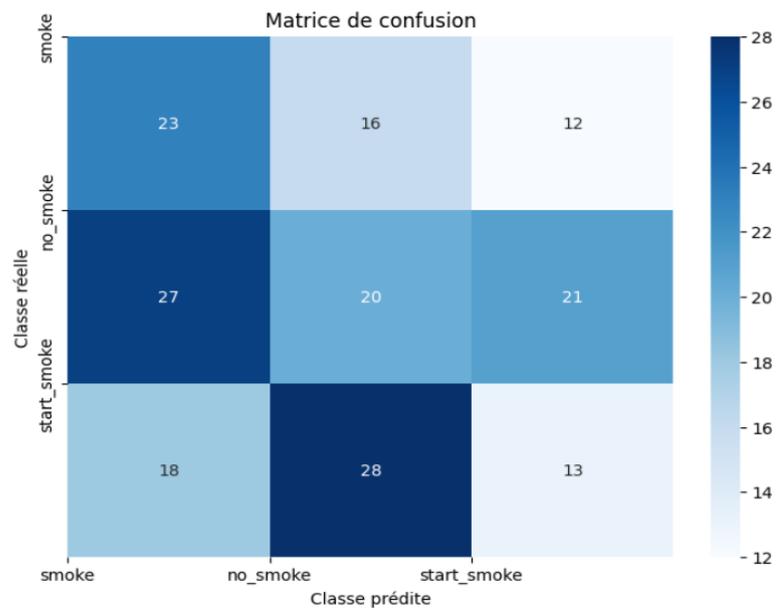
L'objectif est de comparer les performances des modèles dans les deux scénarios afin d'évaluer l'impact de l'augmentation de la taille de la base de données sur les résultats obtenus.

### III.6.1 Résultats obtenus par notre modèle:

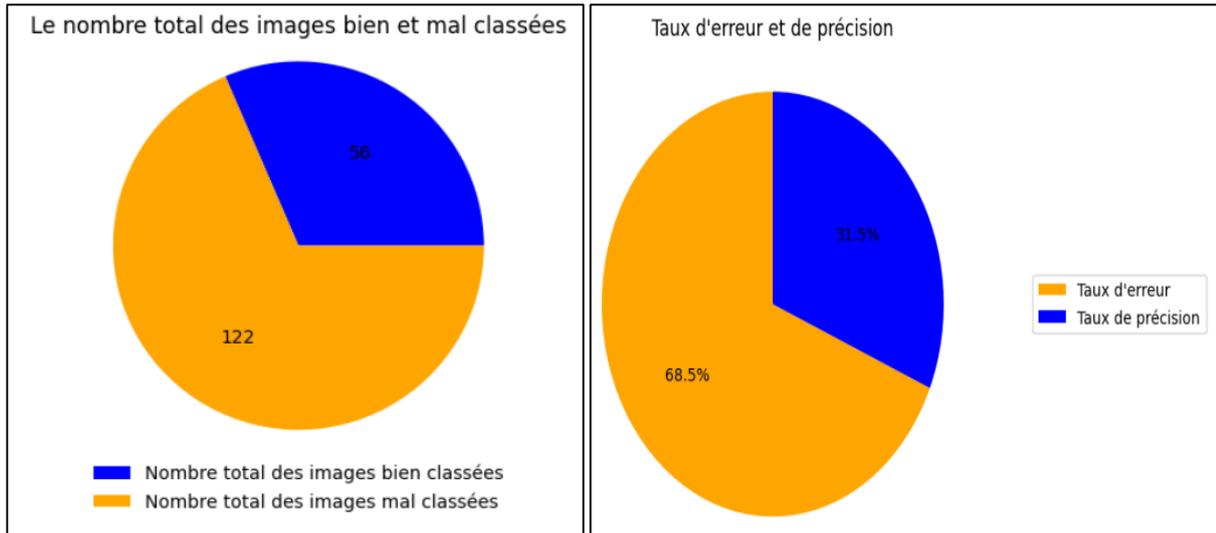
#### a. Résultats sur 893 images



**Figure III.6** Précision et Erreur pour notre modèle sur 893 images

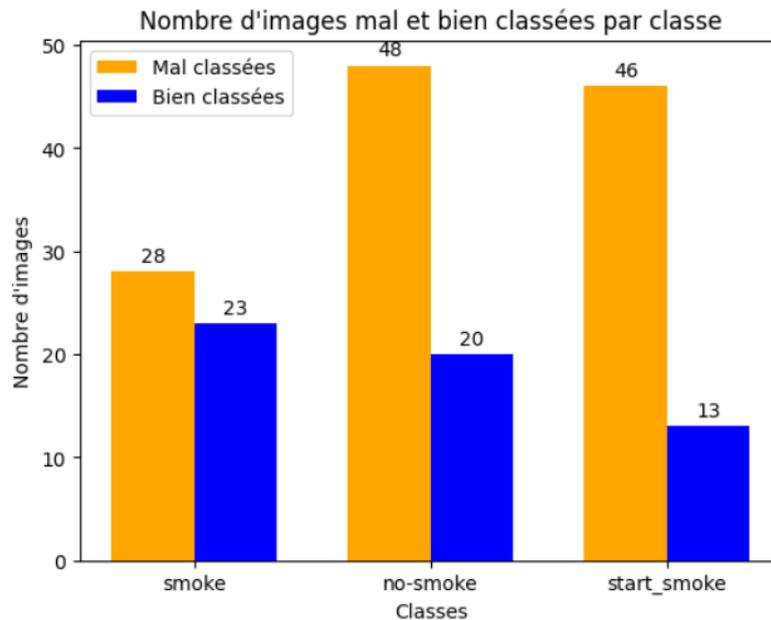


**Figure III.7** Matrice de confusion du notre modèle



**Figure III.8** Nombre total des images mal et bien classé

**Figure III.9** Taux d'erreur et de précision



**Figure III.10** Nombre des images mal et bien classé pour notre modèle

Après avoir examiné les résultats obtenus, nous pouvons noter les observations suivantes :

D'après la **Figure III.6** la précision de l'apprentissage et de la validation reste constante et ne montre pas d'amélioration significative au fur et à mesure que le nombre d'époques augmente. La précision d'apprentissage est relativement basse, atteignant environ 0.33.

Les courbes de l'erreur d'apprentissage et de validation démontrent une tendance à la baisse régulière au fur et à mesure que le nombre d'époques augmente. Cela suggère que le modèle parvient à améliorer ses prédictions et à réduire l'erreur au fil du temps. Toutefois, il est important de noter que malgré cette amélioration, les valeurs de l'erreur d'apprentissage et de validation demeurent relativement élevées.

Les résultats indiquent que le modèle ne parvient pas à obtenir de bonnes performances en termes de précision et d'erreur.

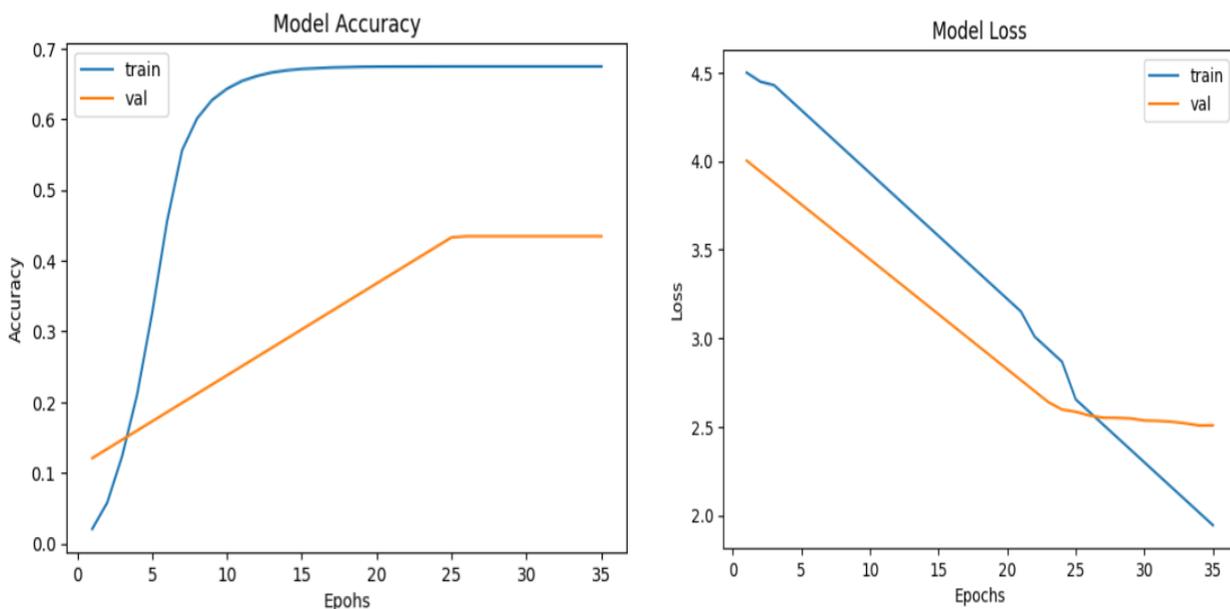
D'après la **Figure III.8** et **Figure III.9** nous constatons que le modèle présente une précision de **31,5%** et une erreur de **68,5%**. Sur un ensemble total de 178 images évaluées, seulement 56 ont été correctement classées, tandis que 122 ont été mal classées. Ces résultats révèlent une performance limitée du modèle en ce qui concerne la classification des images, avec un taux d'erreur élevé.

La matrice de confusion de la **Figure III.7** résume les prédictions du modèle sur les images de test pour les trois classes :

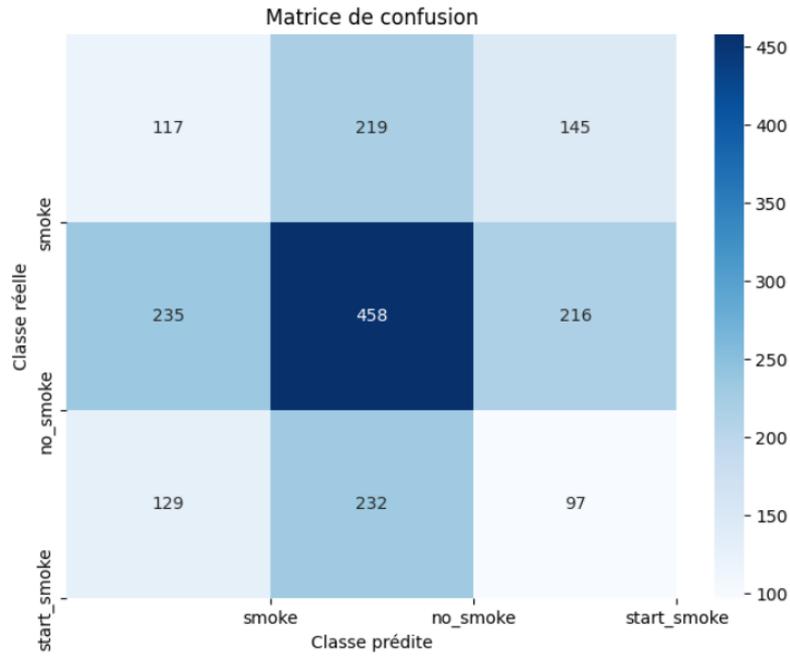
- Classe 1 « **smoke** »: le modèle a correctement prédit 23 échantillons, mais en a mal classé 16 comme appartenant à la classe 2 « no\_smoke » et 12 comme appartenant à la classe 3 « start\_smoke ».
- Classe 2 « **no\_smoke** »: le modèle a correctement prédit 20 échantillons, mais en a mal classé 27 comme appartenant à la classe 1 « smoke » et 21 comme appartenant à la classe 3 « start\_smoke ».
- Classe 3 « **start\_smoke** »: le modèle a correctement prédit 13 échantillons, mais en a mal classé 18 comme appartenant à la classe 1 « smoke » et 28 comme appartenant à la classe 2 « no\_smoke ».

D'après la **Figure III.10** nous remarquons que le modèle a eu des difficultés à distinguer correctement les échantillons des classes 1 et 2, ainsi que les échantillons des classes 2 et 3. Cependant, une observation importante est que la classe « no\_smoke » présente un nombre plus élevé d'images mal classées par rapport aux autres classes.

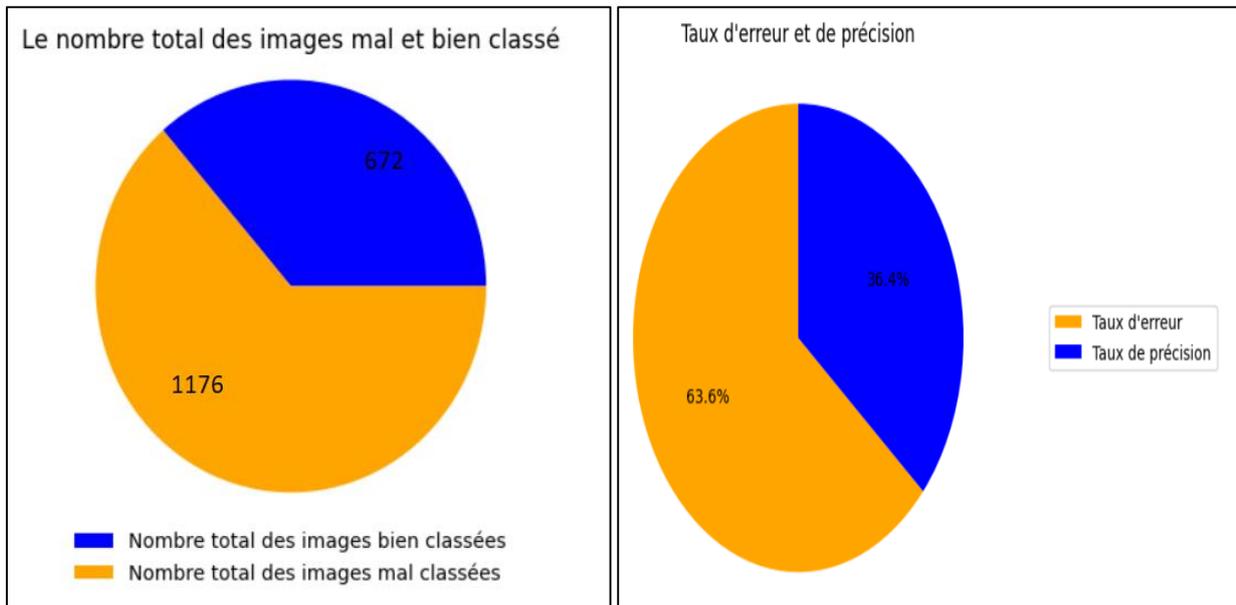
## **b. Résultats sur 9238 images**



**Figure III.11** Précision et Erreur pour notre modèle

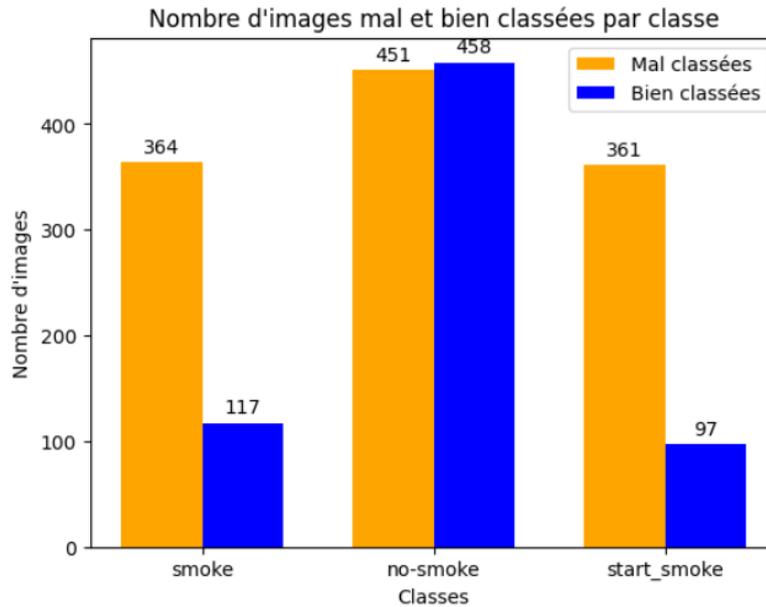


**Figure III.12** Matrice de confusion du notre modèle



**Figure III.13** Nombre total des images mal et bien classé

**Figure III.14** Taux d'erreur et de précision



**Figure III.15** Nombre des images mal et bien classé pour notre modèle

Après avoir examiné les résultats obtenus, nous pouvons noter les observations suivantes :

D'après la **Figure III.11** les courbes d'accuracy (précision) pour l'apprentissage et la validation montrent une tendance croissante avec l'augmentation du nombre d'époques. Cependant, la précision atteint un plateau assez tôt, autour de **0.67**, et ne semble pas s'améliorer davantage par la suite.

De même les courbes de loss (erreur) pour l'apprentissage et la validation montrent une tendance décroissante avec l'augmentation du nombre d'époques, aussi l'erreur diminue de manière significative au début de l'apprentissage, puis converge vers une valeur proche de **2.5**. La courbe de validation suit une tendance similaire, mais avec une légère augmentation après environ 25 époques.

D'après les résultats obtenus, le modèle a une performance limitée dans la classification des données et qu'il commet des erreurs dans une grande proportion des prédictions.

D'après la **Figure III.13** et **Figure III.14** nous observons que la précision de 36,4% et une erreur de 63,6% pour le modèle. Sur un total de 1848 images évaluées, 672 ont été bien classées, tandis que 1176 ont été mal classées. Ces résultats indiquent que le modèle présente une performance limitée dans la classification des images, avec un pourcentage élevé d'erreurs.

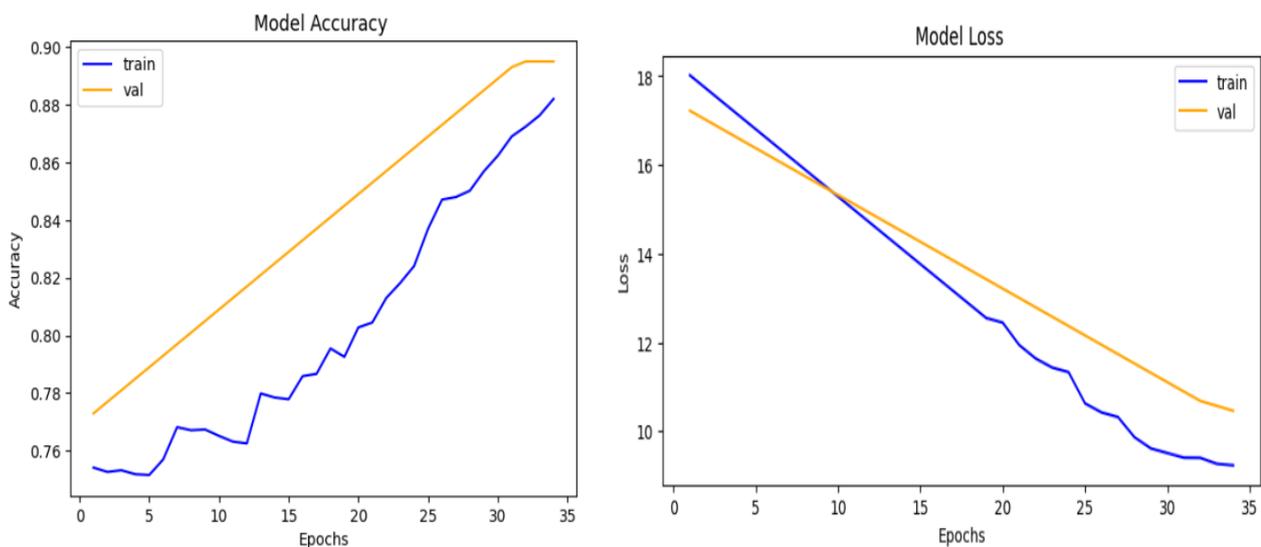
La matrice de confusion de la **Figure III.12** résume les prédictions du modèle sur les images de test pour les trois classes :

- Classe 1 « **smoke** »: le modèle a correctement prédit 117 échantillons, mais en a mal classé 219 comme appartenant à la classe 2 « no\_smoke » et 145 comme appartenant à la classe 3 « start\_smoke ».
- Classe 2 « **no\_smoke** »: le modèle a correctement prédit 458 échantillons, mais en a mal classé 235 comme appartenant à la classe 1 « smoke » et 216 comme appartenant à la classe 3 « start\_smoke ».
- Classe 3 « **start\_smoke** »: le modèle a correctement prédit 97 échantillons, mais en a mal classé 129 comme appartenant à la classe 1 « smoke » et 232 comme appartenant à la classe 2 « no\_smoke ».

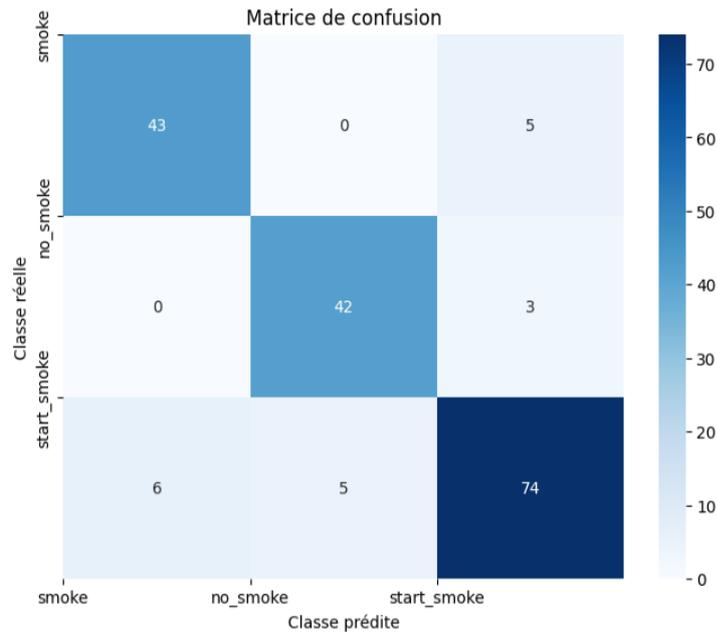
D'après la **Figure III.15** nous remarquons que le modèle a eu des difficultés à distinguer correctement les échantillons des classes 1 et 2, ainsi que les échantillons des classes 2 et 3. Cependant, il a obtenu une meilleure performance dans la classification des échantillons de la classe 2 « no\_smoke » par rapport aux autres classes.

### III.6.2 Résultats obtenus par le modèle 2 « VGG16 » :

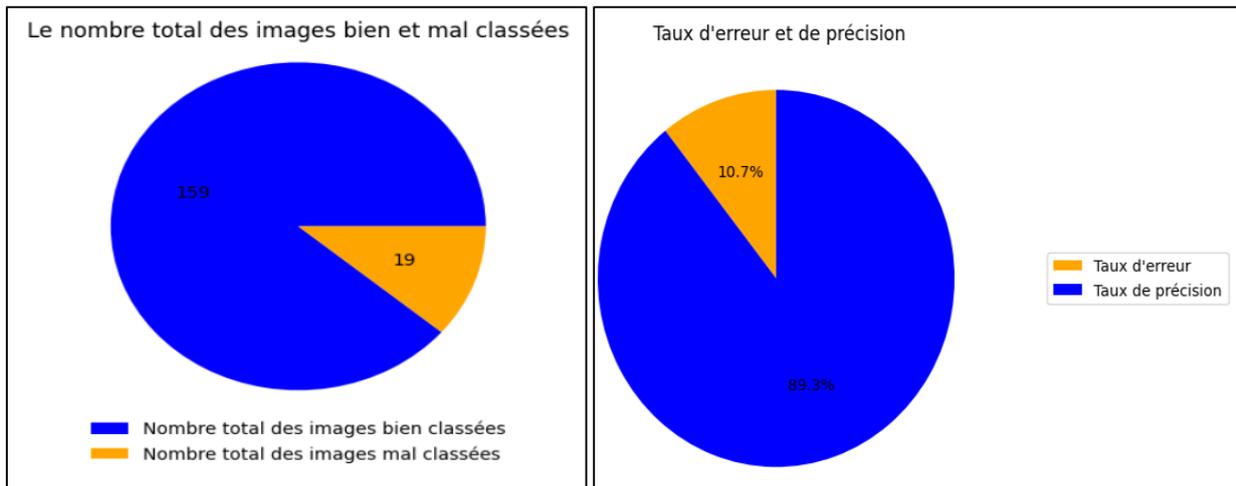
#### a. Résultats sur 893 images



**Figure III.16** Précision et Erreur du **VGG16** sur 893 images

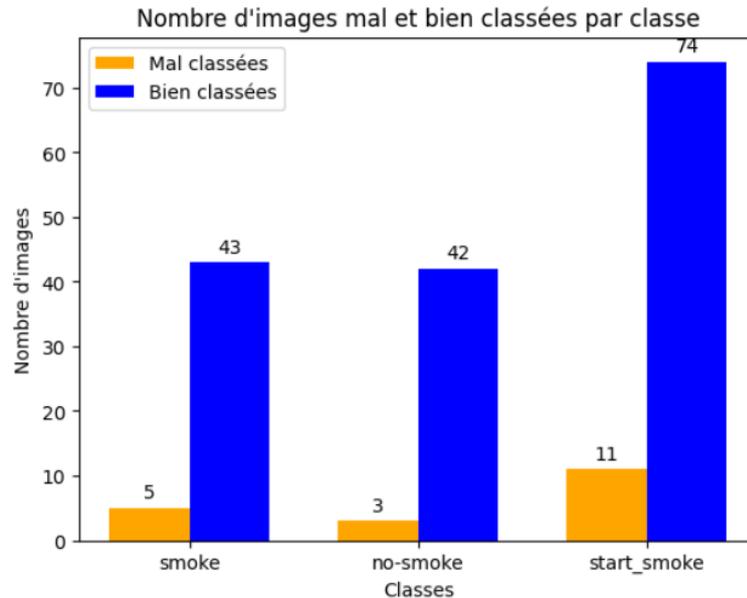


**Figure III.17** Matrice de confusion du **VGG16**



**Figure III.18** Nombre total des images mal et bien classé

**Figure III.19** Taux d'erreur et de précision



**Figure III.20** Nombre des images mal et bien classé par VGG16

Après avoir analysé les résultats obtenus, les observations suivantes ont été relevées :

D'après la **Figure III.16** la précision de l'apprentissage (train) augmente progressivement à mesure que le nombre d'époques augmente. Cela suggère que le modèle apprend et s'améliore au fil du temps, la précision de validation (val) suit une tendance similaire à celle de l'apprentissage jusqu'à atteindre une valeur de **0.896**.

De même, l'erreur d'apprentissage (train) diminue régulièrement à mesure que le nombre d'époques augmente, l'erreur de validation (val) suit une tendance similaire à celle de l'apprentissage, bien que légèrement plus élevée que la perte d'apprentissage jusqu'à atteindre une valeur d'environ **9.65**.

Les résultats indiquent que le modèle progresse avec l'augmentation du nombre d'époques, avec des courbes de précision et de perte qui convergent vers une performance optimale.

D'après la **Figure III.18** et **Figure III.19** nous observons que le nombre total d'images mal classées est 19, ce qui correspond à un taux d'erreur de **10.7%**. En revanche, toutes les autres images (soit un total de 159) ont été correctement classées, ce qui se traduit par un taux de précision de **89.3%**.

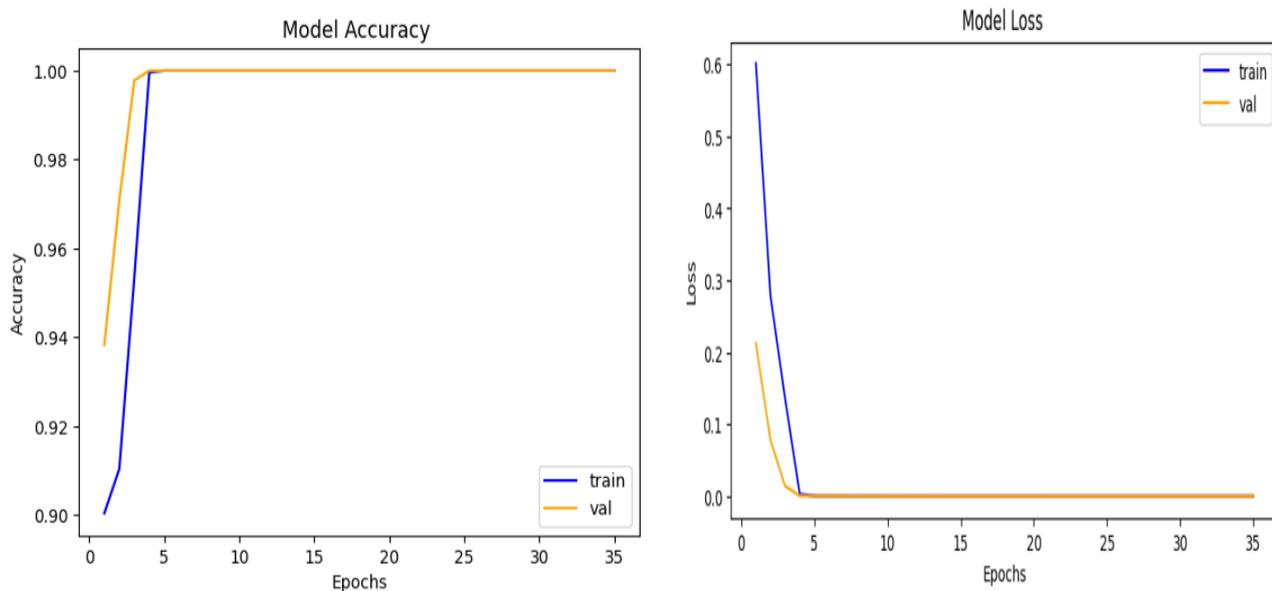
La matrice de confusion de la **Figure III.17** synthétise les résultats des prédictions effectuées par le modèle sur les images de test pour les trois classes:

- Classe 1 « **smoke** » : Il y a 43 échantillons qui appartiennent à cette classe et qui ont été correctement classés comme tels. Il y a 5 échantillons de cette classe qui ont été mal classés comme appartenant à la classe 3 « start\_smoke ».
- Classe 2 « **no\_smoke** » : Il y a 42 échantillons qui appartiennent à cette classe et qui ont été correctement classés comme tels. Il y a 3 échantillons de cette classe qui ont été mal classés comme appartenant à la classe 3 « start\_smoke ».

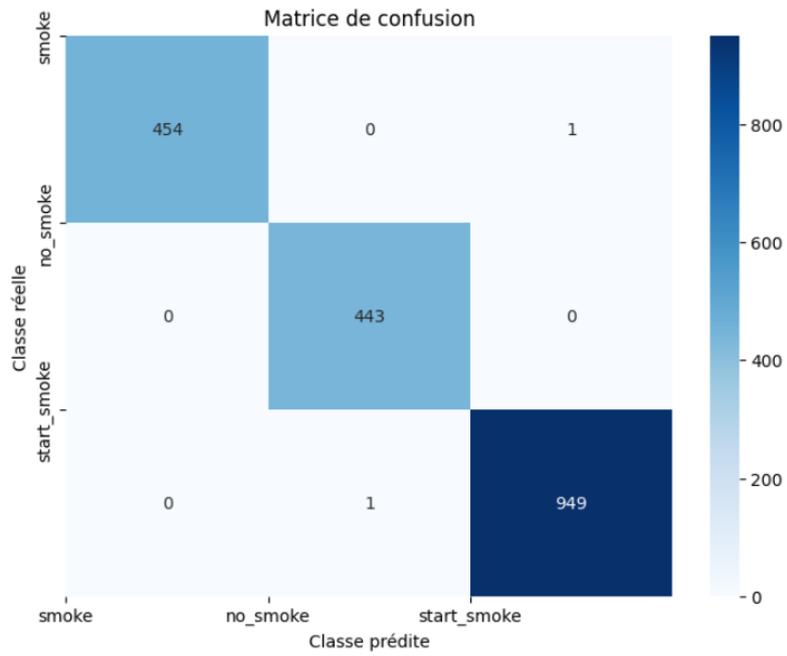
- Classe 3 « **start\_smoke** » : Il y a 74 échantillons qui appartiennent à cette classe et qui ont été correctement classés comme tels. Il y a 6 échantillons de cette classe qui ont été mal classés comme appartenant à la classe 1 « smoke » et 5 échantillons qui ont été mal classés comme appartenant à la classe 2 « no\_smoke ».

D'après la **Figure III.20** la plupart des échantillons ont été correctement classés dans leurs classes respectives, nous remarquons que la classe "start\_smoke" présente le plus grand nombre d'images mal classées, ce qui indique que le modèle a plus de difficultés à prédire correctement cette classe.

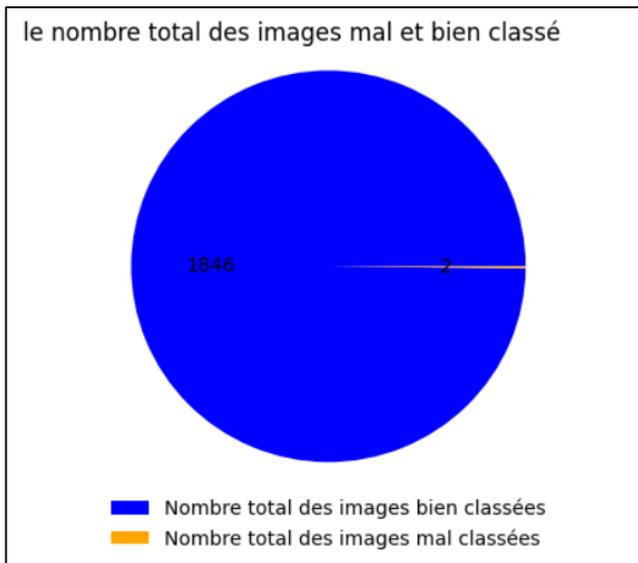
### **b. Résultats sur 9238 images**



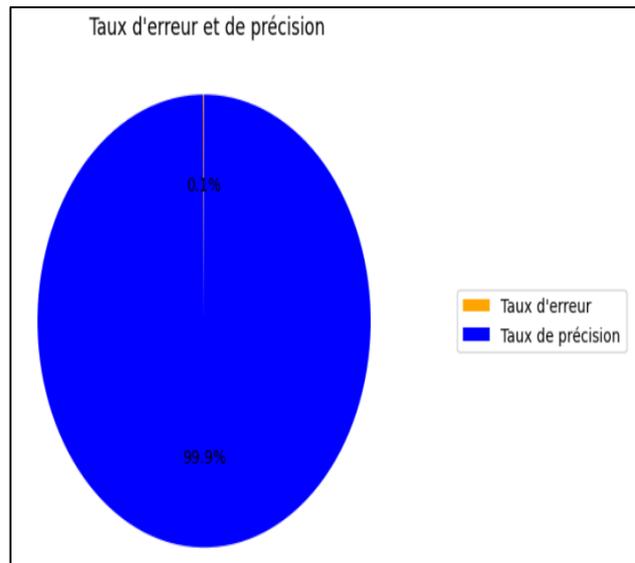
**Figure III.21** Précision et Erreur du **VGG16** sur 9238 images



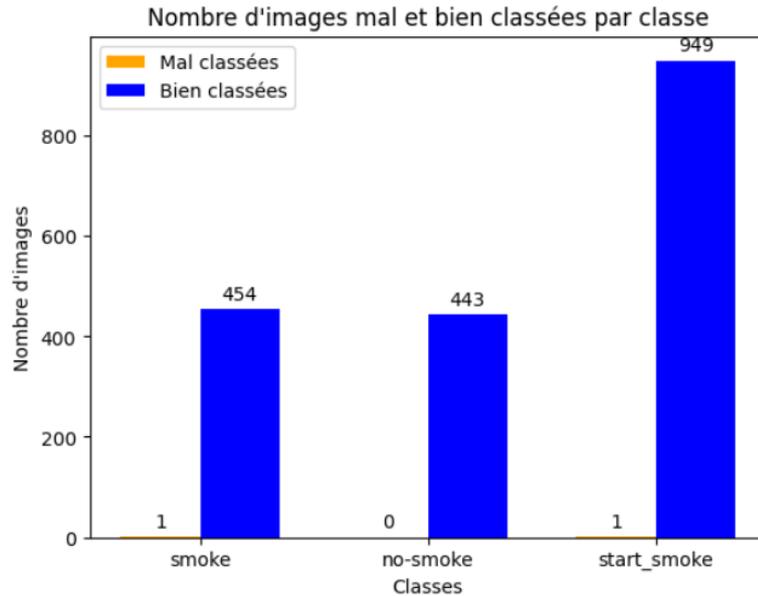
**Figure III.22** Matrice de confusion du VGG16



**Figure III.23** Nombre total des images mal et bien classé



**Figure III.24** Taux d'erreur et de précision



**Figure III.25** Nombre des images mal et bien classé par VGG16

Après avoir analysé les résultats obtenus, les observations suivantes ont été relevées :

D'après la **Figure III.21** la précision de l'apprentissage et de la validation augmente progressivement avec le nombre d'époques, de même, l'erreur d'apprentissage et de validation diminue progressivement avec le nombre d'époques.

Cela indique que le modèle apprend progressivement et devient de plus en plus précis dans la classification des données et aussi signifie qu'il parvient à réduire l'écart entre ses prédictions et les vérités terrain au fil de l'apprentissage.

D'après la **Figure III.23** et **Figure III.24** nous observons que le nombre total d'images mal classées est de seulement 2, ce qui correspond à un taux d'erreur de **0,1%**. En revanche, toutes les autres images (soit un total de 1846) ont été correctement classées, ce qui se traduit par un taux de précision de **99,9%**.

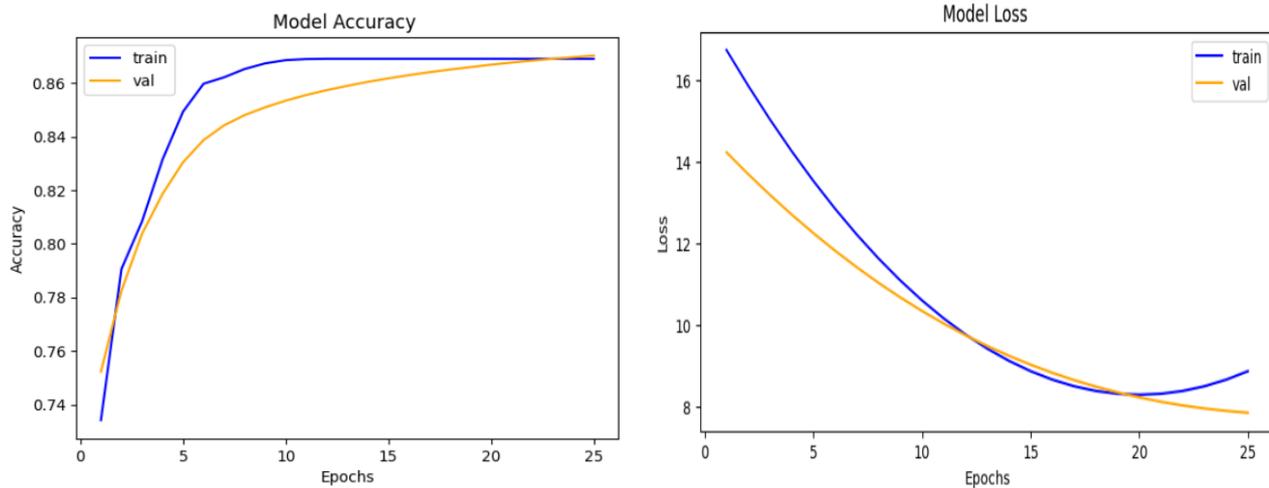
La matrice de confusion de la **Figure III.22** synthétise les résultats des prédictions effectuées par le modèle sur les images de test pour les trois classes:

- Classe 1 « **smoke** » : Il y a 454 échantillons qui appartiennent à cette classe et qui ont été correctement classés comme tels. Il y a 1 échantillon de cette classe qui a été mal classé comme appartenant à la classe 3 « start\_smoke ».
- Classe 2 « **no\_smoke** » : Il y a 443 échantillons qui appartiennent à cette classe et qui ont été correctement classés comme tels. Il n'y a aucun échantillon de cette classe qui a été mal classé comme appartenant à une autre classe.
- Classe 3 « **start\_smoke** » : Il y a 949 échantillons qui appartiennent à cette classe et qui ont été correctement classés comme tels. Il y a 1 échantillon de cette classe qui a été mal classé comme appartenant à la classe 2 « no\_smoke ».

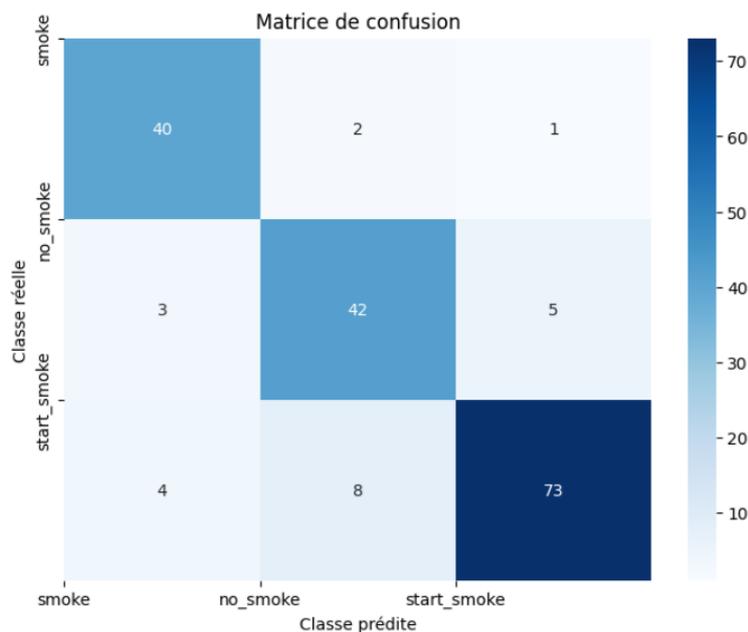
D'après la **Figure III.25** la plupart des échantillons ont été correctement classés dans leurs classes respectives. Les erreurs de classification sont relativement faibles, avec seulement 1 échantillon mal classé dans chaque classe, à l'exception de la classe 1 « no\_smoke » où tous les échantillons ont été correctement classés. Cela suggère une bonne performance globale du modèle de classification.

### III.6.3 Résultats obtenus par le modèle 3 « VGG19 » :

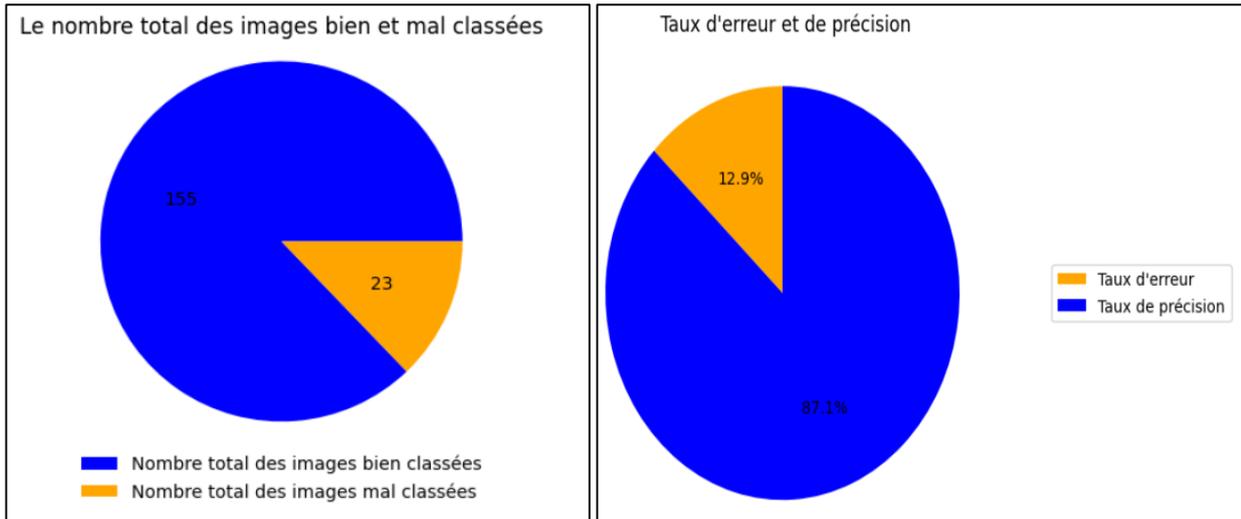
#### a. Résultats sur 893 images



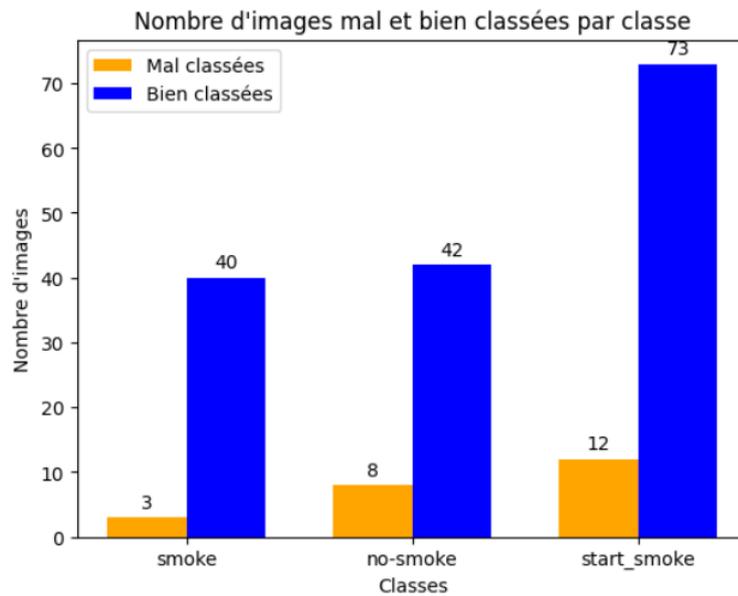
**Figure III.26** Précision et Erreur du **VGG19** sur 893 images



**Figure III.27** Matrice de confusion du **VGG19**



**Figure III.28** Nombre total des images mal et bien classé      **Figure III.29** Taux d'erreur et de précision



**Figure III.30** Nombre des images mal et bien classé par VGG19

Après avoir analysé les résultats obtenus, les observations suivantes ont été relevées :

D'après la **Figure III.26** les résultats de la précision d'apprentissage (train) et de la précision validation (val) montrent une tendance à l'augmentation avec le nombre d'époques. Cela suggère que le modèle s'améliore au fil du temps et apprend à mieux prédire les étiquettes correctes pour les données d'entraînement et de validation.

De même, l'erreur d'apprentissage (loss) montre une diminution régulière de l'erreur à mesure que le nombre d'époques augmente, l'erreur de validation (val) suit une tendance similaire à celle de l'apprentissage d'apprentissage jusqu'à atteindre une valeur d'environ 7.83.

Les résultats montrent que le modèle a une performance satisfaisante en termes d'apprentissage et de généralisation.

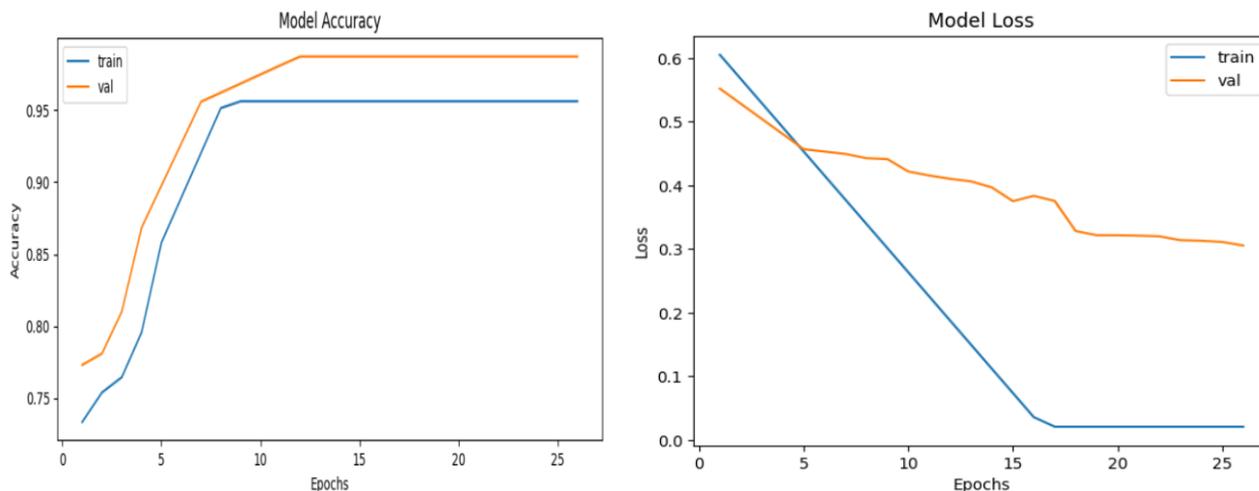
D'après la **Figure III.28** et **Figure III.29** nous observons que le nombre total d'images mal classées est 23, ce qui correspond à un taux d'erreur de **12.9%**. En revanche, toutes les autres images (soit un total de 155) ont été correctement classées, ce qui se traduit par un taux de précision de **87.1%**.

La matrice de confusion de la **Figure III.27** résume les prédictions du modèle sur les images de test pour les trois classes :

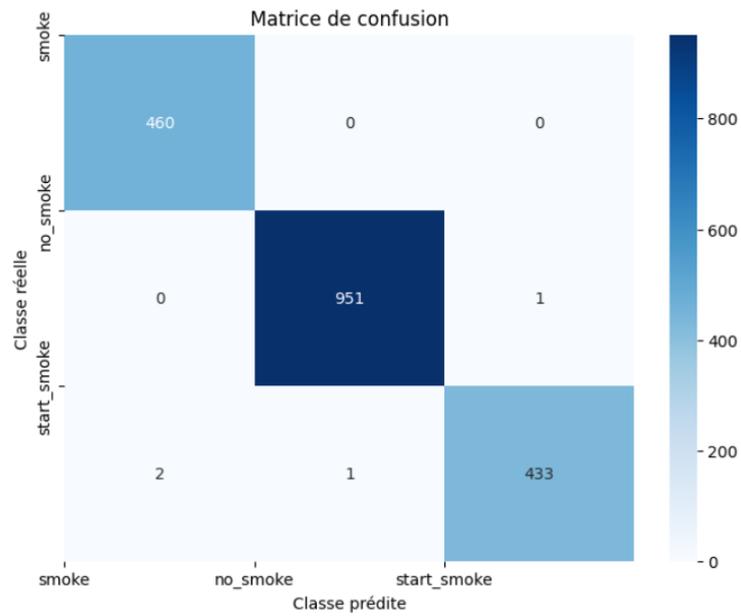
- Classe 1 « **smoke** »: le modèle a correctement prédit 40 échantillons, mais en a mal classé 2 comme appartenant à la classe 2 « no\_smoke » et 1 comme appartenant à la classe 3 « start\_smoke ».
- Classe 2 « **no\_smoke** »: le modèle a correctement prédit 42 échantillons, mais en a mal classé 3 comme appartenant à la classe 1 « smoke » et 5 comme appartenant à la classe 3 « start\_smoke ».
- Classe 3 « **start\_smoke** »: le modèle a correctement prédit 73 échantillons, mais en a mal classé 4 comme appartenant à la classe 1 « smoke » et 8 comme appartenant à la classe 2 « no\_smoke ».

D'après la **Figure III.30** la grande majorité des échantillons ont été correctement classés dans leurs classes respectives. Cependant, une observation importante est que la classe « start\_smoke » présente un nombre plus élevé d'images mal classées par rapport aux autres classes.

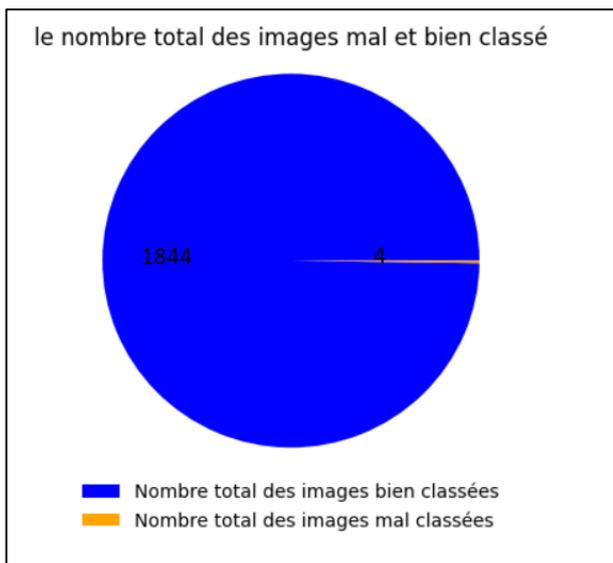
## **b. Résultats sur 9238 images**



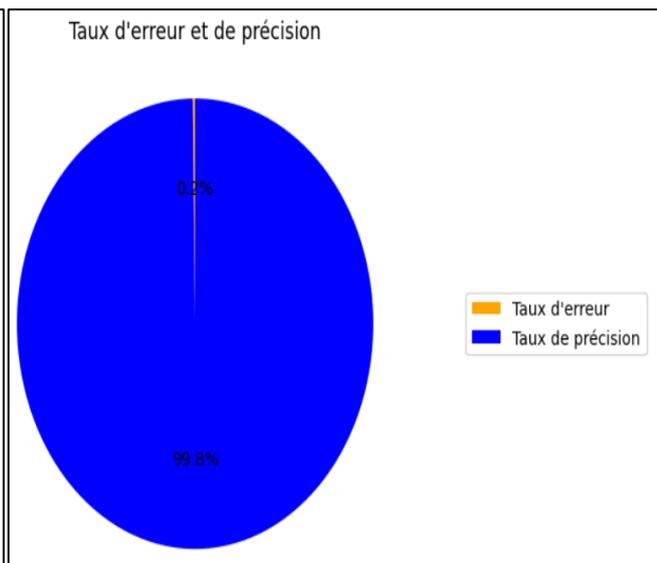
**Figure III.31** Précision et Erreur du **VGG19** sur 9238 images



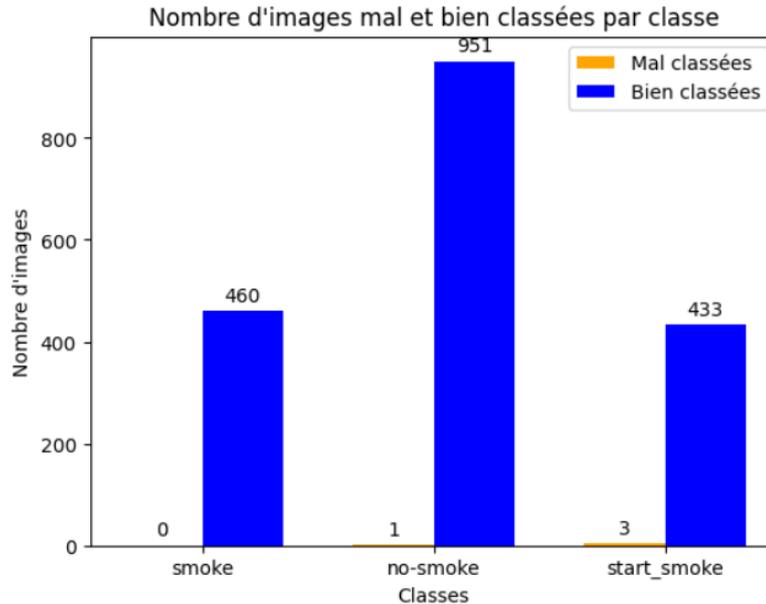
**Figure III.32** Matrice de confusion du VGG19



**Figure III.33** Nombre total des images mal et bien classé



**Figure III.34** Taux d'erreur et de précision



**Figure III.35** Nombre des images mal et bien classé par VGG19

Après avoir examiné les résultats obtenus, nous pouvons noter les observations suivantes :

D'après la **Figure III.31** la précision de l'apprentissage (train) augmente de manière significative au fil des époques, et la précision de la validation (val) augmente également rapidement jusqu'à atteindre une valeur de 0.9879.

De même, l'erreur d'apprentissage (train) et l'erreur de validation (val) diminue rapidement au début et atteignant une valeur d'environ 0.305.

Ces résultats indiquent que le modèle a atteint une performance très élevée avec une précision de 98.79% sur les données de validation. L'erreur est également très faible, ce qui montre que le modèle est capable de bien généraliser et de faire des prédictions précises sur de nouvelles données.

D'après la **Figure III.33** et **Figure III.34** nous constatons que le nombre total des images mal classées est très faible, seulement 4, ce qui représente un taux d'erreur de seulement 0,2%. En revanche, toutes les autres images, soit un total de 1844, ont été bien classées, ce qui correspond à un taux de précision élevé de **99,8%**.

La matrice de confusion de la **Figure III.32** résume les prédictions du modèle sur les images de test pour les trois classes :

- Classe 1 « **smoke** »: Sur les 460 échantillons de cette classe, tous ont été correctement classés. Il n'y a aucun échantillon de cette classe qui a été mal classé comme appartenant à une autre classe.
- Classe 2 « **no\_smoke** »: Sur les 952 échantillons de cette classe, 951 ont été correctement classés. Il y a 1 échantillon de cette classe qui a été mal classé comme appartenant à la classe 3 « **start\_smoke** ».

- Classe 3 « **start\_smoke** »: Sur les 436 échantillons de cette classe , 433 ont été correctement classés . Il y a 3 échantillons de cette classe qui ont été mal classés comme appartenant, un à la classe 2 « no\_smoke », et les deux autres à la classe1 « smoke ».

D’après la **Figure III.35** la majorité des échantillons ont été correctement classés dans leurs classes respectives. Les erreurs de classification sont minimales, avec seulement 4 échantillons mal classés dans chaque classe, à l'exception de la classe 1 « smoke » où tous les échantillons ont été correctement classés. Cette observation met en évidence la bonne performance globale du modèle de classification.

### III.6.4 Comparaison des résultats :

Les résultats obtenus sur les trois modèles, présentés dans le tableau ci-dessous :

|           | Architecture Utilisée |                   |                 | Nombre d'images | Nombre époque | Précision d'apprentissage | Précision de validation | Erreur | Temps d'exécution  |
|-----------|-----------------------|-------------------|-----------------|-----------------|---------------|---------------------------|-------------------------|--------|--------------------|
|           | Couche de Convolution | Couche de Pooling | Fully connected |                 |               |                           |                         |        |                    |
| Modèle 01 | 5                     | 2                 | 3               | 893             | 35            | 34,52                     | 33,02%                  | 68,50% | 660 s (11 min)     |
|           |                       |                   |                 | 9238            |               | 67,23%                    | 48,64%                  | 63,60% | 1920 s (32 min)    |
| Modèle 02 | 13                    | 5                 | 3               | 893             | 35            | 89,05%                    | 89,61%                  | 10,70% | 900 s (15 min)     |
|           |                       |                   |                 | 9238            |               | 100%                      | 98,34 %                 | 0,10%  | 6120 s (1 h 42min) |
| Modèle 03 | 16                    | 5                 | 3               | 893             | 25            | 87,75%                    | 88,23%                  | 12,90% | 420 s (7 min)      |
|           |                       |                   |                 | 9238            |               | 95,03%                    | 98,41%                  | 0,20%  | 4800 s (1h 20min)  |

**Tableau III.1 Comparaison des trois modèles (paramétrage et résultats)**

Le tableau III.1 présente l'architecture utilisée dans chaque modèle, le nombre d'images, le nombre d'époques, les résultats obtenus en termes de précision d'apprentissage et de validation, ainsi que l'erreur et le temps d'exécution. L'utilisation de GPU a permis de réduire le temps d'exécution malgré la taille importante de la base de données.

En général, les réseaux de neurones convolutionnels (CNN) de grande taille et profonds produisent de bons résultats. Nous confirmons cette information avec le tableau III.1 qui montre une nette amélioration des résultats pour les réseaux les plus profonds. Par exemple en passant du modèle 1 avec 5 couches de convolution au modèle 2 avec 13 couches de convolutions, l'erreur de classification passe de 68,50% à 10,70% et ceci pour la plus petite base de données.

L'amélioration des performances va en croissant avec l'augmentation du nombre d'époques (sans atteindre le sur-apprentissage), comme on peut le constater entre le modèle 3 plus profond que le modèle 2 mais qui devient moins performant que ce dernier avec seulement 25 époques contre 35 pour le modèle 2.

La taille de l'ensemble d'apprentissage est également un facteur déterminant dans les CNN. Nous le constatons également dans ce même tableau : prenons, par exemple le modèle 3 où l'erreur passe de 12,90% avec une base de 893 à 0,20% sur une base de 9238. Il est donc important d'avoir un ensemble d'apprentissage de grande taille pour obtenir de meilleurs résultats. Une base d'apprentissage plus importante permet au modèle d'apprendre à reconnaître une plus grande variété de motifs et de caractéristiques, ce qui peut conduire à des performances améliorées.

Enfin et toujours selon le tableau III.4, c'est le modèle 02 « **VGG16** » qui a donné les meilleurs résultats avec une précision d'apprentissage de 100%, une erreur de 0.1% et ceci sur la plus grande base de données. Le choix du nombre d'époques, la taille de la base de données et de couches de convolution a joué un rôle clé dans ces performances. Notons également le temps d'exécution qui augmente avec la taille de la base. Un compromis précision/temps d'exécution peut être discuté.

### **III.7 Conclusion**

Dans ce chapitre, nous avons présenté une approche de classification basée sur l'utilisation de réseaux de neurones convolutionnels (CNN). Pour ce faire, nous avons expérimenté, avec python et une programmation en ligne sous Google Colab, trois modèles avec différentes architectures sur deux tailles de base de données : le modèle expérimental avec, entre autres, 5 couches de convolutions, le modèle VGG19 et le modèle VGG16.

Nous avons évalué les résultats obtenus en termes de précision et d'erreur. Nous avons constaté que la performance augmente avec la profondeur du modèle, le nombre d'époques et la taille de la base d'apprentissage. La phase d'apprentissage peut prendre un temps significatif quand la base est de grande taille, mais cette taille est nécessaire pour atteindre une grande précision. En ajustant judicieusement ces paramètres on a obtenu les meilleurs résultats avec le modèle **VGG16** dont la précision atteint **99.9%**.

Cette étude confirme la robustesse de la classification par les réseaux de neurones convolutionnels en général et du VGG16 en particulier. On peut donc insérer ce module dans la chaîne de visualisation laparoscopique pour détecter les images avec fumée et enclencher un processus d'aspiration de fumée.

## Conclusion générale

Ce projet de fin d'études traite la détection et la classification de fumée dans les images laparoscopiques avec les réseaux de neurones convolutionnels (CNN) utilisés pour la classification.

Notre objectif principal était de trouver des réseaux (modèles) qui fonctionneraient au mieux avec le problème de la classification de la fumée dans les images laparoscopiques.

Au départ de ce mémoire, nous avons introduit les notions fondamentales de la chirurgie laparoscopique, qui constitue l'environnement de travail dans lequel nous avons extrait les images utilisées. Nous avons également présenté les principaux aspects de cette technique n mettant l'accent sur les complications qui peuvent être évités avec des méthodes intelligentes de détection tel que la détection de fumée sur laquelle nous nous sommes engagés.

Par la suite, nous avons exploré les techniques d'analyse d'image en particulier les plus récents domaines du Deep learning. L'état de l'art de la classification des images médicales confirme la domination de ces techniques intelligentes en comparaison avec les anciennes méthodes basées sur l'extraction manuelles des caractéristiques. Cependant la laparoscopie est encore gérée par les anciennes techniques telles que les SVMs.

Nous nous sommes donc investis pour explorer la pertinence de l'utilisation des réseaux de neurones convolutionnels (CNN) pour la laparoscopie. Nous avons expérimenté, avec python et une programmation en ligne sous Google Colab, trois modèles, dont deux étaient pré-entraînés (VGG16 et VGG19), tandis que le premier était un modèle que nous avons créé de toutes « pièces ». Nous avons détaillé l'architecture de chaque modèle et l'influences des différents composants et du paramétrage dans le processus de classification, notamment la couche de convolution, la couche de pooling et la couche fully connected.

Nous avons également discuté l'importance du choix du nombre d'époques, de la taille de la base de données pour obtenir des résultats optimaux dans la classification basée sur les CNN. Nous avons remarqué que l'utilisation de GPU a permis de réduire le temps d'exécution malgré la taille importante de la base de données.

En finalité, les résultats ont prouvé que le modèle VGG16 avec un bon paramétrage, est le plus performant avec une précision **99.9%** et une erreur **0.1%**. Ces résultats sont très encourageants pour introduire ce modèle dans le système laparoscopique.

---

## Les références

---

- Allodocteurs.fr. (2015). *La coelioscopie : une révolution chirurgicale*. Consulté le Mars 05, 2023, sur Allodocteurs.fr: <https://www.allodocteurs.fr/se-soigner-chirurgie-coelioscopie-la-coelioscopie-une-revolution-chirurgicale-216.html>
- Ameli.fr. (2019). *Comment se déroule une coelioscopie (ou laparoscopie) ?* Consulté le Février 07, 2023, sur Ameli.fr: <https://www.ameli.fr/assure/sante/examen/exploration/deroulement-coelioscopie>
- Andreas Leibetseder, M. J. (2017, Septembre 08). *Image-Based Smoke Detection in Laparoscopic Videos*. Consulté le Février 05, 2023, sur Springer Link: [https://link.springer.com/chapter/10.1007/978-3-319-67543-5\\_7](https://link.springer.com/chapter/10.1007/978-3-319-67543-5_7)
- Arora, S. &. (2022, Février 10). *Deep Learning for Brain Tumor Classification from MRI Images*. Consulté le Février 08, 2023, sur IEEE Xplore: <https://ieeexplore.ieee.org/document/9702609>
- B.Estelle. (2018, Septembre 12). *Coelioscopie : examen médical, indications et déroulement*. Consulté le Février 10, 2023, sur Santé sur le Net: <https://www.sante-sur-le-net.com/maladies/examens-medicaux/coelioscopie/>
- Beghdadi, A. C. (2020). Towards a video quality assessment based framework for enhancement of laparoscopic videos. *Medical Imaging 2020: Image Perception, Observer Performance, and Technology Assessment*.
- Belhadjer H, S. B. (2018). Classification des images avec les réseaux de neurones Convolutionnels. Tizi Ouzo, Informatique.
- Bizzotto, E. (2023, Avril 24). *4 choses à savoir avant une coelioscopie*. Consulté le Février 15, 2023, sur Santé Magazine: <https://www.santemagazine.fr/traitement/4-choses-a-savoir-avant-une-coelioscopie-361884>
- Boukhlof, D. (2005). Résolution de problèmes par écosystèmes : Application au traitement d'images. Université Mohamed Khider , Biskra: <http://thesis.univ-biskra.dz/id/eprint/2271> .
- Chalret Du Rieu, M. (2022, Août 24). *Coelioscopie : définition, déroulement, suites*. Consulté le Mars 03, 2023, sur Elsan Care: <https://www.elsan.care/fr/pathologie-et-traitement/maladie-digestive/c%C5%93lioscopie-definition-deroulement-suites>.
- Delmont Imaging. (s.d.). *Instruments de coelioscopie*. Consulté le Mars 12, 2023, sur Delmont Imaging: <https://www.delmont-imaging.com/produits-instruments-de-coelioscopie/>

- Docteur Benchimol. (2021, Septembre 11). *Coelioscopie*. Consulté le Mars 05, 2023, sur Docteur Benchimol: <https://www.docteur-benchimol.com/coelioscopie.html>.
- Ducrotoy. (s.d.). *La coelioscopie*. Consulté le Février 22, 2023, sur Centre Hospitalier Belvédère: <https://www.ch-belvedere.fr/chirurgie/chirurgie-gynecologique/la-coelioscopie/>
- Estelle. (2018, Septembre 12). *Coelioscopie : examen médical, indications et déroulement*. Consulté le Février 10, 2023, sur Santé sur le Net: <https://www.sante-sur-le-net.com/maladies/examens-medicaux/coelioscopie/>
- Estelle, B. (2018, Septembre 12). *Coelioscopie : examen médical, indications et déroulement*. Consulté le Février 10, 2023, sur Santé sur le Net: <https://www.sante-sur-le-net.com/maladies/examens-medicaux/coelioscopie/>
- Institut Mutualiste Montsouris. (s.d.). *Conseils après une chirurgie par voie coelioscopique (Hystérectomie, chirurgie des ovaires, cure de prolapsus par promontofixation, chirurgie des ganglions)*. Consulté le Mars 05, 2023, sur Institut Mutualiste Montsouris: <https://imm.fr/wp-content/uploads/2020/08/Conseils-apres-une-chirurgie-par-voie-coelioscopique.pdf>
- János Veress, K. S. (2022, Septembre 12). *La coelochirurgie : innovation et risque en chirurgie*. Consulté le Mars 05, 2023, sur INA (Institut National de l'Audiovisuel): <https://enseignants.lumni.fr/fiche-media/00000000679/la-coelochirurgie-innovation-et-risque-en-chirurgie.html>
- Laurence, P. (2021, Mars 04). *La chirurgie sous coelioscopie*. Consulté le Avril 03, 2023, sur ActuSoins: <https://www.actusoins.com/344852/la-chirurgie-sous-coelioscopie.html>
- Li, Y., & Zhu, H.-j. (2021, Février 16). *MFENet*. Consulté le Mars 24, 2023, sur IEEE Xplore: <https://ieeexplore.ieee.org/document/9312373>
- Meziani, I. (s.d.). *Segmentation d'image*. Consulté le Mars 20, 2023, sur DataFranca: [https://datafranca.org/wiki/Segmentation\\_d%27image](https://datafranca.org/wiki/Segmentation_d%27image)
- Miguel Molina-Moreno, I. G.-D.-S.-d.-M. (2022, Avril). *ACME*. Consulté le Mars 24, 2023, sur ScienceDirect: <https://doi.org/10.1016/j.media.2022.102358>
- Miguel Molina-Moreno, I. G.-D.-S.-d.-M. (2022, Avril). *ACME: Automatic feature extraction for cell migration examination through intravital microscopy imaging*. Consulté le Mars 24, 2023, sur sciencedirect.
- Nour Aldeen Jalal , T. A. (2017, Septembre 07). *Classifying smoke in laparoscopic videos using SVM*. Consulté le Février 06, 2023, sur De Gruyter: <https://www.degruyter.com/document/doi/10.1515/cdbme-2017-0040/html>
- Perdigón Romero, F. &. (2019, Juillet 11). *End-To-End Discriminative Deep Network For Liver Lesion Classification*. Consulté le Février 08, 2023, sur IEEE Xplore: <https://ieeexplore.ieee.org/document/8759257>

Reiter, W. (2020, Septembre 17). *Improving endoscopic smoke detection with semi-supervised noisy student models*. Consulté le Février 05, 2023, sur De Gruyter: <https://www.degruyter.com/document/doi/10.1515/cdbme-2020-0026/html>

Société canadienne du cancer. (s.d.). *Laparoscopie*. Consulté le Mars 12, 2023, sur Société canadienne du cancer: <https://cancer.ca/fr/treatments/tests-and-procedures/laparoscopy#:~:text=La%20laparoscopie%20est%20une%20intervention,lumineuse%20et%20d'une%20lentille>

URPS Infirmière PACA. (2019, Septembre 06). *Le bon usage des antiseptiques chez l'adulte*. Consulté le Mars 07, 2023, sur URPS Infirmière PACA: <https://www.urps-infirmiere-paca.fr/les-bonnes-pratiques/le-bon-usage-des-antiseptiques-chez-ladulte/>

Yann LeCun, L. B. (1998, November). *Gradient-based learning applied to document recognition*. Consulté le Mars 25, 2023, sur IEEE Xplore: <https://ieeexplore.ieee.org/document/726791>

Z. A. Khan, A. B.-C. (2020, October). *Residual Networks based Distortion Classification and Ranking for Laparoscopic Image quality Assessment*. Consulté le Février 03, 2023, sur IEEE- ICIP2020.

Zakaria, M. M. (2017, Juillet 03). *Classification des images avec les réseaux de neurones*. Tlemcen, Informatique.

Zohra, H. K. (2015, Juin 15). *Détection automatique des nodules pulmonaires*. Département de Génie Electrique et Electronique, Tlemcen.

