



N° d'ordre : 15/D3C/2022

N° de série : 03/Elec/2022

THÈSE

EN VUE DE L'OBTENTION DU DIPLOME DE
DOCTORAT 3^{ème} CYCLE

Filière : **Génie biomédical**

Spécialité : **Procédés et dispositifs pour le biomédical**

Présentée par :

BOUYAYA Dallel

Thème

***Analyse des images de capsule
endoscopique***

Soutenue publiquement le 24 /03 /2022 devant le jury :

Président	Mr BENHABILES Mohamed Taoufik	Prof	Univ. Frères Mentouri Constantine 1
Rapporteur	Mr BENIERBAH Said	Prof	Univ. Frères Mentouri Constantine 1
Examineurs	Mme HACHOUF Fella	Prof	Univ. Frères Mentouri Constantine 1
	Mr FERDI Youcef	Prof	Ecole National Supérieur de Biotechnologie Constantine
	Mr MESSAOUDI Kamel	Prof	Univ. Souk Ahras

REMERCIEMENTS

*Avant tout, Je remercie **ALLAH** le tout Puissant pour la volonté, la santé et la patience qu'il m'a données durant toutes ces longues années pour réaliser cette thèse.*

*Ce travail de recherche a été effectué au Laboratoire de traitement du signal de l'Université des frères Mentouri Constantine 1, sous la direction scientifique de Monsieur le professeur **Said. BENIERBAH**, que je tiens à remercier très profondément pour avoir accepté de m'encadrer et me proposer le sujet de cette thèse. Je tiens à exprimer ma profonde gratitude pour ses précieux conseils et son suivi pour finaliser les travaux de cette thèse.*

*Je tiens à remercier très vivement Monsieur **Mohamed Toufik. BENHABILES**, Professeur à l'Université des Frères Mentouri Constantine 1, pour m'avoir fait l'honneur d'accepter de présider le jury de cette thèse.*

*J'exprime également mes vifs remerciements à Madame **Fella. HACHOUF** Professeur à l'Université des Frères Mentouri Constantine 1, Monsieur **Youcef. FERDI** Professeur à l'école National Supérieur de Biotechnologie Constantine et Monsieur **Kamel. MESSOUDI** Professeur à l'université de Souk Ahras, pour avoir accepté d'être examinateurs de cette thèse et pour le temps qu'ils ont consacré à la lecture, l'analyse et à l'évaluation de ce travail.*

Je souhaite remercier du plus profond de mon cœur mes chères parents et toute ma famille qui m'ont soutenu durant mes études et qui m'ont toujours encouragé.

Je voudrais exprimer ma reconnaissance envers mes amies et collègues qui m'ont apporté leur support moral et intellectuel tout au long de ces années, un grand merci.

Enfin, je remercie tous ceux qui, de près ou de loin, ont contribué à la réussite de ce travail.

Résumé

La capsule endoscopique sans fil (WCE) est un dispositif non invasif, introduit en 2000, qui est utilisé par les médecins dans le diagnostic des maladies du tractus gastro-intestinal (GI). Malgré ces ressources limitées, une capsule peut fonctionner pendant plusieurs heures et transmettre des dizaines de milliers d'images. Cependant, ces images sont à faible qualité et fréquence d'image. En plus, cet outil ne parviendra pas à visualiser tout le tractus GI, car le temps de vie de la batterie est limité. Le but principal du présent travail est de développer des algorithmes de traitement automatique de ces images, pour un aide au diagnostic et à la réduction de la consommation d'énergie. Pour une première contribution, nous proposons dans un premier temps, une classification automatique des lésions et des organes digestifs dans les images WCE. Cette classification est réalisée en utilisant deux techniques d'apprentissage telles que l'apprentissage à partir de zéro d'un CNN proposé et un apprentissage par transfert des CNN pré-entraînées. Dans un deuxième temps, nous présentons une nouvelle méthode de classification pour détecter automatiquement les différentes maladies du tractus GI. Il s'agit d'un algorithme d'apprentissage profond basé sur la concaténation des caractéristiques de deux réseaux de neurones convolutifs pré-entraîné. Dans une deuxième contribution, nous proposons un système de compression intelligent, qui répond aux problèmes de limitations énergétiques de la WCE. Le principe est d'inclure une boucle de rétroaction de classification, basée sur l'apprentissage profond, pour déterminer l'importance des images transmises. Cette classification est utilisée avec un algorithme de compression prédictif permettant une gestion intelligente de l'énergie limitée de la capsule. Le but d'un tel système est d'augmenter la durée de vie de la batterie ou d'obtenir des images de haute qualité dans des zones spécifiques. D'après les résultats obtenus, nous concluons que notre système est efficace et offre une bonne optimisation énergétique de la capsule.

Mots clés : la capsule endoscopique sans fil, apprentissage profond, apprentissage par transfert, classification, compression.

Abstract

Wireless capsule endoscopic (WCE) is a non-invasive device, introduced in 2000, that is used by physicians in the diagnosis of diseases of the gastrointestinal (GI) tract. Despite these limited resources, one capsule can operate for several hours and transmit tens of thousands of images. However, these images are of low quality and frequency. In addition, this tool will not be able to visualize the entire GI tract, as the battery life is limited. The main goal of the present work is to develop algorithms for the automatic processing of these images, to help in the diagnosis and in the reduction of the energy consumption. As a first contribution, we propose an automatic classification of lesions and digestive organs in WCE images. This classification is achieved by using two learning techniques such as learning from scratch of a proposed CNN and a transfer learning of pre-trained CNNs. In a second step, we present a new classification method to automatically detect different diseases of the (GI) tract. It is a deep learning algorithm based on features concatenation of two pre-trained convolutional neural networks. In a second contribution, we propose an intelligent compression scheme, which addresses the energy limitation issues of WCE. The principle is to include a classification feedback loop, based on deep learning, to determine the importance of transmitted images. This classification is used in conjunction with a predictive compression algorithm to intelligently manage the limited energy of the capsule. The goal of such a system is to increase the battery life or to obtain high quality images in specific areas. Based on the results obtained, we conclude that our system is efficient and provides good energy optimization of the capsule.

Keywords: wireless capsule endoscopic, deep learning, transfer learning, classification, compression.

ملخص

الكبسولة التنظيرية اللاسلكية (WCE) هو جهاز غير جراحي ، تم طرحه في عام 2000 ، ويستخدمه الأطباء في تشخيص أمراض الجهاز الهضمي. على الرغم من هذه الموارد المحدودة ، يمكن للكبسولة أن تعمل لعدة ساعات وتنقل عشرات الآلاف من الصور. ومع ذلك ، فإن هذه الصور منخفضة الجودة ومعدل الإطارات. بالإضافة إلى ذلك ، لن تتمكن هذه الأداة من تصور الجهاز الهضمي بالكامل ، لأن عمر البطارية محدود. الهدف الرئيسي من هذا العمل هو تطوير خوارزميات للمعالجة التلقائية لهذه الصور ، للمساعدة في التشخيص وتقليل استهلاك الطاقة. للمساهمة الأولى ، نقتراح أولاً تصنيفًا تلقائيًا للآفات وأعضاء الجهاز الهضمي في صور WCE. يتم تحقيق هذا التصنيف باستخدام طريقتين للتعلم مثل التعلم من الصفر لشبكة CNN المقترحة ونقل التعلم لشبكات CNN المدربة مسبقًا. ثانيًا ، نقدم طريقة تصنيف جديدة للكشف تلقائيًا عن أمراض الجهاز الهضمي المختلفة. إنها خوارزمية تعلم عميق تعتمد على تسلسل ميزات شبكتين عصبيتين تلافيفيتين مدربتين مسبقًا . في مساهمة ثانية ، نقتراح نظام ضغط ذكيًا يعالج مشاكل الحد من الطاقة في WCE. المبدأ هو تضمين حلقة التغذية المرتدة للتصنيف ، بناءً على التعلم العميق ، لتحديد أهمية الصور المنقولة. يستخدم هذا التصنيف مع خوارزمية ضغط تنبؤية تسمح بإدارة ذكية للطاقة المحدودة للكبسولة. الغرض من هذا النظام هو زيادة عمر البطارية أو الحصول على صور عالية الجودة في مناطق معينة. من النتائج التي تم الحصول عليها ، نستنتج أن نظامنا فعال ويقدم تحسينًا جيدًا للطاقة للكبسولة.

الكلمات المفتاحية: الكبسولة التنظيرية اللاسلكية ، التعلم العميق ، نقل التعلم ، التصنيف ، الضغط

Table des matières

Liste des figures	VII
Liste des tableaux.....	X
Liste d'abréviations.....	XI
Introduction générale.....	1

Chapitre 1 : La capsule endoscopique

1.1	Introduction	4
1.2	Historique.....	4
1.3	Système de capsule endoscopique.....	6
1.3.1	La capsule.....	7
1.3.1.1	Capteur d'image	7
1.3.1.2	Système d'illumination.....	8
1.3.1.3	Système de transmission.....	8
1.3.1.4	L'alimentation	10
1.3.2	Le dispositif externe d'enregistrement.....	11
1.3.3	La station de travail	11
1.4	Différent types des capsules disponibles dans le commerce.....	12
1.5	Les applications.....	13
1.5.1	La Capsule de l'intestin grêle.....	13
1.5.2	La Capsule œsophagienne	14
1.5.3	La Capsule colique	14
1.6	Les limites et les perspectives	15
1.7	Techniques de compression d'images WCE.....	16
1.7.1	Techniques de compression d'images sans perte.....	17
1.7.1.1	Compression d'images basée sur JPEG-LS	17
1.7.1.2	Techniques basées sur le DPCM	17
1.7.2	Techniques de compression d'images avec perte	18
1.7.2.1	Compression d'images basées sur la DCT	18
1.8	Conclusion	19

Chapitre 2 : Notions de base sur l'apprentissage profond

2.1	Introduction	20
2.2	L'apprentissage automatique.....	20
2.2.1	Paradigmes d'apprentissage	20
2.2.1.1	Apprentissage supervisé	20
2.2.1.2	Apprentissage non-supervisé	22
2.2.1.3	Apprentissage semi-supervisé	22
2.2.1.4	Apprentissage par renforcement	22
2.3	Les réseaux de neurones artificiels.....	22
2.3.1	Fonctionnement d'un neurone artificiel	23
2.3.2	Réseaux de neurone monocouche	26
2.3.3	Réseaux de neurone multicouche	27
2.3.4	Apprentissage et rétropropagation.....	27
2.3.4.1	Notations.....	27
2.3.4.2	Coût d'erreur.....	28
2.3.4.3	Descente de gradient et perceptron multicouche	28
2.3.4.4	Apprentissage et rétropropagation des erreurs	32
2.3.5	Problème de sur-apprentissage et sous-apprentissage.....	33
2.4	Les réseaux de neurones profonds.....	36
2.4.1	Réseaux de neurones convolutifs	36
2.4.2	Architecture d'un CNN	36
2.4.2.1	Couche de convolution (CONV)	37
2.4.2.2	Couche de mise en commun	39
2.4.2.3	Couche entièrement connectée (FC).....	39
2.5	Apprentissage par transfert.....	40
2.6	DCNN architectures	41
2.7	L'évaluation de la performance des modèles d'apprentissage automatique	42
2.7.1	La matrice de confusion	42
2.7.2	La courbe ROC.....	44
2.8	Conclusion	45

Chapitre 3 : Classification des images WCE en utilisant l'apprentissage profond

3.1	Introduction	47
3.2	Etat de l'art sur la classification des images WCE.....	47

3.3	Classification en utilisant l'apprentissage à partir de zéros	49
3.3.1	Base de données	50
3.3.2	Augmentation de données	51
3.3.3	La validation croisée	53
3.3.4	Les paramètres d'apprentissage utilisés :	53
3.3.5	Résultats et discussion.....	54
3.4	Classification en utilisant la méthode d'apprentissage par transfert.....	58
3.4.1	Paramètres d'apprentissage	59
3.4.2	Résultats et discussion.....	59
3.5	Classification en utilisant la concaténation de caractéristiques.....	70
3.5.1	Description de la méthode proposée pour la classification des maladies GI	71
3.5.2	Base de données	71
3.5.3	Augmentation de données	74
3.5.4	Paramètres d'apprentissage	75
3.5.5	Résultats et discussion.....	75
3.5	Conclusion	83

Chapitre 4 : Système de compression des images WCE proposé

4.1	Introduction	84
4.2	Etat de l'art sur la compression des images WCE.....	84
4.3	La méthode proposée	87
4.3.1	Le codeur.....	89
4.3.1.1	Conversion de RVB en YUV	89
4.3.1.2	Sous échantillonnage	91
4.3.1.3	Codage prédictif DPCM	92
4.3.1.4	Codage Golomb-Rice	93
4.3.2	Le décodeur	95
4.4	Résultats et discussion.....	96
4.4.1	Effet de sous-échantillonnage sur la classification	96
4.4.2	Analyse des performances de la compression proposée	97
4.5	Conclusion	105
	Conclusion générale.....	106
	Bibliographies.....	108

Liste des Figures

Figure 1. 1 Le premier prototype de la capsule.	5
Figure 1. 2 Système de capsule endoscopique : capsule, dispositif d'enregistrement et post de travail.	7
Figure 1. 3 Composants de la capsule endoscopique.	7
Figure 1. 4 Schéma de la technique de transmission radiofréquence.	9
Figure 1. 5 Schéma de la technique de transmission via le corps humain.	9
Figure 1. 6 Schéma de la technique de transmission via stockage de données intégrées.	10
Figure 1. 7 Enregistreur de données.	11
Figure 1. 8 Images du (a) estomac et (b) intestin grêle capturée par une capsule.....	13
Figure 1. 9 (a) capsule l'intestin grêle, (b) capsule œsophage, (c) capsule colique.....	15
Figure 2.1 Exemple d'une régression linéaire.	21
Figure 2.2 Exemple d'une classification avec des données d'entrées x.	21
Figure 2.3 Modèle d'un neurone artificiel.	24
Figure 2.4 Tracé de la fonction sigmoïde.....	24
Figure 2.5 Tracé de La fonction tangente hyperbolique	25
Figure 2.6 Tracé de La fonction Rectified Linear Unit.....	26
Figure 2.7 Schéma d'un réseau de neurones monocouche.	26
Figure 2.8 Schéma d'un réseau de neurones multicouches.	27
Figure 2. 9 Illustration des deux étapes successives de propagation avant puis rétropropagation pour calculer les z et δ intervenant dans les dérivée partielle dE/dw_{ji} . Dans cette figure, on a isolé le neurone j, et représenté uniquement les neurones qui lui sont connectés dans la couche précédente et dans la couche suivante.	31
Figure 2.10 Les différents niveaux d'apprentissage.	33
Figure 2.11 Fonction de coût calculée sur l'ensemble d'apprentissage et sur l'ensemble de validation en fonction du nombre d'époque.	34
Figure 2.12 Principe de dropout.	35
Figure 2.13 Architecture standard d'un réseau de neurones convolutionnel.	37
Figure 2.14 Principe d'une opération de convolution.	38
Figure 2.15 Types de mise en commun.....	39
Figure 2. 16 Exemple de courbes ROC pour des classes multiples.	45
- Figure 3.2 Schéma fonctionnel de la méthodologie proposée.	49

Figure 3.1 Schéma fonctionnel de la méthodologie proposée.	50
Figure 3.3 Exemple des images de notre base de données.....	51
Figure 3.5 Résultat de l'application des techniques d'augmentation des données : (a) image originale, (b) retournement horizontal, (c) correction de la luminosité, (d) rotation, (e) ajustement du contraste, (f) retournement vertical.....	52
Figure 3.4 Equilibrage des données en utilisant les techniques d'augmentation des données.	52
Figure 3.6 Validation croisée à 5 échantillons, dite 5-folds.....	53
Figure 3.7 Matrice de confusion de la classification des lésions.	55
Figure 3.8 Matrice de confusion de la classification des organes digestifs.	56
Figure 3.9 Résultats de prédiction obtenus par la classification des lésions.....	57
Figure 3.10 Résultats de prédiction obtenus par la classification des organes digestifs.....	57
Figure 3.11 Organigramme de l'expérience globale.....	60
Figure 3.12 Matrice de confusion de la classification de lésions obtenues par différents modèles DCNN.....	61
Figure 3. 13 Courbes de précision et de perte de différents modèles pour la classification des lésions.	65
Figure 3.14 Matrice de confusion de la classification des organes digestifs obtenus par.....	66
Figure 3. 15 Courbes de précision et de perte de différents modèles pour la classification des organes digestifs.....	70
Figure 3.16 Schéma fonctionnel de la méthode proposée.....	70
Figure 3.17 Schéma de l'architecture proposée	71
Figure 3.18 Exemples d'images des ensembles de données Kvasir-Capsule.....	74
Figure 3. 19 Résultats de classification obtenus par la méthode proposée pour différentes époques.	76
Figure 3. 20 Cartes de caractéristiques des couches convolutives apprises par notre modèle. .	77
Figure 3. 21 Comparaison des résultats de classification obtenus par notre méthode et classificateurs individuelles.	78
Figure 22 Courbes de précision et de perte pour (a) mobilenet, (b) densenet169 et (c) méthode proposée.	79
Figure 3. 23 Comparaison des courbes ROC de différents modèles.....	81
Figure 3.24 Matrice de confusion de la classification des maladies GI.....	82
Figure 4. 1 Schéma de principe de la méthode de codage proposée	88
Figure 4.2 Les différentes composantes YUV d'une image endoscopique	89
Figure 4.3 Histogramme des composantes RVB et YUV d'une image endoscopique.....	90

Figure 4.4 Schéma de sous-échantillonnage YUV444, YUV422 et YUV420.	91
Figure 4.5 Valeurs de la différence $e(i)$ des composantes R, V et B.	92
Figure 4.6 Valeurs de la différence $e(i)$ des composantes Y, U et V.	92
Figure 4. 7 Matrice de confusion de la classification des lésions et de l'anatomie.	98
Figure 4.8 Nombre d'images 444 gagnées qui sont obtenues par sous-échantillonnage (444,422) et (444,420), avant compression pour chaque séquence de test.	102
Figure 4. 9 Un exemple des positions des images classées, pour la séquence anatomique test. La classe 1 correspond à la région d'intérêt (côlon) et la classe 0 aux autres régions (estomac, intestin grêle).	104

Liste des Tableaux

Tableau 1. 1 Principales caractéristiques des capsules WCE disponibles dans le commerce. [23].....	12
Tableau 2. 1 Exemple de matrice de confusion pour une classification binaire	43
Tableau 3.1 Résultats de la classification des lésions et des organes digestifs évalués sur 5 folds par validation croisée.	54
Tableau 3.2 Performance de la classification de lésions obtenues par l'apprentissage par transfert.	60
Tableau 3.3 Performance de la classification des organes digestifs obtenus par l'apprentissage par transfert.	65
Tableau 3. 4 Le nombre d'images dans les différentes classes pour les jeux de données Kvasir- Capsule.....	72
Tableau 3. 5 Performance de classification de la méthode proposé sur différentes classes. ...	76
Tableau 3.6 Comparaison des résultats de classification obtenus par différents CNN pré- entraînés.	83
Tableau 4. 1 Résultats de la précision avec 95% CI du réseau de classification proposé sur les ensembles de données de lésions et d'anatomie.....	97
Tableau 4. 2 Nombre d'images normales et anormales, pour chaque séquence de test, tel que déterminé par notre classificateur.	99
Tableau 4. 3 Taux de compression et PSNR obtenus par notre système pour les séquences de test WCE.....	100
Tableau 4. 4 Une comparaison du CR et de la qualité des différents schémas de sous- échantillonnage	101
Tableau 4. 5 Bits gagnés par le système de compression proposé.....	103

Liste d'abréviations

CCD	Charge coupled device
CE	Capsule endoscopique
CMOS	Complementary metal oxide semiconductor
CNN	Convolutional neural network
CR	Compression ratio
DB	Decibel
DCT	Discrete Cosine Transform
DL	Deep learning
DPCM	Differential pulse-code modulation
DWT	discrete wavelet transform
FN	Faux négatif
FP	Faux positif
GI	Gastro-intestinale
GPU	Graphics processing unit
GR	Golomb rice
IA	Intelligence artificielle
LED	White light emitting diode
MCC	Matthews correlation coefficient
ML	Machine learning
MLP	Multi layer perceptron
PSNR	Peak signal-to-noise ratio
RELU	Rectified Linear Unit
RNA	Réseau de neurone artificiel
ROC	Receiver Operating Characteristic
SGD	Descente de gradient stochastique
TFP	Taux faux positif
TVP	Taux vrai positif
VN	Vrai négatif
VP	Vrai positif
VPP	Valeur prédictive positive
WCE	Wireless capsule endoscopic

Introduction générale

Les maladies gastro-intestinales (GI), telles que les polypes, les hémorragies et les cancers de l'intestin, sont devenues de grandes menaces pour la santé de la population. La meilleure façon de détecter les maladies gastro-intestinales est de visualiser directement le tube digestif, ce qui fait de l'endoscopie une technologie de diagnostic efficace [1]. Les systèmes d'endoscope traditionnels peuvent fournir une localisation précise des lésions, puisque la caméra est déplacée vers l'avant et vers l'arrière, par des humains. De plus, ces systèmes disposent de sources d'énergie illimitées qui leur permettent de produire des images à haute résolution. Cependant, ils sont douloureux et ne peuvent pas atteindre l'intestin grêle. Pour résoudre ce problème, la WCE a été développée. Elle a été proposée pour la première fois par Given Imaging (Intromedic) en 2000 [2], [3], et a été suivie par de nombreuses autres sociétés telles que Olympus [4] et Sayaka [5].

La WCE permet d'examiner l'ensemble du tube digestif, y compris l'intestin grêle, de manière non invasive et sans douleur. Le principe consiste en un dispositif ayant la forme d'une pilule et composé de quelques composants miniatures. Après avoir été avalé par un patient, la WCE prend des images du tube digestif pendant plusieurs heures et les transmet à un dispositif d'enregistrement de données externe. Plus tard, les médecins visionneront les images enregistrées, afin de fournir un diagnostic final. Bien que les résultats de la WCE soient prometteurs, certaines améliorations doivent encore être apportées à cet appareil, notamment l'augmentation de la durée de vie de sa batterie, de la qualité de ses images et de sa fréquence d'images (image par second). Le problème principal du WCE vient de ses réserves d'énergie réduites et limitées qui proviennent de petites piles boutons. La majeure partie de cette énergie est consommée par le module de communication du WCE [6]. Pour cela, il est nécessaire de développer et d'utiliser des techniques de compression efficaces, avec une très faible complexité, pour réduire le nombre des bits transmis. Dans les capteurs sans fil, le calcul peut consommer moins que la transmission, et en réduisant le nombre de bits transmis, nous réduisons l'énergie nécessaire à la transmission [7], [8].

Cependant, la compression seule n'est pas suffisante, car la WCE a d'autres contraintes qui affectent son fonctionnement et son utilisation dans le diagnostic. La WCE se déplace dans une seule direction et même lorsqu'une lésion importante est observée, elle ne peut pas faire l'objet d'une attention plus particulière, comme c'est le cas en endoscopie filaire. De plus, le

mouvement est irrégulier et parfois la WCE s'arrête de bouger et produit les mêmes images. Enfin, toutes ces images sont codées de la même manière, indépendamment de leur importance pour le diagnostic. En fait, même après avoir obtenu toutes les images, leur lecture prend du temps. En général, un WCE produit environ 86 000 images pour un patient, au cours d'un examen [9]. La lecture de ces images révèle que seule une petite fraction d'entre elles est réellement importante pour le diagnostic. Cependant, elles ont toutes été traitées de la même manière et ont consommé une énergie similaire.

Récemment, la détection automatique des lésions dans les images de la WCE a été un domaine de recherche actif afin de réduire la charge de lecture des images pour les médecins [10]. Cependant, ces opérations sont effectuées après que les images ont été transmises et enregistrées. Si ces informations étaient connues avant, elles auraient pu être utilisées pour améliorer la qualité des images importantes ou pour produire davantage d'images dans les régions importantes. Dans cette thèse, nous proposons de réaliser une classification intelligente des images WCE pendant la compression et la transmission. Notre système suivra la transmission de chaque image et décidera de son importance pendant sa transmission. L'objectif est de réduire l'utilisation de l'énergie en effectuant un codage avec perte pour les images moins importantes. Ainsi, sur la base du résultat de la classification, nous réduirons les bits transmis par les images moins importantes et les utiliserons pour transmettre plus d'images ou pour augmenter la qualité des images importantes. Cette technique améliore l'efficacité énergétique du système, tout en maintenant une qualité élevée pour les images d'intérêt ou en augmentant leur nombre.

Cette thèse est divisée en quatre chapitres. Le premier chapitre concerne une description générale des capsules endoscopiques, où l'on a présenté les différents types de capsules, leurs compositions, leurs modes d'utilisation ainsi que leurs limitations. Ensuite, nous avons montré l'importance de la compression dans les capsules endoscopiques comme une solution pour réduire la consommation d'énergie. Pour cela, on a présenté les différentes techniques de compression qui peuvent être utilisés pour les images de capsule endoscopique.

Le deuxième chapitre explore l'utilisation de l'apprentissage automatique pour classer des images et ses méthodologies. Nous abordons également divers concepts importants couvrants les principes fondamentaux de l'apprentissage automatique tels que, les types d'apprentissage, le modèle du réseau de neurones qui est à la base de l'apprentissage profond ainsi que les

Introduction générale

notions des réseaux de neurones convolutionnels qui est la base de notre travail, et enfin nous présentons les techniques d'apprentissage par transfert et leurs architectures standards.

Dans le troisième chapitre, nous proposons une classification automatique des lésions et des organes digestifs dans les images WCE, cette classification est réalisée en utilisant deux techniques d'apprentissage telles que l'apprentissage à partir de zéro d'un CNN proposé et un apprentissage par transfert des CNN pré-entraînées. On a proposé aussi une nouvelle méthode de classification pour détecter automatiquement les différentes maladies du tractus gastro intestinale. Il s'agit d'un algorithme d'apprentissage profond basé sur la combinaison de deux réseaux de neurones convolutifs pré-entraînés, à savoir MobileNet et DenseNet169. Les caractéristiques extraites produites par la méthode proposée sont ensuite introduites dans un classificateur multicouche (MLP). Ensuite nous allons discuter les résultats obtenus.

Dans le quatrième chapitre, nous proposons un système de compression intelligent qui inclut une classification basée sur l'apprentissage profond dans la boucle de rétroaction d'un système de compression basé sur la prédiction. Nous avons implémenté à la fois les opérations de compression et de classification intelligente utilisées pour coder de nombreuses séquences de test d'images WCE. Ensuite, nous allons montrer notre méthode en détail, et enfin, nous allons analyser les résultats de la compression pour voir la faisabilité de tels systèmes de compression intelligents et leurs gains possibles.

Une conclusion générale résume les principales réalisations de notre thèse et elle identifie certaines orientations pour les futurs travaux dans le domaine de la capsule endoscopique.

Chapitre 1

La capsule endoscopique

1.1 Introduction

Les maladies gastro-intestinales (GI) sont devenues une menace énorme pour la santé humaine. Le cancer gastrique, par exemple, se classe au quatrième rang parmi les types de cancers les plus courants dans le monde et est la deuxième cause de décès par cancer dans le monde [11]. La détection et le diagnostic précoces des maladies GI deviennent plus cruciaux. La meilleure façon de détection est de visualiser directement le tractus GI, ce qui fait de l'endoscopie une technologie de diagnostic directe et efficace [1]. Les techniques traditionnelles d'endoscopie peuvent fournir une localisation précise des lésions. Cependant, elles sont douloureuses et ne peuvent pas atteindre l'intestin grêle. Cette limitation est un facteur qui a contribué à au développement d'un nouveau type d'endoscopie, appelée la capsule endoscopique sans fil (WCE). Elle a été introduite en 2000 par Given Imaging Incorporation et mise en œuvre en 2001 [12]. Cette nouvelle technique permet de visualiser l'ensemble du tube digestif, y compris l'intestin grêle, sans douleur ni sédation. Dans ce chapitre, on présentera un bref historique sur la capsule endoscopique (CE), son principe de fonctionnement et technologie, ainsi les différents types disponibles dans le commerce, et en fin leurs limites et perspectives.

1.2 Historique

Les techniques endoscopiques traditionnelles (c'est-à-dire celles qui exploitent des endoscopes flexibles) peuvent être utilisées de manière efficace et fiable dans différentes zones de l'appareil gastro-intestinal (Œsophage, estomac, gros intestin ou côlon et une partie de l'intestin grêle) avec des capacités diagnostiques, thérapeutiques et chirurgicales [13]. Cependant, la rigidité de l'instrument, due à ses dimensions (diamètre de 11 à 13 mm pour un coloscope standard), et l'intégration de mécanismes d'actionnement traversant l'instrument entraînent une accessibilité limitée, rendant les procédures endoscopiques significativement traumatisantes et mal tolérées par les patients [14]. La douleur ou les problèmes de sédation rendent les patients peu enclins à subir une endoscopie, limitant considérablement l'omniprésence des campagnes de dépistage de masse. Un inconvénient médical important de l'endoscopie flexible est que certaines zones du tractus gastro-intestinal ne peuvent pas être atteintes, comme la majeure partie de l'intestin grêle [15].

En 1981, un ingénieur en électro optique, Gavriel Iddan en collaboration avec le gastroentérologue/inventeur Paul Swain a conçu l'idée d'une capsule de caméra sans fil pour l'imagerie de l'ensemble du tractus GI. Les limitations des technologies disponibles ont empêché le développement d'une capsule de caméra pouvant être avalée. En 1990, Iddan a utilisé la technologie CCD (Charge Coupled Device) dans les capteurs d'images remplaçant

Chapitre 1 : La capsule endoscopique

ainsi la fibre optique. Cette technologie CCD permet à Iddan d'utiliser seulement une caméra en la laissant se déplacer, naturellement, dans le tractus gastro-intestinal. Ensuite, l'idée d'équiper la caméra par un transmetteur sans-fil a été rapidement développée. Puis une question importante relative au temps nécessaire à la caméra pour traverser le petit intestin s'est posée ? La batterie miniature qui existait alors ne permettait que de 10 à 15 minutes de travail, sans tenir compte de l'énergie nécessaire à la transmission et à l'éclairage.

Ces défis énormes ont presque empêché Iddan de continuer le projet, et il a décidé de s'attaquer à ces obstacles un par un. Il a commencé par fabriquer son premier prototype en utilisant une fenêtre optique "axicon", en supposant que cette forme sera continuellement nettoyée pendant le déplacement. Il a ajouté une caméra miniature CCD et une source lumineuse qui a montré expérimentalement que cela produisait des images raisonnables. Certains de ces premiers modèles sont présentés à la figure 1.1. En 1993, avec le développement de nouveaux capteurs CCD, qui consomment moins d'énergie, une nouvelle conception a été proposée qui divise l'appareil en 3 parties : un capteur CCD, un transmetteur-enregistreur et un poste de travail. Cette conception permet d'analyser les images hors ligne et non en temps réel. Un réseau de capteurs est ajouté à la conception pour la localisation de l'appareil. En 1997, Iddan change le capteur d'image CCD par un capteur CMOS qui consomme moins d'énergie et augmente le temps de vie de la capsule.



Figure 1. 1 Le premier prototype de la capsule. [14]

Chapitre 1 : La capsule endoscopique

Dans La même année, Iddan et Scapa ont travaillé avec un autre groupe de Paul Swain, et ses collègues. Le groupe de Paul Swain a travaillé sur la transmission sans fil et sur la capsule pH. En 1994, Paul Swain a soulevé la possibilité d'un endoscope sans fil dans une conférence donnée lors du congrès mondial de gastro-entérologie. Plus tard, en 1996, le groupe Swain a commencé à tester la transmission d'images vidéo à travers le corps humain. En 1999, les deux groupes ont procédé aux tests expérimentaux dans la clinique de Scapa avec une capsule de taille 33 x11 mm. Les résultats obtenus ont montré que le temps de vie ne dépasse pas les 2 h. Plus tard, et avec un très bon transmetteur la capsule a transmis des images de bonne qualité avec un temps de vie allant jusqu'à 6 h.

En 2000, grâce aux brevets de G. Iddan, la société Given Imaging Ltd. a introduit « l'endoscopie par capsule sans fil » (WCE), qui implique l'ingestion d'une caméra miniaturisée de la taille d'une pilule qui navigue passivement dans le tractus gastro-intestinal au moyen des contractions péristaltiques naturelles et prend des images de la paroi environnante [16]. WCE permet l'inspection du système digestif sans douleur ni besoin de sédation, évitant les risques et la mauvaise tolérance de l'endoscopie conventionnelle. De cette manière, la WCE présente l'avantage supplémentaire potentiel d'encourager les patients à se soumettre à des examens médicaux du tractus GI augmentant éventuellement le succès des campagnes de dépistage de masse [17].

Dans les années qui ont suivi, Given Imaging a proposé des prototypes d'une deuxième et troisième génération pour les capsules du petit intestin (PillCam SB2, PillCam SB3), Ces capsules présentent quelques améliorations technologiques. Telle que, la source de lumière, le temps de vie de batterie, l'angle de vue et la fréquence des images (image par seconde). La société a proposé aussi deux générations d'une capsule œsophagienne à deux faces (PillCam Eso2), deux générations de la capsule colique (PillCam Colón 2) avec une fréquence d'image adaptée et une très large gamme de vue. Ensuite, le progrès technique a conduit à l'introduction de certaines versions (2ème et 3ème génération) de CE pour l'intestin grêle, l'œsophage et le côlon. Ensuite, plusieurs autres prototypes ont été fabriqués par des sociétés japonaises, chinoises, coréennes et américaines. Telles que Sayaka, Olympus, Philips.

1.3 Système de capsule endoscopique

Le système WCE typique, montré dans la figure 1.2, se compose de trois composants principaux

- (a) une capsule à usage unique.
- (b) un enregistreur de données.

(c) un poste de travail.



Figure 1. 2 Système de capsule endoscopique : capsule, dispositif d'enregistrement et post de travail. [18]

1.3.1 La capsule

C'est un dispositif jetable qui se présente sous la forme d'une grande gélule, elle est construite d'un matériau biocompatible résistant à l'action des enzymes digestifs. Elle contient plusieurs composants, y compris un capteur d'image, une source de lumière, un système de communication (transmetteur), et une source d'énergie (batterie) (figure 1.3).

1.3.1.1 Capteur d'image

Deux types de capteurs d'images sont utilisés dans les capsules endoscopiques à savoir : le CCD et le CMOS. Les principales caractéristiques de ces technologies sont : Le capteur d'image CCD (Charge Coupled Device) transfère une charge presque complète de pixels, créé par une conversion photoélectrique signifiant que le bruit est rare, mais un différentiel de haute tension

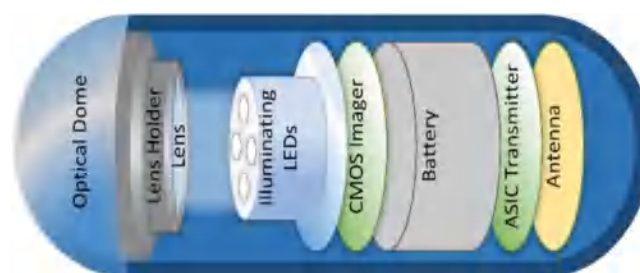


Figure 1. 3 Composants de la capsule endoscopique. [19]

est nécessaire pour améliorer l'efficacité du transfert augmentant aussi la consommation d'énergie, et c'est pour cette raison que, actuellement, la plupart des capsules endoscopiques n'utilisent pas ce capteur. Le capteur d'image CMOS (CoMplementary Oxide Semiconductor) convertit la charge en une tension pour chaque pixel. Il utilise une matrice de pixels pour convertir la lumière en des signaux électriques. Ce signal est faible et doit être amplifié à un niveau utilisable, où chaque pixel possède son propre circuit amplificateur. Le capteur CMOS contient un nombre de connecteurs plus faibles, une fiabilité accrue, une consommation d'énergie réduite et il est plus compacte. Dans les capsules endoscopiques, en avant du capteur d'image (CMOS ou CCD), on trouve une lentille asphérique qui est focalisée manuellement sur le dôme optique, de sorte que l'intestin se trouve contre le dôme optique. Le choix des lentilles permet un examen plus efficace de l'intestin. Les améliorations de la qualité d'image peuvent être obtenues grâce à une attention particulière au positionnement de la lentille par rapport au capteur. La position de la pupille d'entrée de la lentille optique doit coïncider avec le centre de la courbure de la fenêtre du dôme optique pour empêcher toute lumière réfléchie d'entrer dans le capteur d'image [20].

1.3.1.2 Système d'illumination

À l'intérieur de la capsule, un système d'éclairage miniaturisé de type LED (white light emitting diode) est utilisé pour éclairer le tractus digestif. En général, la répartition de la lumière spatiale n'est pas uniforme sur le plan de l'objet devant la capsule, ce qui entraîne une réduction de la qualité d'image. Dans certaines régions du plan de l'objet, la luminance est très petite et dans d'autres régions, elle pourrait être très grande. Afin de résoudre ce problème, plusieurs LED ont été disposées dans différentes positions et angles, pour améliorer l'uniformité lumineuse de la capsule. En raison des contraintes d'espace dans la capsule, les LED doivent être petites et doivent fournir une intensité lumineuse suffisante pour obtenir des images de bonnes qualités. Leurs intensités lumineuses devraient également pouvoir fournir suffisamment de lumière pour le pire scénario où la lumière nécessaire est très grande. C'est le cas dans le cæcum et dans le côlon ascendant. La puissance requise par les LED dépend du nombre et du type de LED utilisé.

1.3.1.3 Système de transmission

Toutes les images capturées sont ensuite traitées par l'une des trois manières suivantes : transmission par radiofréquence, communication avec le corps humain ou stockage intégré des données [21].

- Transmission par radiofréquence

Chapitre 1 : La capsule endoscopique

Tous les modèles PillCam, l'EndoCapsule et la capsule OMOM utilisent la même technique de transmission. Les images compressées sont envoyées par radiofréquence à travers une antenne intégrée, à un réseau de capteurs. Le patient porte ce réseau de capteurs directement sur la peau ou intégré dans une ceinture ou un gilet de capteurs. Le réseau est connecté à un dispositif de stockage portable (Figure 1.4). Avec les dernières générations de WCE, chaque fabricant utilisant cette méthode offre la possibilité d'une visualisation en temps réel grâce à un écran intégré ou connecté. Une fois l'étude terminée, le dispositif de stockage est connecté via une station d'accueil à l'ordinateur d'analyse pour le transfert et le traitement des données en vue de leur analyse. Comme ces types de capsules ne stockent pas de données par elles-mêmes, elles peuvent être jetées après excrétion.



Figure 1. 4 Schéma de la technique de transmission radiofréquence. [21]

- Communication avec le corps humain

La capsule MiroCam, transmet ses données via un boîtier bipolaire extérieur spécial aux électrodes du capteur sur la peau, en utilisant le corps humain comme conducteur (communication avec le corps humain). Cette méthode utilise moins d'énergie que la transmission par radiofréquence [22]. Semblable à la transmission par radiofréquence, le patient



Figure 1. 5 Schéma de la technique de transmission via le corps humain. [21]

Chapitre 1 : La capsule endoscopique

porte un dispositif de stockage portable, qui est ensuite connecté à l'ordinateur d'analyse (Figure 1.5). À la fin de l'analyse, la capsule n'a pas besoin d'être récupérée par le médecin et peut être jetée.

- Stockage intégré des données

La Capsocam stocke les images capturées dans une puce de stockage Flash EPROM intégrée, évitant ainsi le besoin d'un émetteur intégré et d'un récepteur externe. Une fois l'analyse terminée, le patient rend la capsule au personnel médical pour le téléchargement des données via un port USB dans une station d'accueil (Figure 1.6). Avec ce système, il n'y a aucune possibilité de visualisation en temps réel.



Figure 1. 6 Schéma de la technique de transmission via stockage de données intégrées. [21]

1.3.1.4 L'alimentation

Les CE commercialisées utilisent de petites batteries d'oxyde d'argent (généralement, il y en a deux), qui représentent plus de la moitié du poids de la capsule. Dans toutes les applications de la CE, il est impératif de tenir compte des compromis de la durée de vie de la batterie et de la toxicité de la technologie de la batterie. Un choix judicieux de composants et une conception électrique efficace sont importants pour améliorer la performance des capsules. Il existe seulement certains types de piles, disponibles dans le commerce, qui peuvent s'insérer dans l'espace limité de la capsule. Donc, l'alimentation électrique doit être conçue de manière très efficace. La batterie fournit un courant continu suffisant pour permettre le fonctionnement fiable de tous les éléments électroniques à l'intérieur de la capsule. En plus de la tension de fonctionnement, le courant de décharge continu, le poids faible et les dimensions petites sont aussi des facteurs décisifs lors du choix d'une alimentation électrique efficace. Donc, un paramètre important à considérer est la résistance interne de la batterie, car elle doit fournir suffisamment de courant de décharge continue de 28-30mA pour les composants électroniques et une impulsion de 400ms/80mA pour déclencher le micro-chauffeur simultanément. Plus la

résistance interne est faible, moins la batterie est limitée dans sa capacité à fournir les pics de puissance nécessaire. En outre, plus la batterie est petite (c'est-à-dire son diamètre), plus la résistance interne est élevée.

1.3.2 Le dispositif externe d'enregistrement

C'est une unité de réception enregistrement externe qui reçoit les données transmises par la capsule (figure 1.7). Ce module est de la taille d'un baladeur, alimenté par batteries et porté à la ceinture par le patient durant l'examen. Le module d'enregistrement comprend un récepteur, un processeur et un disque dur pour enregistrer les données. Tous ces éléments sont renfermés dans un boîtier en plastique. La ceinture ambulatoire permet au patient de poursuivre ses activités quotidiennes normales.



Figure 1. 7 Enregistreur de données. [21]

1.3.3 La station de travail

C'est un ordinateur conçu pour le traitement des données et la présentation des images capturées par la capsule, où un logiciel est installé pour analyser les données enregistrées. Ce logiciel est fourni par le vendeur correspondant. Différents vendeurs de WCE ont développé divers logiciels avec différentes fonctionnalités, par exemple, le logiciel RAPID Reader est développé et fourni par PillCam. L'affichage permet aux médecins de suivre le chemin parcouru par la capsule, de visualiser les lésions et d'enregistrer des images importantes pour le diagnostic.

1.4 Différent types des capsules disponibles dans le commerce

Un certain nombre de CE sont maintenant disponibles dans le commerce. La plupart de ces capsules sont déjà de la deuxième ou la troisième génération, avec un développement continu. Ils diffèrent, par exemple, par le nombre de caméras, le taux d'imagerie, les méthodes de stockage ou de transmission des données, la puissance de la batterie et le domaine d'utilisation. Les principales caractéristiques des diverses capsules disponibles dans le commerce sont résumées dans le tableau 1.1 [23]. Il ressort clairement du tableau que les capsules à base de capteurs d'images sont plus abondantes sur le marché. D'autres capsules utilisent différents capteurs, tels qu'un capteur de température, un capteur de pH ou un capteur de pression, pour mesurer différents paramètres physiologiques.

Tableau 1. 1 Principales caractéristiques des capsules WCE disponibles dans le commerce. [23]

WCE Capsules Based on Image Sensor											
Device	Company	Target Region	Size (mm)	Mass (g)	IL (LEDs)	BL (h)	FoV (°)	FR	Image Sensor	Res.	TM
PillCam SB	Given imaging/ medtronic	Small bowel	26 × 11	< 4	6	8	140	2	1 CMOS	256 × 256	RF
PillCam SB2		Small bowel	26 × 11	2.9	4	8	156	2 or 4	1 CMOS	256 × 256	RF
PillCam SB3		Small bowel	26 × 11	3	4	12	156	2 or 2–6	1 CMOS	256 × 256	RF
PillCam ESO		Esophagus	26 × 11	2.9	6	0.5	172	35	2 CMOS	256 × 256	RF
PillCam ESO2		Esophagus	26 × 11	2.9	2 × 4	0.5	169	18	2 CMOS	256 × 256	RF
PillCam Colon		Colon	31 × 11	2.9	2 × 6	10	172	4	2 CMOS	256 × 256	RF
PillCam Colon2		Colon	31 × 11	2.9 ± 1	2 × 4	10	360	4–35	2 CMOS	256 × 256	RF
Miro Cam	Intromedic	Small bowel	24.5 × 10.8	3.25	6	12	170	3	1 CMOS	320 × 320	HBC
Endo Capsule	Olympus	Small bowel	26 × 11	3.3	6	12	160	2	1 CCD	1920 × 1080	RF
OMOM	JinShan	Small bowel	27.9 × 13	≤ 6	4	8 ± 1	140 ± 10	0.5 or 2	1 CMOS	640 × 480	RF
Capso Cam	Capso Vision	Small bowel	31 × 11	4	16	~15	360	3 or 5 p.c.	4 CMOS	1920 × 1080	USB p

WCE Capsule Based on Other Sensors							
Device	Company	Target Region	Size (mm)	Mass (g)	BL (h)	Sensor(s)	Measuring Parameter
SmartPill	Given Imaging/ Medtronic	GI tract	11.7 × 26.8	4.5	120	Pressure (0–350 mmHg), pH (0.05–9) and temperature (25–49°C)	Pressure, pH and temperature
CorTemp	HQ Inc.	GI tract	10.37 × 22.32	2.75	168–240	Temperature sensor (30–45°C)	Temperature
VitalSense	Philips	GI tract	8.7 × 23	1.6	240	Jonah™ core temperature sensors (25°C–60°C)	Core body temperature, dermal temperature, heart rate, respiration rate
e-Celsius	BMedical Pty Ltd	GI tract	8.9 × 17.7	1.7	480	Temperature sensor (25°C–45°C)	Core body temperature
Check-Cap	Check Cap Ltd	Colo-rectal	34 × 11.5	–	–	X-ray based 3D imaging	3-Dimensional lining of colon

FoV: field of view; FR: frame rate; Res.: resolution; TM: transmission module; RT: real time; p.c: per camera; RF: radiofrequency; HBC: human body communication; IL: illuminant; BL: battery life.

1.5 Les applications

L'utilisation de la WCE est devenue la première procédure du diagnostic dans certaines pathologies du système GI à commencer par l'œsophage jusqu'au colon. La figure 1.8 montre des exemples visuels des images extraites d'une séquence vidéo obtenue par un examen, en utilisant une WCE. Mais le premier but des capsules endoscopiques est l'intestin grêle, qui reste comme une boîte noire dans le système GI. Plus souvent, la capsule dans l'intestin grêle est utilisée chez les patients présentant des saignements présumés ou pour identifier les signes de la maladie de Crohn active [24]. La WCE peut être utilisée également pour les patients atteints de la maladie cœliaque, de syndromes de polypose et d'autres troubles de l'intestin grêle. Il y a aussi d'autres capsules spécifiques comme les capsules

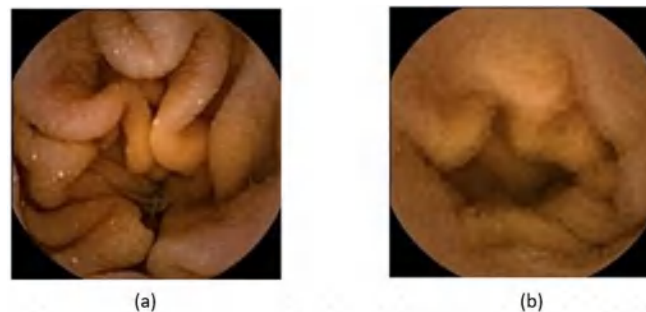


Figure 1. 8 Images du (a) estomac et (b) intestin grêle capturée par une capsule.

œsophagiennes ou coliques. Les capsules œsophagiennes peuvent être utilisées pour diagnostiquer l'œsophagite, l'œsophage de Barrett et les varices, mais on ne peut pas compter sur elles dans l'identification de la maladie gastro-duodénale [24]. Les capsules coliques peuvent être utilisées pour diagnostiquer le cancer du côlon et aussi des maladies traditionnelles du côlon. Donc, en quelques années seulement, la CE a évolué très rapidement, et elle est devenue un outil inestimable pour l'examen de presque l'ensemble du tractus GI, comme l'intestin grêle, l'œsophage et colon, et les résultats des diagnostics ont dépassé de loin les anticipations [24].

1.5.1 La Capsule de l'intestin grêle

La figure 1.9 (a) montre une capsule sans fil de l'intestin grêle (Pillcam SB2), qui est habituellement utilisée dans un cadre ambulatoire. Certains centres médicaux utilisent un régime de liquide clair pendant 24 heures avant l'étude pour améliorer la visualisation dans l'intestin grêle. Plusieurs études suggèrent que l'utilisation d'une préparation intestinale totale ou partielle la nuit avant l'étude donne une meilleure visualisation de l'intestin grêle [25], [26]. Au moment de la procédure, le système de détection est appliqué sur la paroi abdominale et

relié à l'enregistreur de données porté par le patient. La capsule est activée dès qu'elle est retirée de son support magnétique. Après la prise de la capsule, les patients sont invités à tenir compte des symptômes et à surveiller les lumières de l'enregistreur de données afin de confirmer que le signal est reçu. Les patients sont encouragés à éviter les exercices ou les activités susceptibles de provoquer le décollement des capteurs. Un régime de liquides transparents est autorisé après 2 heures et un repas léger après 4 heures. Le système d'enregistrement des données utilisables peut être déconnecté du patient après la fin de la durée de vie de la batterie. La capsule est jetable et conçue pour être excrétée. L'enregistreur de données est ensuite connecté à une station de travail pour le transfert des images acquises.

1.5.2 La Capsule œsophagienne

La capsule PillCam ESO2, montrée dans la figure 1.9 (b), est un système de la CE actuellement disponible pour les applications œsophagiennes. Les dimensions de la capsule, la fréquence de transmission, le champ de vision et la taille minimale de l'objet sont similaires au PillCam SB2. Cependant, la durée de vie de la batterie de la capsule est de seulement 20 minutes (contre 8-12 heures pour les capsules de l'intestin grêle), les caméras sont situées aux deux extrémités de la capsule et la capsule prend 18 images par seconde (contre 2-3 images par seconde pour la capsule d'intestin grêle). Au moment de l'examen, le patient devrait jeûner pendant 2 heures. Les patients sont équipés de trois capteurs thoraciques qui sont connectés à l'enregistreur de données. Comme dans le protocole traditionnel, le patient consomme 100 ml d'eau en position debout, puis la capsule est activée en position couchée avec 20 gouttes de Siméthicone qui peut être administrée à l'aide d'une seringue. Un protocole de 5 minutes en gestion est recommandé, comprenant un enregistrement de 2 minutes avec le patient en position couchée, puis l'angle du lit est augmenté à 30 ° pendant 2 minutes, et à 60 ° pendant une minute, suivi d'une position droite jusqu'à 15 minutes pour maximiser le temps de la capsule à capturer les images, alors qu'elle traverse l'œsophage [27]. Comme le PillCam SB2, les images sont transmises à l'enregistreur de données et ensuite au poste de travail pour examen et interprétation via le logiciel propriétaire.

1.5.3 La Capsule colique

La deuxième version de la CE pour le colon (Pillcam Colon 2) comme montrée dans la figure 1.9 (c) a été fabriquée pour augmenter la sensibilité pour la détection des polypes colorectaux. Pillcam Colon 2 permet d'acquérir des images jusqu'à un maximum de 35 images par seconde et offre une couverture de 172 ° du colon. La société Given Imaging conseille d'utiliser PillCam Colon 2 comme un outil fiable et sûr pour la visualisation du colon chez les patients. En bref,

Chapitre 1 : La capsule endoscopique

une méta-analyse de 8 études de la capsule colique de première génération (PillCam Colon, Given Imaging) a révélé une sensibilité par patient pour les polypes de toutes tailles et les résultats sont significatifs où 71 % à 68 % par rapport à la coloscopie conventionnelle [28]. En outre, les essais polyvalents de la capsule colique de la deuxième génération ont montré une sensibilité pour les polypes supérieurs ou égaux à 10 mm de 88 % à 89 %, par rapport à la coloscopie conventionnelle [29], [30].

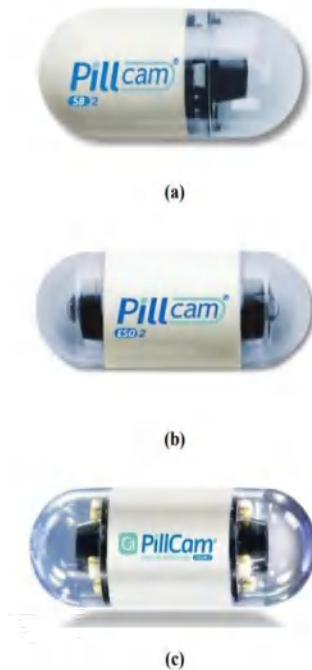


Figure 1. 9 (a) capsule l'intestin grêle, (b) capsule œsophage, (c) capsule colique

1.6 Les limites et les perspectives

Malgré le développement rapide de la technologie de la CE, il reste encore de nombreux problèmes en suspens qui doivent être améliorés, à savoir l'incapacité de contrôler la locomotion, une qualité d'image faible, aucune capacité vidéo en temps réel à haute fréquence d'images, des ressources d'alimentation limitées, une télémétrie à faible débit de données, et des capacités de traitement limitées. Ces questions ouvertes rendent les techniques endoscopiques traditionnelles toujours supérieures en termes de capacités diagnostiques. Les principaux défis de la conception du WCE sont d'acquérir et de transmettre un nombre acceptable d'images avec la meilleure qualité possible, en utilisant le moins de matériel et de batterie possible. Comme la transmission sans fil consomme la majeure partie de l'énergie, il est possible de réduire la consommation de cette énergie en diminuant le nombre de bits transmis [6]. Cette ressource énergétique limitée a imposé aux systèmes WCE d'utiliser de petites tailles comme 256x256 pixels et de très petites fréquences d'images de seulement 2 à 4 images par seconde, afin de

pouvoir travailler longtemps. L'augmentation de ces valeurs ne nous permettra pas d'atteindre les régions intestinales éloignées. Par exemple, les capsules de l'œsophage peuvent fonctionner avec 18 (ou même 35) images par seconde, mais pendant seulement 20 minutes [31]. Même avec de petites images, il est nécessaire d'ajouter un algorithme de compression efficace. La compression permet de supprimer toute information redondante, tout en conservant les mêmes images. Cependant, même cette opération doit être effectuée avec une faible complexité et une faible consommation.

En outre, la position et la taille des lésions ne sont pas connues et la capsule se déplace dans une seule direction, avec des mouvements irréguliers, ce qui signifie qu'il n'y a pas de possibilité de réexaminer une région si elle est manquée. Ainsi, la WCE est obligé de travailler avec une fréquence d'images régulière et d'utiliser de petites images qui lui permettent de couvrir uniformément toutes les régions du tractus GI, et de travailler sur la plus grande distance possible. Il en résulte deux problèmes. Premièrement, une compression efficace est nécessaire pour toutes les images. Cela permettra au WCE de travailler plus longtemps ou d'augmenter la taille de l'image ou la fréquence d'images, tout en consommant la même énergie. Deuxièmement, même après avoir décodé toutes les images, leur lecture prendra beaucoup de temps pour finalement déterminer que seul un petit nombre d'images est vraiment important pour le diagnostic [32].

1.7 Techniques de compression d'images WCE

En fonction de la qualité de l'image reconstruite, les algorithmes de compression d'images peuvent être divisés en deux catégories : les algorithmes avec perte et les algorithmes sans perte ou presque sans perte. Dans la compression sans perte, il est possible de reconstruire l'image exactement comme l'image originale, tandis que dans la compression presque sans perte, un certain degré de distorsion est accepté à condition qu'il n'y ait pas de perte d'informations précieuses [33] [34]. Le principe de la compression sans perte est la réduction de la redondance dans l'image en exploitant la corrélation spatiale et le codage entropique approprié [35]. Mais selon l'observation de Clunie dans [36] sur un grand nombre d'algorithmes de compression, le meilleur taux de compression reste inférieur à 4 pour les images en niveaux de gris alors qu'il est inférieur à 12 pour les images en couleurs [35]. D'autre part, la compression avec perte peut réduire le contenu informatif de l'image [37] sans pour autant affecter directement la résolution [35].

1.7.1 Techniques de compression d'images sans perte

Les techniques de compression d'images sans perte dépendent principalement de la décorrélation et du codage entropique. La redondance spatiale entre les pixels peut être supprimée en utilisant des techniques de prédiction. Dans le codage prédictif, les pixels sont décorrélés par un bon prédicteur. Il s'agit principalement de la modulation delta, la modulation différentielle par impulsions codées (DPCM) ou la compression JPEG. Cependant, la plupart des techniques de compression endoscopique prédictive prennent en compte la modulation DPCM et l'estimation simple. Toutes ces approches sont principalement sans perte ou presque sans perte, à une ou deux exceptions près.

1.7.1.1 Compression d'images basée sur JPEG-LS

Tous les schémas de compression basés sur JPEG-LS pour WCE ont deux composantes principales : Le moteur JPEG-LS et une étape de prétraitement, qui consiste en un filtre passe-bas (LPF) et une transformation. Les algorithmes de compression basés sur JPEG-LS offrent le meilleur taux de compression sans perte et la plus grande vitesse de compression pour les images médicales qui ont un fond similaire [38].

JPEG-LS, une méthode de compression sans perte basée sur l'algorithme LOCO-I [39], consiste en deux phases indépendantes et distinctes appelées modélisation et codage [40]. Dans JPEG-LS, la redondance spatiale de l'image est réduite par un modèle de prédiction appelé Median Edge Detector (MED), où un algorithme simple de détection des bords prédit la valeur du pixel suivant. Un choix approprié d'un codeur parmi une collection de codeurs de Golomb réduit la redondance du codage. Le JPEG-LS améliore encore les performances en utilisant le codage run length coding pour les régions plates. Une description détaillée de l'algorithme JPEG-LS peut être trouvée dans [40].

1.7.1.2 Techniques basées sur le DPCM

Le DPCM [41], [42] est une procédure largement connue de quantification différentielle des signaux de communication, Leur principe repose sur la prédiction de la valeur d'un pixel en fonction d'un certain nombre de pixels voisins. On peut alors supprimer l'information redondante du pixel et ne coder que l'erreur de prédiction, déterminée par la différence entre la valeur du pixel à coder et sa prédiction. Une technique basée sur la prédiction est préférable pour l'endoscopie sans fil [43]–[45]. En raison de l'absence d'objet de couleur bleutée dans le tractus gastro-intestinal, la variation de la composante U dans l'image endoscopique est inférieure à celle de la composante V. Ce phénomène rend le sous-échantillonnage plus favorable lorsqu'on utilise l'espace couleur YUV. Le sous-échantillonnage peut ne pas aboutir

à une compression sans perte pure, mais il est possible d'obtenir un mode quasi sans perte. Par exemple, dans [43], le sous-échantillonnage est utilisé dans l'espace couleur YUV, où les composantes Y, U et V sont sous-échantillonnées dans le rapport 8:1:2. La compression qui en résulte peut fonctionner en mode quasi sans perte.

Les algorithmes basés sur la DPCM sont prometteurs car ils peuvent fonctionner à la fois en mode sans perte et en mode avec perte acceptable. En mode avec perte, il fournit l'un des taux de compression les plus élevés parmi toutes les techniques basées sur la prédiction. En mode sans perte ou presque sans perte, le taux de compression est meilleur que celui des algorithmes basés sur JPEG-LS. Cependant, un choix judicieux de l'espace couleur, du sous-échantillonnage et de l'encodage peut fournir de meilleurs résultats, ce qui est prometteur pour obtenir un taux de compression élevé permettant des fréquences d'images élevées. En particulier, un meilleur filtre de prédiction doit être réalisé au lieu de soustraire le pixel précédent. De plus, un DPCM comme [45] comprenant une transformation de l'espace couleur et un sous-échantillonnage pourrait donner plus de compression.

1.7.2 Techniques de compression d'images avec perte

1.7.2.1 Compression d'images basées sur la DCT

La transformée en cosinus discrète (DCT) [46] est principalement employée comme schéma de compression avec perte. Elle est largement utilisée dans de nombreux codecs de compression d'images et de vidéos [47]. Une DCT typique prend un bloc de 8x8 en niveaux de gris et applique la compression. Le même processus est répété pour chaque canal de couleur. La DCT est employée dans la norme JPEG [48] et souvent utilisée pour la compression des images WCE [49]–[54]. Un algorithme typique comprend généralement le démosaïquage de l'image brute (facultatif), la transformation de l'espace couleur, la transformation directe bidimensionnelle et la quantification suivie d'un codage entropique. Dans toutes ces mises en œuvre, les capteurs d'image accèdent aux pixels de l'image par blocs de 4x4 ou 8x8. La DCT est bien mieux établie que les transformées en ondelettes, mais elle nécessite toujours des calculs complexes qui peuvent être trop importants pour les applications WCE [55], [56]. Par conséquent, une version entière simplifiée de la DCT (également connue sous le nom de iDCT) devient populaire [56]–[58]. La quantification peut être redondante dans cette version de la DCT car cette transformation représente déjà une perte. L'iDCT a déjà prouvé son efficacité dans l'application WCE [57] au prix d'un stockage temporaire car les images sont divisées en blocs 8x8. L'un des avantages de l'iDCT est que la partie quantification peut être partiellement éliminée car elle est intégrée dans la transformation DCT. Cependant, à faible débit binaire, cet algorithme produit

des artefacts significatifs dans les trames reconstruites, ce qui entraîne une mauvaise qualité d'image.

1.8 Conclusion

WCE joue un rôle important dans le diagnostic des maladies du tractus GI, grâce à une procédure indolore. C'est une procédure sûre et largement utilisée pour le dépistage de l'intestin grêle, qu'un endoscope filaire ne peut pas atteindre. Dans ce chapitre, on a fait une description générale des capsules, leur composition, ainsi que leur différent type et ces applications et aussi leurs limites : comme le manque de contrôle, de l'éclairage, de la résolution, de ressources d'alimentation limitées, de la prise de décision en temps réel, etc. Ensuite, nous avons montré l'importance de la compression efficace dans les capsules endoscopiques comme une solution pour réduire la consommation d'énergie. Pour cela, on a présenté les différentes techniques de compression qui peuvent utilisés pour les images de CE. En conclusion, un système intelligent de gestion des données est l'un des objectifs importants qui doivent être abordés dans le développement de futurs systèmes d'endoscopie par capsule

Chapitre 2

**Notions de base sur l'apprentissage
profond**

2.1 Introduction

L'apprentissage profond (Deep Learning) fait partie des méthodes d'apprentissage automatique (machine learning), qui est un sous ensemble de l'intelligence artificielle. Dans ce chapitre nous allons expliquer ce qu'est l'apprentissage automatique et ce qu'il permet de faire. Nous présentons ensuite le modèle du réseau de neurones qui est à la base de l'apprentissage profond. Nous montrerons ensuite les notions des réseaux de neurones convolutionnels et leur architecture. Le but de ce chapitre est de comprendre le fonctionnement de l'apprentissage profond afin de pouvoir développer des algorithmes robustes et efficaces pour traiter des données. Dans le cadre de notre travail, les données sont des images de WCE.

2.2 L'apprentissage automatique

L'apprentissage est un concept pour définir l'acquisition de connaissance et réutiliser ces nouvelles connaissances. Pour nous les humains, notre apprentissage se fait tout au long de notre vie. Nous apprenons à partir de la perception de l'environnement avec les cinq sens, les expériences de la vie, de répétition des événements avec la mémoire, et de notre jugement (libre arbitre ou intelligence). Pour les machines, comme elles ne sont pas dotées de sens ni de jugement dynamique, elles obéissent aux instructions dans un programme avec des données d'entrée et une donnée sortie ou réponse du programme.

Arthur Samuel a décrit l'apprentissage automatique comme un domaine qui permet à une machine d'avoir des connaissances sans être formellement programmé ou sans intervention humaine [59].

Tom Mitchell quant à lui a défini l'apprentissage comme un programme informatique qui apprend "à partir d'une donnée d'expérience E une tâche spécifique T avec une mesure de performance P " [60].

2.2.1 Paradigmes d'apprentissage

Les algorithmes d'apprentissage automatique apprennent selon des objectifs particuliers qui peuvent être divisés en trois principales catégories : l'apprentissage supervisé, l'apprentissage non-supervisé et l'apprentissage par renforcement. Il existe également un paradigme d'apprentissage semi-supervisé qui mélange les apprentissages supervisés et non-supervisé.

2.2.1.1 Apprentissage supervisé

L'apprentissage supervisé concerne les modèles prédictifs dont le rôle est d'apprendre à prédire les cibles pour des problèmes de régression ou de classification.

Chapitre 2 : Notions de base sur l'apprentissage profond

- Dans le cas d'un problème de régression :

Les cibles prennent des valeurs réelles, ainsi $y \in \mathbb{R}$. Par exemple, une tâche de prédiction pourrait être de tenter de prédire la valeur boursière des actions d'une compagnie en fonction de son historique et de textes de nouvelles dans les médias. La figure 2.1 montre un exemple de régression linéaire, la ligne rouge représente l'hypothèse et les points bleus sont des ensembles de données.

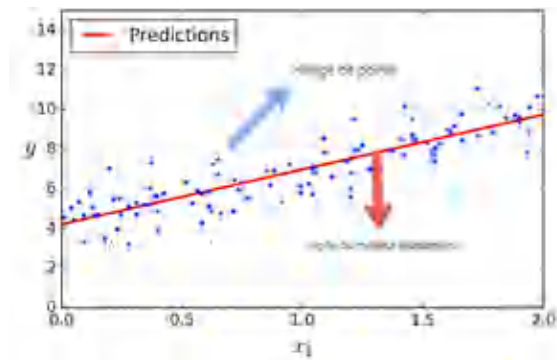


Figure 2.1 Exemple d'une régression linéaire. [61]

- Dans le cas d'un problème de classification,

Les cibles prennent des valeurs entières, plus communément appelées classes ou catégories. Si le problème de classification est binaire, alors $y \in \{0, 1\}$. Si le problème de classification est multiclasse, alors $y \in \{1, \dots, L\}$ où L correspond au nombre de classes. Par exemple, une tâche de classification pourrait être la prédiction du stade de la maladie d'un patient à partir de plusieurs métriques issues de tests cliniques. Un exemple de classification binaire (c'est-à-dire, à deux classes) est présenté dans la Figure 2.2. Les points bleus et les points rouges représentent des données appartenant à deux classes différentes. Le but d'un classificateur est d'apprendre la frontière entre ces deux classes.

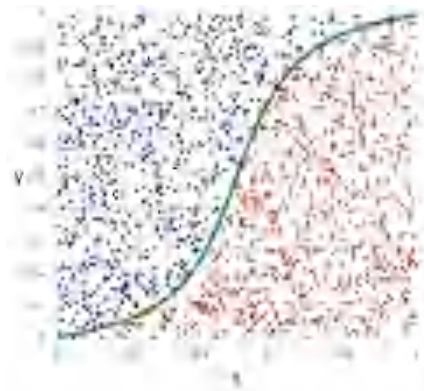


Figure 2.2 Exemple d'une classification avec des données d'entrées x . [62]

2.2.1.2 Apprentissage non-supervisé

L'apprentissage non-supervisé concerne les modèles descriptifs où il n'existe pas de cible explicite. L'objectif est d'entraîner un modèle qui soit capable de déceler par lui-même des facteurs qui expliqueraient au mieux les données. En d'autres mots, l'idée est d'apprendre une distribution de probabilité qui modélise le jeu de données. Plusieurs applications en découlent comme l'estimation de densité, la génération de nouveaux exemples, le partitionnement de données en sous-groupes, ou encore la réduction de dimensionnalité.

2.2.1.3 Apprentissage semi-supervisé

L'apprentissage semi-supervisé concerne le cas où le jeu de données D est partiellement étiqueté. L'objectif est d'entraîner un modèle qui soit capable de tirer parti à la fois des cibles présentes mais aussi des données non étiquetées.

2.2.1.4 Apprentissage par renforcement

L'apprentissage par renforcement fait référence à une classe de problèmes d'apprentissage automatique, qui consiste à apprendre à partir d'expériences successives. Ce qu'il convient de faire de façon à trouver les meilleures solutions. Autrement dit, les machines intelligentes essaient plusieurs situations afin de pouvoir déterminer les actions les plus avantageuses et ne se contentent pas de recevoir des instructions sur les actions à appliquer. C'est ce qui distingue cette méthode des autres techniques d'apprentissage. L'apprentissage par renforcement est un modèle d'apprentissage comportemental. L'algorithme dans ce cas reçoit les informations en analysant des données, pour pouvoir orienter l'utilisateur vers les meilleurs résultats. Dans ce type d'apprentissage, le système n'est pas entraîné à partir d'un ensemble de données mais il apprend par essais et erreurs, ce qui le différencie des autres types d'apprentissage supervisé.

2.3 Les réseaux de neurones artificiels

Les réseaux de neurones sont des modèles d'apprentissage automatique capables de représenter une relation entre des données d'un espace d'entrée X et un espace de sortie Y . Ils sont utilisés dans de nombreux domaines, comme la vision assistée par ordinateur [63]–[65], le traitement du langage naturel [66], l'analyse audio [67], [68], mais également pour développer des IA capables de jouer à des jeux [69] ou utilisées comme assistant personnel (tel que Alexa d'Amazon, Siri d'Apple, Cortana de Microsoft ou l'Assistant de Google).

L'origine des réseaux de neurones vient de l'essai de modélisation mathématique du cerveau humain. Les premiers travaux datent de 1943 et sont l'œuvre de W.M. Culloch et W. Pitts. Ils supposent que l'impulsion nerveuse est le résultat d'un calcul simple effectué par chaque

neurone et que la pensée née grâce à l'effet collectif d'un réseau de neurone interconnecté. Un réseau de neurones est un assemblage de constituants élémentaires interconnectés (appelés « neurones » en hommage à leur modèle biologique), qui réalisent chacun un traitement simple mais dont l'ensemble en interaction fait émerger des propriétés globales complexes. Chaque neurone fonctionne indépendamment des autres de telle sorte que l'ensemble forme un système massivement parallèle. L'information est stockée de manière distribuée dans le réseau sous forme de coefficients synaptiques ou de fonctions d'activation, il n'y a donc pas de zone de mémoire et de zone de calcul, l'une et l'autre sont intimement liés. Les réseaux de neurones artificiels consistent en des modèles plus ou moins inspirés du fonctionnement cérébral de l'être humain en se basant principalement sur le concept de neurone [70].

2.3.1 Fonctionnement d'un neurone artificiel

Un "neurone" est une fonction algébrique non linéaire et bornée, dont la valeur dépend des paramètres appelés coefficients ou poids. Les variables de cette fonction sont habituellement appelées "entrées" du neurone, et la valeur de la fonction est appelée sa "sortie". Un neurone est donc avant tout un opérateur mathématique [70].

Le modèle mathématique d'un neurone artificiel est illustré à la figure 2.3. Un neurone est essentiellement constitué d'un intégrateur qui effectue la somme pondérée des entrées. Le résultat s de cette somme est ensuite transformé par une fonction d'activation appelée aussi fonction de transfert f qui produit la sortie y du neurone., les n entrées du neurone correspondent au vecteur $x = [x_1, x_2, x_3, \dots, x_n]^T$, alors que $w = [w_{11}, w_{21}, w_{31}, \dots, w_{n1}]^T$ représente le vecteur des poids du neurone. La sortie s de l'intégrateur est donnée par l'équation suivante :

$$s = \sum_{i=1}^n w_{i1} x_i \pm b = w_{11}x_1 + w_{21}x_2 + \dots + w_{n1}x_n \pm b \quad (2.1)$$

Que l'on peut aussi écrire sous forme matricielle :

$$s = w^T x \pm b \quad (2.2)$$

Alors, la sortie y du neurone est :

$$y = f(s) = f(w^T x \pm b) \quad (2.3)$$

En remplaçant w^T par W d'une seule ligne, on obtient :

$$y = f(Wx \pm b) \quad (2.4)$$

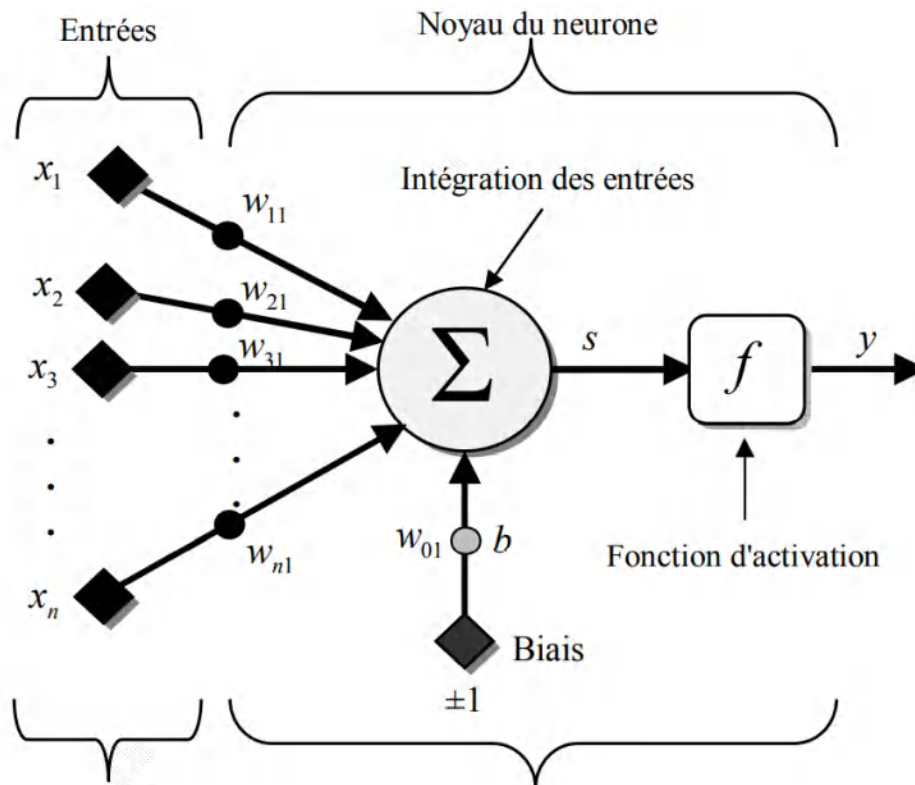


Figure 2.3 Modèle d'un neurone artificiel. [70]

Il existe de nombreux types de fonction d'activation, leurs valeurs de sortie peuvent être dans l'intervalle $[0, +1]$ (ou $[-1, +1]$). Les trois principales fonctions d'activation sont les suivantes :

- Fonction sigmoïde :

La fonction sigmoïde est une classe de fonctions d'activations ayant une forme de S, Sa plage est comprise entre 0 et 1 (figure 2.4).

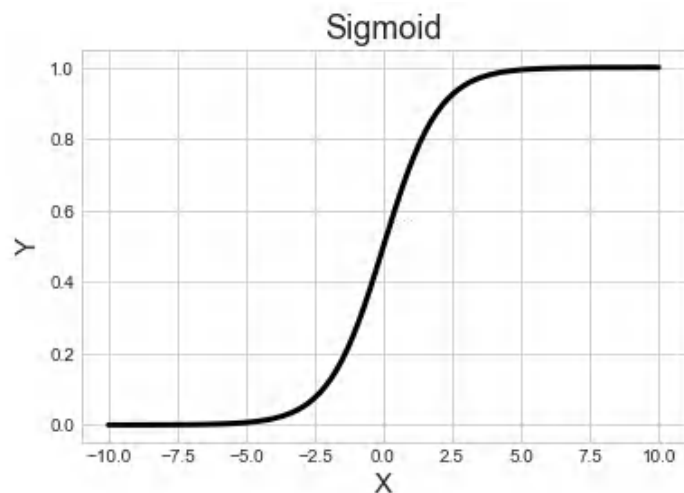


Figure 2.4 Tracé de la fonction sigmoïde

Chapitre 2 : Notions de base sur l'apprentissage profond

Elle est facile à utiliser et possède toutes les propriétés intéressantes des fonctions d'activation : non linéaire, continuellement différentiable, monotone et plage de sortie fixe. Sa formule mathématique est :

$$y = \frac{1}{1 + e^{-x}} \quad (2.5)$$

- La fonction tangente hyperbolique (tanh) :

Cette fonction faisant partie de la famille des sigmoïdes, comprise entre (-1 et 1).

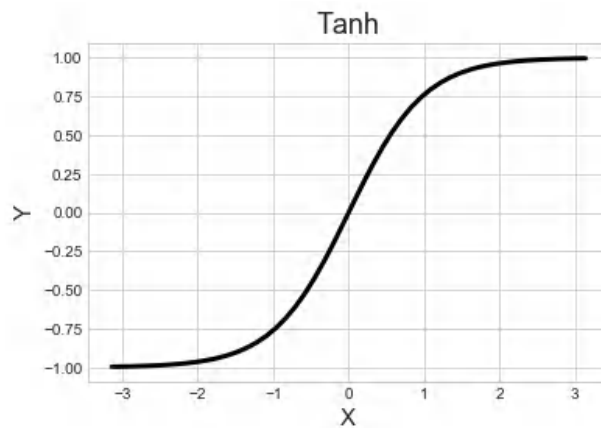


Figure 2.5 Tracé de La fonction tangente hyperbolique

Sa formule mathématique est :

$$y = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2.6)$$

- La fonction d'activation ReLU (Rectified Linear Unit) :

La ReLU est la fonction d'activation la plus utilisée. Elle est définie comme :

$$y = \begin{cases} 0, & \text{si } x < 0 \\ x, & \text{si } x \geq 0 \end{cases} \quad (2.7)$$

La fonction ReLU n'est pas linéaire, ce qui signifie que nous pouvons facilement propager les erreurs en arrière et que plusieurs couches de neurones sont activées par la fonction ReLU.

La fonction d'activation de la couche de sortie dépend du type de problème que l'on veut résoudre. Lorsqu'on souhaite prédire si oui ou non l'objet étudié appartient à une classe donnée, on choisira généralement une fonction sigmoïde, nous donnant une probabilité. Si on veut déterminer la classe de l'objet étudié parmi un panel de K classes possibles, on choisira généralement un *softmax*, défini par l'équation 2.8. Cette fonction prend en entrée \mathbf{x} un vecteur

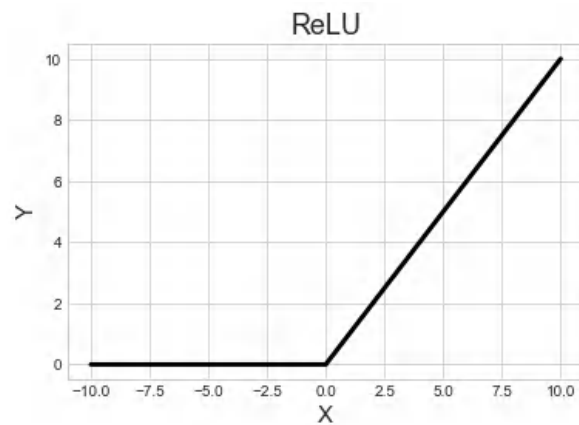


Figure 2.6 Tracé de La fonction Rectified Linear Unit

de réels de dimension K et produit un vecteur de réels strictement positifs, de dimension K , dont la somme des valeurs vaut 1. On obtient ainsi la probabilité pour l'objet d'appartenir à chacune des classes. Sa formule est la suivante [71]:

$$\text{Softmax}(x)_j = \frac{e^{x_j}}{\sum_{k=1}^K e^{x_k}} \forall j \in \{1, \dots, K\} \quad (2.8)$$

2.3.2 Réseaux de neurone monocouche

Ce réseau ne contient aucune couche cachée, ce qui signifie qu'il ne se compose qu'une couche d'entrée et une couche de sortie. Ces structures lui permettent d'être considéré comme un classifieur linéaire : autrement dit, il peut classer les données en fonction de deux caractéristiques.

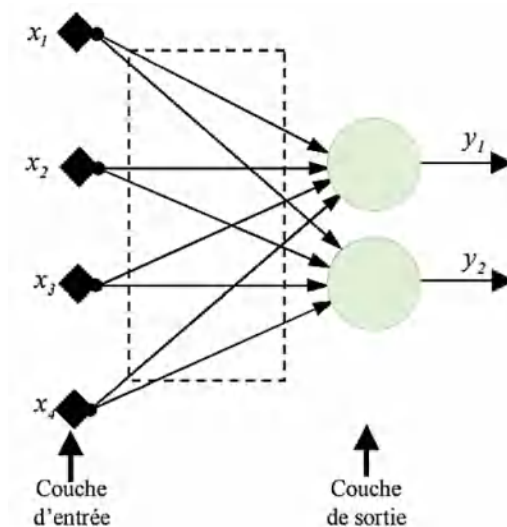


Figure 2.7 Schéma d'un réseau de neurones monocouche.

2.3.3 Réseaux de neurone multicouche

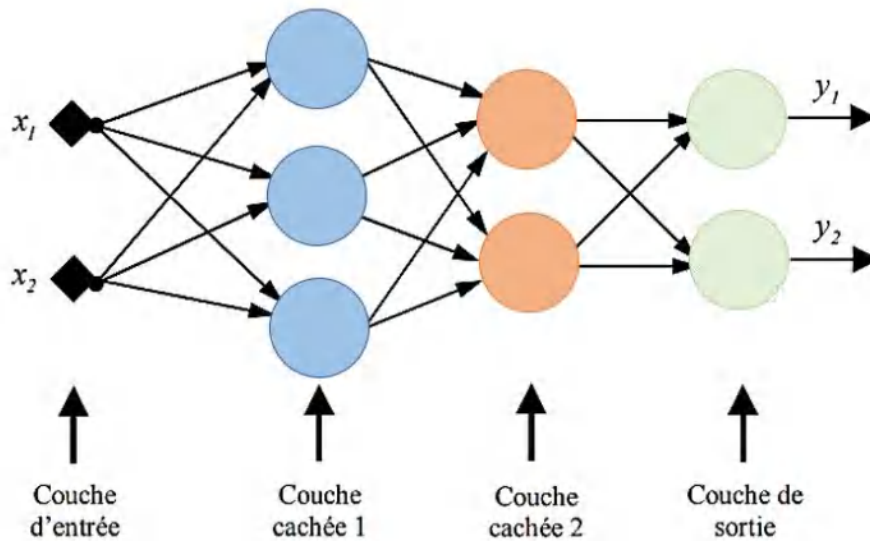


Figure 2.8 Schéma d'un réseau de neurones multicouches.

Il se compose d'une couche d'entrée, d'une couche de sortie et d'une ou plusieurs couches cachées (voir figure 2.8). Chaque neurone d'une couche possède des connexions dirigées vers les neurones de la couche suivante. Les perceptrons multicouches (MLP) sont beaucoup plus utiles et une bonne raison est qu'ils sont capables d'apprendre des représentations non linéaires [72]. MLP est largement utilisé pour résoudre des problèmes nécessitant un apprentissage supervisé. Les applications incluent la reconnaissance vocale, la reconnaissance d'images et la traduction automatique [73].

2.3.4 Apprentissage et rétropropagation

Pour entraîner un réseau de neurones, c'est-à-dire apprendre les poids de connexion, nous allons chercher à minimiser l'erreur de prédiction sur le jeu d'entraînement. Donc, l'entraînement est un processus itératif. Après chaque observation, nous allons ajuster les poids de connexion de sorte à réduire l'erreur de prédiction faite par le réseau de neurones dans son état actuel. Pour cela, nous allons utiliser l'algorithme de la descente de gradient.

Pour alléger les notations, on suppose que la première composante de chaque observation est égale à 1, ce qui permet de traiter les biais comme des poids pour simplifier les équations [74].

2.3.4.1 Notations

$(x_n, y_n)_{n=1 \dots N}$ désigne la base d'apprentissage faite de N observations étiquetées. Les x_n sont des vecteurs de dimensions d (d neurones en entrée), et les étiquettes y_n sont, selon les situations :

Chapitre 2 : Notions de base sur l'apprentissage profond

- Pour un problème de classification biclassé : $y_n = 0$ ou 1 (un neurone en sortie avec une activation σ);
- Pour un problème de classification à $K > 2$ classes y_n est un vecteur de dimension K fait de 0 et un seul 1 ($s=K$ neurones de sorties dans une couche SoftMax).

La sortie du perceptron sur l'entrée $x_n=(x_{n1}, \dots, x_{nd})$ est notée : $\hat{y}_n=(y_{n1}, \dots, y_{ns})$. Dans tous les cas, le but ultime est que la prédiction \hat{y}_n ne soit pas très éloigné de l'étiquette d'apprentissage y_n .

2.3.4.2 Coût d'erreur

Le risque empirique (coût d'erreur moyen sur la base d'apprentissage) est la moyenne des coûts des erreurs E_n sur chaque observation étiquetée (x_n, y_n) :

$$E(w) = \frac{1}{N} \sum_{n=1}^N E_n(w) \quad (2.9)$$

Il s'agit d'une fonction de w , l'ensemble des poids du réseau. Il faut choisir la fonction E_n en fonction du problème traité. Les choix les plus populaires sont :

- Pour un problème de classification biclassé :

$$E_n(w) = -y_n \log \hat{y}_n - (1 - y_n) \log(1 - \hat{y}_n) \quad (2.10)$$

- Pour un problème de classification à $K > 2$ classes :

$$E_n(w) = - \sum_{j=1}^k y_{nj} \log \hat{y}_{nj} \quad (2.11)$$

Ce coût est l'entropie croisée entre les distributions de probabilité discrètes (y_{n1}, \dots, y_{nk}) et $(\hat{y}_{n1}, \dots, \hat{y}_{nk})$. Ici, par convention : $0 \log(0) = 0$. L'objectif est de choisir les poids w qui minimisent l'erreur moyenne $E(w)$.

2.3.4.3 Descente de gradient et perceptron multicouche

L'algorithme de descente de gradient permettant de minimiser le risque empirique $E(w)$ met à jour chaque poids w_{ji} (poids du neurone i de la couche k vers le neurone j de la couche $k+1$) de la manière suivante :

$$w_{ji} \leftarrow w_{ji} - \eta \frac{\partial E}{\partial w_{ji}}(w) \quad (2.12)$$

Chapitre 2 : Notions de base sur l'apprentissage profond

Où le paramètre positif η est le pas de descente, appelé « taux d'apprentissage » lorsqu'on entraîne un classifieur.

Par linéarité de la dérivation :

$$\frac{\partial E}{\partial w_{ji}} = \frac{1}{N} \sum_{n=1}^N \frac{\partial E_n}{\partial w_{ji}}(w)$$

Nous allons donc commencer par calculer les dérivées partielles $\partial E / \partial w_{ji}$.
Le signal arrivant au neurone j est :

$$a_j = \sum_i w_{ji} z_i$$

Où $z_i (= \sigma(a_i))$ est le signal sorti par le neurone i connecté à j

La formule de dérivation composée nous indique (ici, $a_i = g(w_{ji}) \in \mathbb{R}$) :

$$\frac{\partial E_n}{\partial w_{ji}} = \frac{\partial E_n}{\partial a_j} \frac{\partial a_j}{\partial w_{ji}}$$

On a $\frac{\partial a_j}{\partial w_{ji}} = z_i$, donc, en posant $\delta_j = \frac{\partial E_n}{\partial a_j}$:

$$\frac{\partial E_n}{\partial w_{ji}} = \delta_j z_i \tag{2.13}$$

Il s'avère qu'il existe une astuce pour calculer efficacement les δ_j en chaque neurone : l'algorithme de rétropropagation des erreurs (*backpropagation*). Il s'appuie sur le calcul des δ_j aux neurones de sortie, puis calcule, de proche en proche, les δ_j des couches cachées jusqu'aux neurones de la couche cachée la plus proche de l'entrée du réseau.

Calcul de δ_j aux neurones de sortie

Nous montrons qu'aux neurones de la couche de sortie, $\delta_j = \frac{\partial E_n}{\partial a_j}$ est calculable en fonction des sorties \hat{y} du réseau, qui dépendent bien entendu de la fonction de coût adoptée.
Pour reprendre les fonctions de coût introduite précédemment :

-Problème de classification biclassé : $E_n(w) = -y_n \log \hat{y}_n - (1 - y_n) \log(1 - \hat{y}_n)$ où $\hat{y}_n = \sigma(a)$ est la sortie du neurone de la dernière couche, avec σ la fonction sigmoïde.

Chapitre 2 : Notions de base sur l'apprentissage profond

Comme σ vérifie $\log(\sigma(a)) = -\log(1 + e^{-a})$ et $\log(1 - \sigma(a)) = -a - \log(1 + e^{-a})$, on peut écrire $E_n = y_n \log(1 + e^{-a}) + (1 - y_n)a + (1 - y_n) \log(1 + e^{-a}) = a - y_n a + \log(1 + e^{-a})$, donc :

$$\delta = \frac{\partial E_n}{\partial a} = 1 - y_n - \frac{e^{-a}}{1 + e^{-a}} = \frac{1}{1 + e^{-a}} - y_n$$

Soit :

$$\delta = \hat{y}_n - y_n \quad (2.14)$$

-Problème de classification à $s > 2$ classes : $E_n = -\sum_{l=1}^s y_{nl} \log(\hat{y}_{nl})$ où $\hat{y}_{nl} = e^{a_l} / \sum_m e^{a_m}$ (Couche SoftMax). Autrement écrit :

$$E_n = -\sum_{l=1}^s y_{nl} \left(a_l - \log \left(\sum_m e^{a_m} \right) \right)$$

Par conséquent :

$$\delta_j = \frac{\partial E_n}{\partial a_j} = - \left(y_{nj} - \frac{e^{a_j}}{\sum_m e^{a_m}} \right) \sum_{l=1}^s y_{nl}$$

Comme pour tout n , $\sum_l y_{nl} = 1$, on peut conclure. Au neurone de sortie j :

$$\delta_j = \hat{y}_{nj} - y_{nj} \quad (2.15)$$

Pour toutes les fonctions de coût vues ici, on voit que δ_j est calculé en un neurone de sortie comme la différence entre la sortie attendue y et la sortie observée \hat{y} .

Calcul de δ_j aux neurones cachés

On calcule δ_j en un neurone j appartenant à une couche k . On applique la règle de dérivation composée en exprimant la dépendance de E_n par rapport au vecteur des (a_l) représentant les entrées des neurones de la couche $k + 1$ liés au neurone j :

$$\delta_j = \frac{\partial E_n}{\partial a_j} = \sum_l \frac{\partial E_n}{\partial a_l} \frac{\partial a_l}{\partial a_j}$$

Chapitre 2 : Notions de base sur l'apprentissage profond

Ici, l parcourt l'ensemble des neurones de la couche $k + 1$ auxquels est connecté j de la couche k .

De plus :

$$\frac{\partial a_l}{\partial a_j} = \frac{\partial}{\partial a_j} \sum_i w_{li} \sigma(a_i) = w_{lj} \sigma'(a_j)$$

où i parcourt l'ensemble des neurones de la couche k auquel le neurone l de la couche $k + 1$ est connecté : le neurone j est l'un des i , les autres ne dépendent pas des a_j . σ' désigne la dérivée de l'activation σ .

On a donc obtenu la formule :

$$\delta_j = \sigma'(a_j) \sum_l w_{lj} \delta_l \quad (2.16)$$

Conclusion : calcul des δ_j

On peut calculer tous les δ sur le réseau par *rétropropagation des erreurs*. On commence par calculer les erreurs comme différence entre sortie attendue y_n et sortie du réseau \hat{y}_n , puis on propage les δ_j vers les couches d'entrée à l'aide de la formule (2.16).

L'algorithme de calcul des δ_j avec propagation avant suivie de rétropropagation est illustré par la figure 2.9.

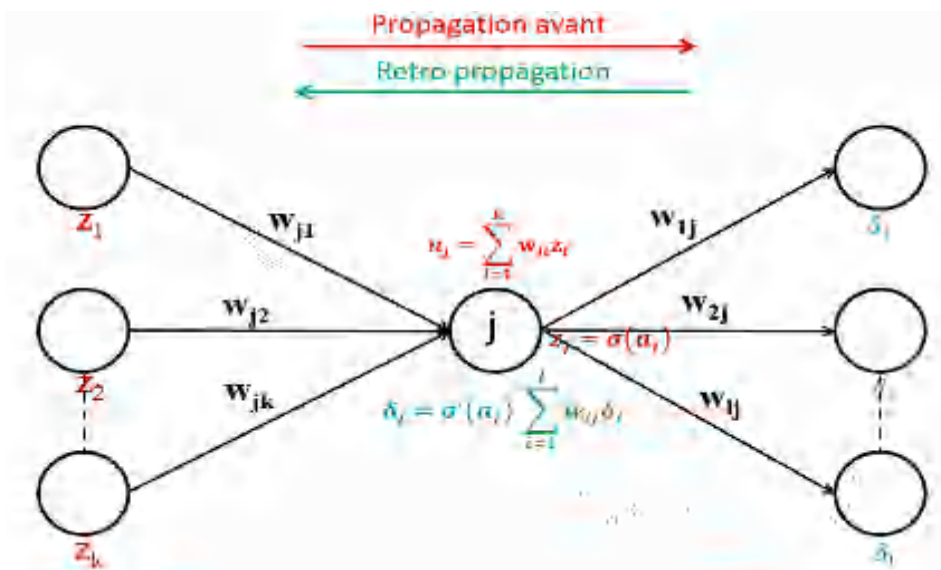


Figure 2. 9 Illustration des deux étapes successives de propagation avant puis rétropropagation pour calculer les z et δ – intervenant dans les dérivée partielle dE/dw_{ji} . Dans cette figure, on a isolé le neurone j , et représenté uniquement les neurones qui lui sont connectés dans la couche précédente et dans la couche suivante. [74]

2.3.4.4 Apprentissage et rétropropagation des erreurs

L'apprentissage (c'est-à-dire l'adaptation des poids w du réseau de manière à minimiser le coût moyen d'erreur $E(w)$) peut donc se faire à l'aide d'un algorithme de descente, en itérant les deux étapes suivantes jusqu'à satisfaction d'un critère d'arrêt :

1. Pour chaque observation étiquetée (x_n, y_n) en entrée :
 - (a) Propagation avant pour calculer les a_j et z_j en chaque neurone, ainsi que la sortie \hat{y}_n du réseau ;
 - (b) Calcul des δ_j aux neurones de sortie : $\delta_j = \hat{y}_{nj} - y_{nj}$
 - (c) Rétropropagation des δ : en chaque neurone caché, calcul de $\delta_j = \sigma'(a_j) \sum_l w_{lj} \delta_l$
 - (d) Calcul des dérivées partielles : $\frac{\partial E_n}{\partial w_{ji}}(w) = \delta_j z_i$
2. Pour chaque poids w_{ji} , calcul de $\frac{\partial E}{\partial w_{ji}}(w) = \frac{1}{N} \sum_{n=1}^N \frac{\partial E_n}{\partial w_{ji}}(w)$ et mise à jour :

$$w_{ji} \leftarrow w_{ji} - \eta \frac{\partial E}{\partial w_{ji}}(w)$$

En pratique, il est coûteux en temps de calcul et occupation mémoire de ne faire la mise à jour dans l'étape 2 qu'une fois que l'intégralité de la base d'exemples a été parcourue dans l'étape 1.

En pratique, deux variantes sont utilisées dans les implantations logicielles :

- version *online* / *algorithme du gradient stochastique* : mise à jour des poids pour *chaque* observation de la base d'apprentissage, selon la règle :

$$w_{ji} \leftarrow w_{ji} - \eta \frac{\partial E_n}{\partial w_{ji}} \quad (2.17)$$

-version *mini-batch* / *par mini-lot* : mise à jour des poids après parcours d'un sous ensemble B_m d'observations :

$$w_{ji} \leftarrow w_{ji} - \eta \sum_{n \in B_m} \frac{\partial E_n}{\partial w_{ji}} \quad (2.18)$$

Tel que $\cup_m B_m = \{1, \dots, N\}$

Afin d'accélérer la descente de gradient, de nombreuses autres méthodes ont été proposées pour les réseaux de neurones tels que AdaGrad [75], Adadelta [76], ou Adam [77]. L'une des plus utilisées est la méthode d'optimisation Adam [77]. Cette méthode se base sur la descente de gradient avec de petits batches et reprend les mêmes principes que AdaDelta. L'idée consiste à adapter le taux d'apprentissage à partir d'une estimation du premier et du second moment du gradient (contrairement à la méthode du momentum qui n'utilise que le premier moment). L'estimation du premier et second moment nécessite cependant le maintien et la mise à jour de deux variables supplémentaires pour chaque paramètre du réseau de neurones. Ce type de méthode à l'avantage d'être relativement robuste et permet de s'adapter automatiquement le taux d'apprentissage au cours de l'apprentissage pour chaque poids. [78]

2.3.5 Problème de sur-apprentissage et sous-apprentissage

Lorsqu'on utilise un modèle d'apprentissage automatique, on sépare le jeu de données étiquetées en trois parties : les ensembles d'apprentissage, de validation, et de test. Lors de la phase d'apprentissage le modèle est entraîné sur l'ensemble d'apprentissage. Ensuite, on utilise l'ensemble de test pour mesurer l'erreur du modèle sur des données qu'il n'a jamais vues. L'ensemble de validation sert à contrôler l'apprentissage du modèle, particulièrement le choix des hyper-paramètres. Cette séparation n'est utile que sous la condition que chaque ensemble est indépendant des autres et observe la même distribution.

L'apprentissage d'un réseau de neurones consiste à minimiser l'erreur empirique. On est donc susceptible de se retrouver dans une situation de sur-apprentissage et sous apprentissage, comme illustré par la figure 2.10. Le premier cas illustre un sous-apprentissage. Le modèle

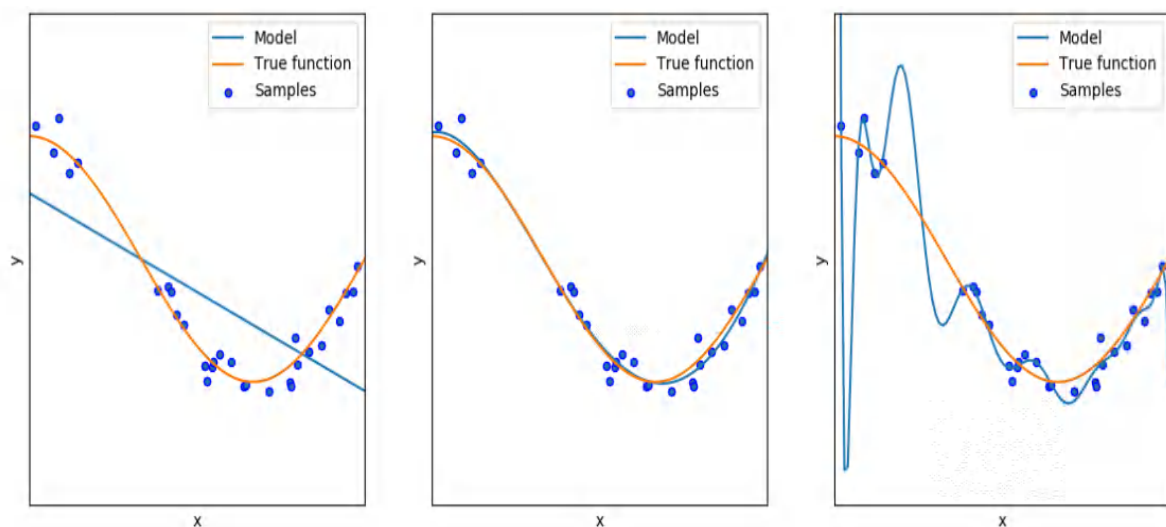


Figure 2.10 Les différents niveaux d'apprentissage. [79]

n'est pas assez complexe. Ne pouvant décrire que des droites, il ne peut pas modéliser la fonction voulue. Le deuxième cas illustre un apprentissage correct. Avec le bon niveau de complexité, le modèle a su approcher la fonction voulue de très près. Le troisième cas illustre un sur-apprentissage. Le modèle trop complexe a surpasser par la plupart des points appartenant aux données d'entraînement. On voit qu'à certains endroits la courbe du modèle (en bleu) est très éloignée de la courbe décrivant la véritable fonction (en orange). Si dans le jeu de test des données se trouvent à cet endroit, l'erreur sera très grande. [71]

Le sur-apprentissage est observé lorsque l'erreur d'apprentissage diminue avec l'augmentation des itérations, tandis que l'erreur de validation commence à augmenter progressivement comme illustré sur la figure 2.11. Ce point de rupture marque la spécialisation du modèle sur les



Figure 2.11 Fonction de coût calculée sur l'ensemble d'apprentissage et sur l'ensemble de validation en fonction du nombre d'époque.

données d'apprentissage au détriment de la capacité de généralisation du modèle (c'est-à-dire, la capacité à traiter des nouvelles données).

On peut utiliser différentes méthodes d'éviter le sur-apprentissage :

Early stopping : La méthode la plus simple pour éviter le sur-apprentissage consiste à arrêter l'apprentissage quand l'erreur sur une base d'observations de test (ou base de validation), indépendante de la base d'apprentissage, commence à augmenter. On peut aussi tout simplement arrêter l'apprentissage après un nombre d'époques fixé a priori [74].

Régularisation : Le sur-apprentissage étant dû à l'apprentissage d'un modèle trop complexe par rapport au problème initial, des techniques pour l'éviter consistent à pénaliser cette complexification du modèle. Pour les réseaux de neurones, cela consiste à pénaliser les poids

Chapitre 2 : Notions de base sur l'apprentissage profond

de connexions trop importants en ajoutant un terme de régularisation à la fonction de coût à minimiser. Ce terme pénalise les poids des connexions trop fortes [78].

Dropout : Le dropout est la technique de régularisation la plus populaire pour les réseaux neuronaux afin d'éviter le problème de sur-apprentissage. Pendant le temps d'entraînement, à chaque itération, un neurone est "dropped" ou désactivé avec une probabilité p ; En d'autres termes, toutes les entrées et les sorties de ce neurone seront désactivées à l'itération actuelle, les neurones abandonnés sont rééchantillonnés avec la probabilité p à chaque étape d'entraînement, donc un neurone abandonné à une étape peut être actif à la prochaine étape. Le paramètre p est appelé taux de dropout et il s'agit généralement d'un nombre autour de 0,5, ce qui signifie que 50% des neurones doivent être abandonnés [80].

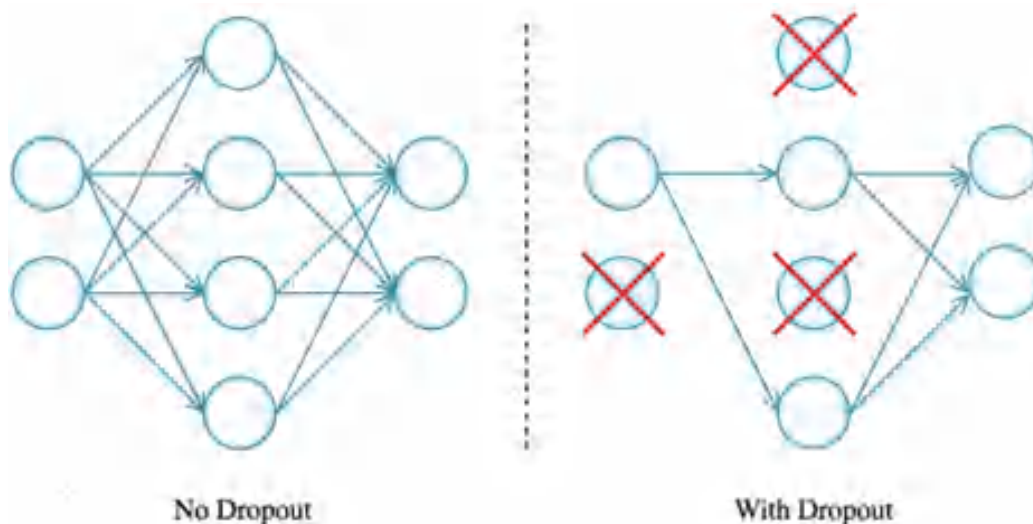


Figure 2.12 Principe de dropout. [80]

Augmentation de données : Il s'agit d'une technique permettant de contourner le manque de données qui réduit les chances du modèle d'obtenir de bonnes performances en généralisation.

Il existe deux types d'augmentation :

- Par transformation des données, ou
- Par création de données synthétiques.

La transformation des données se fait souvent par application de transformations géométriques (e.g. retournement horizontal, rotation), modification du contraste, découpage aléatoire, ajout de bruit blanc, effacement aléatoire des valeurs de pixel d'une zone de l'image. La création de données synthétiques est notamment possible par l'utilisation des réseaux de neurones de type antagoniste génératif (en anglais Generative Adversarial Network).

Généralement, l'augmentation des données est utilisée pendant la phase d'entraînement du modèle, donc à partir du jeu de données d'entraînement. Les jeux de données de validation et de test ne sont pas soumis à cette technique.

2.4 Les réseaux de neurones profonds

Les réseaux de neurones ont très longtemps été limités dans leurs architectures, en particulier concernant leur profondeur, c'est-à-dire, le nombre de couches qu'ils pouvaient apprendre. Cette limitation s'est effondrée dans les années 2010-2012 avec l'arrivée de bases de données bien plus grandes accompagnées de capacités de calcul et de stockage plus importantes [81]. Cette avancée a également été permise par des architectures de réseaux de neurones différentes, plus faciles à apprendre et mieux adaptées à certains types de données.

Le terme profond réfère au nombre de couches des réseaux de neurones entre l'entrée et la sortie. Un réseau avec une seule couche cachée est appelé réseau peu profond, et au contraire, un réseau avec plus de 2 couches cachées est dit profond. De nos jours, il est possible de trouver des réseaux avec une centaine, voir un millier de couches pour les plus profonds [64], [65].

2.4.1 Réseaux de neurones convolutifs

Les réseaux de neurones convolutifs (CNN) ont été introduits par Lecun et al. [82]. La particularité des CNN est l'utilisation de l'opération de convolution dans les premières couches intermédiaires du réseau de neurones. À l'origine, cette opération est utilisée comme filtre dans le domaine de l'image ou du son afin de mettre en évidence des motifs ou réduire un type de bruits.

2.4.2 Architecture d'un CNN

Un CNN classique (figure 2.12) est généralement composé de trois types de couches :

- Les couches convolutives, qui contiennent plusieurs opérations de convolutions appliquées sur la même entrée,
- Les couches de mise en commun,
- Les couches entièrement connectées.

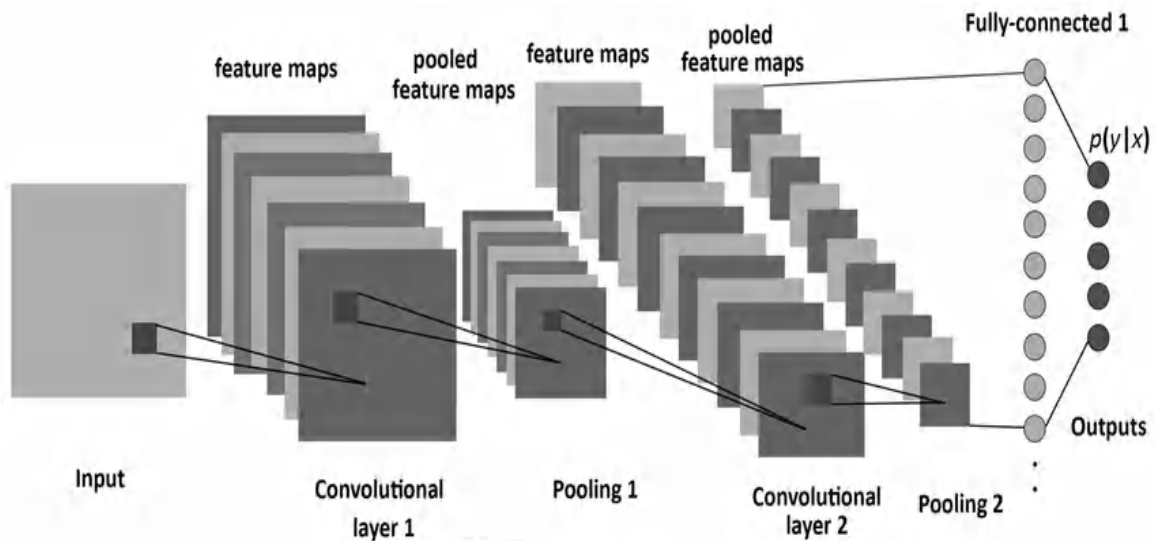


Figure 2.13 Architecture standard d'un réseau de neurones convolutionnel. [83]

2.4.2.1 Couche de convolution (CONV)

Dans le cas d'un CNN, la convolution est effectuée sur les données d'entrée à l'aide d'un filtre (autrement dit noyau) pour produire ensuite une carte de caractéristiques comme montré sur la figure 2.13. On exécute une convolution en faisant glisser le filtre sur l'entrée. À chaque emplacement, une multiplication matricielle par élément est effectuée entre une matrice de filtre de taille 3×3 et une zone de taille 3×3 de la matrice de l'image d'entrée. Les éléments de la matrice résultante sont additionnés, cette somme est la valeur de sortie ("pixel de destination") sur la carte des caractéristiques. Nous faisons ensuite glisser le filtre vers la droite et effectuons la même opération, en ajoutant également ce résultat à la carte des caractéristiques.

Plusieurs filtres sont utilisés pour une entrée et les cartes de caractéristiques résultantes sont réunies pour la sortie finale d'une couche convolutive [84].

L'exemple sur la figure 2.13 représente une opération de convolution en 2D à l'aide d'un filtre 3×3 . Mais en réalité ces circonvolutions sont réalisées en 3D. En réalité, une image est représentée sous la forme d'une matrice 3D avec des dimensions de hauteur, de largeur et de profondeur, où la profondeur correspond aux couches de couleur (RVB). Un filtre à convolution a une hauteur et une largeur spécifiques, comme 3×3 ou 5×5 , et de par sa conception, il couvre toute la profondeur de son entrée, il doit donc également être 3D.

La carte de caractéristiques résultante a une profondeur arbitraire fixée lors de la conception du réseau. On l'a choisi en fixant le nombre de filtre qui seront appliquées à la matrice d'entrée. Le résultat de chaque convolution correspond à une couche dans la carte de caractéristiques de sortie. Le pas de convolution est aussi choisi lors de la conception du réseau. Il régit le

Chapitre 2 : Notions de base sur l'apprentissage profond

glissement du filtre. Lorsqu'un emplacement a été traité, le filtre se déplace d'un nombre de pixels correspondant au pas. La taille de filtre est impaire. On s'intéresse au pixel central et on s'appuie pour cela d'un nombre fixe de pixels dans chaque direction. Lorsque ce pixel central se trouve proche d'un bord de la matrice d'entrée, il peut arriver que le filtre déborde de la carte des caractéristiques. On doit choisir entre ne pas traiter ces pixels, ou ajouter une marge à la matrice d'entrée de manière à ce que le cas ne se produise pas. Cette marge peut ne contenir que des pixels nuls, ou être construite à partir de la matrice d'entrée, typiquement grâce à des symétries.

Ainsi, lors de l'utilisation d'un CNN, les quatre hyperparamètres importants sur lesquels nous devons nous prononcer sont :

- La taille du filtre ;
- Le nombre de filtres (c'est-à-dire combien de filtres voulons-nous utiliser) ;
- Le pas de convolution (stride en anglais) (quelle sont les marches du filtre) ;
- Remplissage de zéros (padding) ;

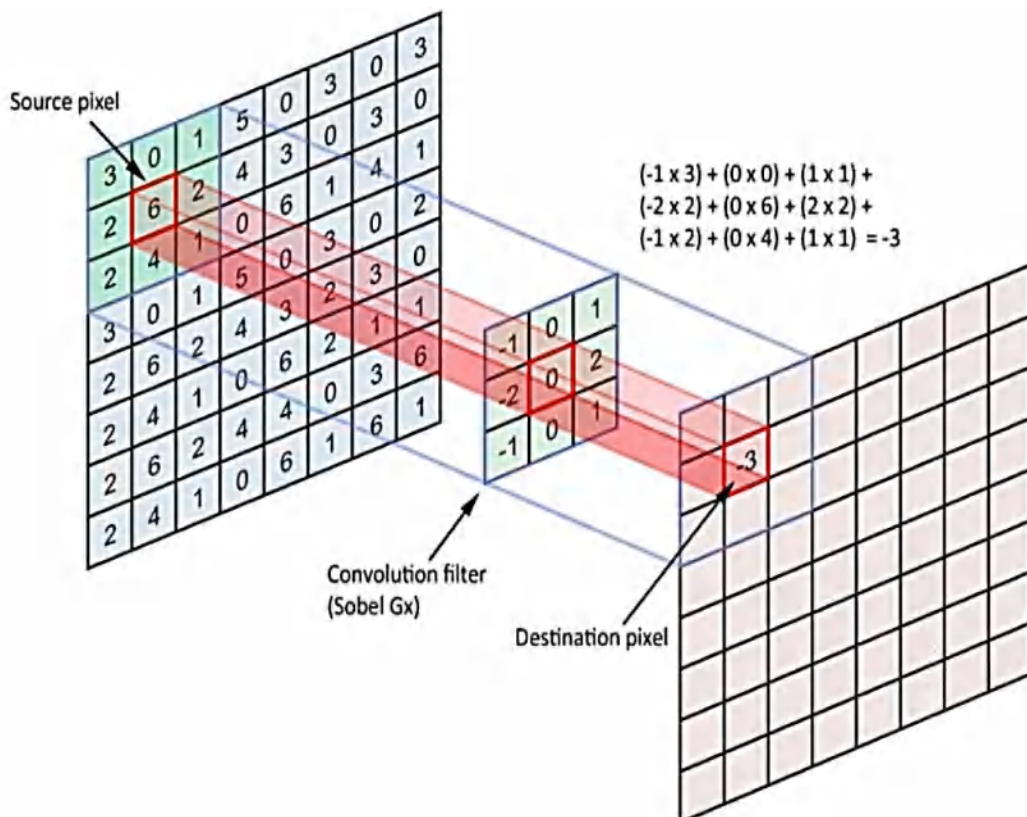


Figure 2.14 Principe d'une opération de convolution. [84]

2.4.2.2 Couche de mise en commun

La mise en commun (pooling en anglais) réduit la taille spatiale des cartes de caractéristiques, tout en préservant la profondeur [80]. Cela est bénéfique car cela diminue la puissance de calcul nécessaire pour traiter les données, tout en extrayant les caractéristiques dominantes dans les cartes de caractéristiques [85]. Il existe plusieurs types de pooling :

- Le « *max pooling* », qui revient à prendre la valeur maximale de la sélection. C'est le type le plus utilisé car il est rapide à calculer (immédiat), et permet de simplifier efficacement l'image.
- Le « *mean pooling* » (ou average pooling), soit la moyenne des pixels de la sélection : on calcule la somme de toutes les valeurs et on divise par le nombre de valeurs. On obtient ainsi une valeur intermédiaire pour représenter ce lot de pixels.
- Le « *sum pooling* », c'est la somme sans avoir divisé par le nombre de valeurs (on ne calcule que leur somme).

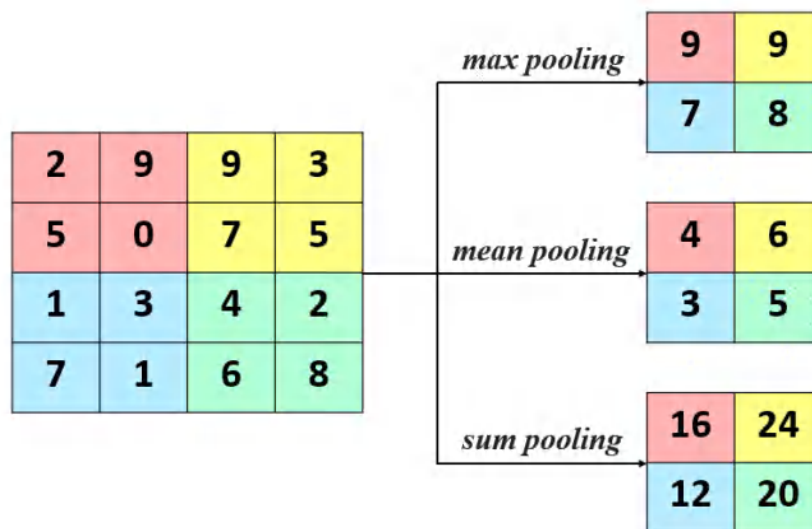


Figure 2.15 Types de mise en commun

2.4.2.3 Couche entièrement connectée (FC)

Ces couches sont placées en fin d'architecture de CNN et sont entièrement connectées à tous les neurones de sorties (d'où le terme fully-connected). Après avoir reçu un vecteur en entrée, la couche FC applique successivement une combinaison linéaire puis une fonction d'activation dans le but final de classifier l'input image. Elle renvoie enfin en sortie un vecteur de taille d correspondant au nombre de classes dans lequel chaque composante représente la probabilité pour l'input image d'appartenir à une classe.

2.5 Apprentissage par transfert

Le Transfer Learning, ou apprentissage par transfert en français, désigne l'ensemble des méthodes qui permettent de transférer les connaissances acquises à partir de la résolution de problèmes donnés pour traiter un autre problème. Le Transfer Learning a connu un grand succès avec l'essor du Deep Learning. En effet, bien souvent, les modèles utilisés dans ce domaine nécessitent des temps de calcul élevés et des ressources importantes. Or, en utilisant des modèles pré-entraînés comme point de départ, le Transfer Learning permet de développer rapidement des modèles performants et résoudre efficacement des problèmes complexes en Computer Vision ou Natural Language Processing (NLP) [76]. Intuitivement, le Transfer Learning s'inspire fortement du processus avec lequel nous apprenons. L'utilisation des méthodes de Transfer Learning en Deep Learning consiste principalement à exploiter des réseaux de neurones pré-entraînés. Généralement, ces modèles correspondent à des algorithmes très performants qui ont été développés et entraînés sur de larges bases de données et qui sont aujourd'hui librement partagés.

Dans ce contexte, on peut distinguer 2 types de stratégies :

- **Utilisation de modèles pré-entraînés comme extracteurs de caractéristiques :**

L'architecture des modèles de deep learning se présente très souvent sous la forme d'un empilement de couches de neurones. Ces couches apprennent différentes caractéristiques en fonction du niveau où elles se situent. La dernière couche (généralement une couche entièrement connectée, dans le cas de l'apprentissage supervisé) est utilisée pour obtenir la sortie finale. Plus la couche se situe en profondeur, plus elle permet d'extraire des features spécifiques.

L'idée est donc de réutiliser un réseau pré-entraîné sans sa couche finale. Ce nouveau réseau fonctionne alors comme un extracteur de caractéristiques fixes pour la réalisation d'autres tâches.

- **Ajustement de modèles pré-entraînés :**

Il s'agit d'une technique plus complexe, dans laquelle non seulement la dernière couche est remplacée pour réaliser la classification ou la régression, mais d'autres couches sont également ré-entraîner de manière sélective. En effet, les réseaux neuronaux profonds sont des architectures hautement configurables avec divers hyperparamètres. De plus, alors que les premières couches capturent les caractéristiques génériques, les dernières couches se concentrent davantage sur la tâche spécifique à accomplir, l'idée est donc de geler (c'est-à-dire fixer les poids) de certaines couches pendant l'entraînement et affiner le reste pour répondre à

la problématique. Cette stratégie permet de réutiliser les connaissances en termes d'architecture globale du réseau et d'exploiter ses états comme point de départ pour l'entraînement. Elle permet donc d'obtenir de meilleures performances avec un temps d'entraînement plus court.

2.6 DCNN architectures

Il existe plusieurs architectures des réseaux de neurones convolutifs. Les plus courants sont :

AlexNet : introduit en 2012 par Alex Krizhevsky [86]. L'architecture de base d'AlexNet consiste en cinq couches convolutives, deux couches de normalisation, trois couches de max-pooling, trois couches entièrement connectées, et une couche linéaire avec une fonction d'activation Softmax en sortie pour s'assurer que chaque neurone prédit les probabilités de classe d'une image particulière.

VGGNet : a été introduit en 2014 par Karen Simonyan et Andrew Zisserman [87] du groupe de géométrie visuelle (VGG) de l'université d'Oxford. Il obtient l'une des meilleures performances lors du concours de reconnaissance visuelle à grande échelle ImageNet (ILSVRC) 2014. VGGNet utilise des filtres plus petits de 3×3 , par rapport au filtre AlexNet 11×11 , afin de fournir une meilleure extraction des caractéristiques des images. L'étude vérifie également que l'utilisation de filtres beaucoup plus petits afin d'augmenter la profondeur du réseau plutôt que sa largeur joue un rôle critique pour obtenir de meilleures performances. Il existe deux versions de cette architecture : VGG16 et VGG19 avec des profondeurs et des couches différentes. Le VGG19 est plus profond que le VGG16. Cependant, le nombre de paramètres pour VGG19 est plus important et l'entraînement du réseau est donc plus coûteux que pour VGG16.

GoogLeNet : en 2014, les chercheurs de Google ont présenté le réseau GoogLeNet [65], qui a pris la première place du concours ImageNet en 2014 pour les défis de classification et de détection. La plus notable est l'utilisation de modules à branches appelés module d'Inception. L'idée de base est de multiplier les filtres avec des tailles différentes. Le module contient 4 branches, 3 couches de convolutions avec des filtres de taille 5×5 , 3×3 et 1×1 , et une opération de mise en commun avec des fenêtres de taille 3×3 . Les opérations de convolution 1×1 consistent à mettre en relation uniquement les différents canaux sur une même position de l'image. Afin de réduire la quantité de calcul et le nombre de paramètres à apprendre, les auteurs proposent d'ajouter des convolutions 1×1 peu coûteuses pour réduire le nombre de canaux avant les opérations de convolution 3×3 et 5×5 .

MobileNet : a été conçu par les chercheurs de Google [88]. L'architecture MobileNet a été construite sur la base de convolutions séparables en profondeur, suivies d'une convolution ponctuelle avec une couche de convolution 1×1 . Dans la couche de convolution standard, chaque noyau est appliqué à tous les canaux de l'image d'entrée. Alors que la convolution en profondeur est appliquée sur chaque canal séparément. Cette approche réduit considérablement le nombre de paramètres une fois comparée aux convolutions standard avec la même profondeur. MobileNet a atteint des performances inspirantes pour diverses applications avec un nombre réduit d'hyperparamètres et de ressources informatiques.

ResNet: elle a été proposée par He et al [64] les réseaux résiduels profonds (ResNet) étaient une idée révolutionnaire qui permettait de développer des réseaux beaucoup plus profonds (des centaines de couches par opposition à des dizaines de couches). Dans ResNets, les couches de convolution sont divisées en blocs résiduels et, pour chaque bloc, une connexion résiduelle est ajoutée, ce qui contourne le bloc correspondant. Ensuite, la sortie du bloc résiduel est fusionnée par sommation avec l'entrée d'origine transmise par la connexion résiduelle. En ajoutant ces connexions résiduelles, le résultat d'une étape d'apprentissage peut être rétro propagé directement aux couches précédentes, sans aucune interférence des couches suivantes.

DenseNet: les réseaux convolutifs denses (DenseNets) proposé par Huang en 2017 [89] DenseNet est assez similaire à ResNet avec quelques différences fondamentales. ResNet utilise une méthode additive (+) qui fusionne la couche précédente (identité) avec la couche suivante, tandis que DenseNet concatène (.) la sortie de la couche précédente avec la couche suivante.

2.7 L'évaluation de la performance des modèles d'apprentissage automatique

Cette section aborde les métriques qui peuvent être utilisés pour tester et évaluer la performance d'un modèle d'apprentissage automatique, pour les problèmes de classification. Dans le contexte des métriques d'évaluation, nous commençons par définir c'est quoi une matrice de confusion. Ensuite, nous présentons les mesures qui sont directement calculées à partir de cette matrice ainsi que la courbe ROC.

2.7.1 La matrice de confusion

Il s'agit d'un tableau de taille $n \times n$ montrant les résultats des modèles prédictifs pour les problèmes de classification. Où 'n' est le nombre de classes dans nos ensembles de données. Dans cette matrice on croise les classes cibles réelles avec les classes prédites obtenues. Ceci nous donne le nombre d'instances qui sont correctement classées et le nombre d'instances mal classées [90].

Chapitre 2 : Notions de base sur l'apprentissage profond

Tableau 2. 1 Exemple de matrice de confusion pour une classification binaire

		Classes Actuels	
		Négatif	Positif
Classes prédites	Négatif	Vrais Négatifs : VN	Faux Négatifs : FN
	Positif	Faux Positifs : FP	Vrais Positifs : VP

VP : c'est le nombre des vrais positifs, le nombre d'instances positives correctement classifiées,

FP : c'est le nombre des faux positifs, le nombre d'instances qui ne sont pas positives et qui sont prédites comme positives,

FN : c'est le nombre des faux négatifs, le nombre d'instances non négatives classifiées comme négatives,

VN : c'est le nombre des vrais négatifs, le nombre d'instances négatives correctement classifiées.

À partir de la matrice de confusion on peut calculer plusieurs métriques.

- **le taux de succès (Accuracy)** : C'est la proportion des instances qui sont correctement classifiées.

$$Accuracy = \frac{VP + VN}{VP + VN + FP + FN} \times 100 \quad (2.19)$$

- **La précision** : C'est le rapport entre les instances positives correctement prédites et le total des instances positives prédites.

$$Précision = \frac{VP}{VP + FP} \times 100 \quad (2.20)$$

- **le taux d'erreur (E)** : c'est l'erreur global de la classification.

$$E = 1 - \text{taux de succès} \quad (2.21)$$

- **la valeur prédictive positive** : c'est la proportion des instances actuelles positives parmi les instances prédites comme positives

$$VPP = \frac{VP}{VP + FN} \quad (2.22)$$

- **la valeur prédictive négative** : c'est la proportion des instances actuelles négatives parmi les instances prédites comme négatives.

$$VPN = \frac{VN}{VN + FN} \quad (2.23)$$

• **le rappel** : appelé aussi sensibilité : c'est le pourcentage des instances positives correctement identifiées.

$$\text{Rappel} = \frac{VP}{VP + FN} \quad (2.24)$$

• **la spécificité** : c'est le pourcentage d'instances négatives correctement identifiées.

$$\text{Spécificité} = \frac{FP}{FP + VN} \quad (2.25)$$

• **la mesure F** : Il s'agit de la moyenne harmonique de la précision et le rappel. Il prend des valeurs entre 0 et 1. Cette mesure est utilisée quand on cherche une balance entre la précision et le rappel. Plus que la valeur de cette mesure s'approche de 1, plus que notre modèle fait une bonne décision, et le contraire est vrai, quand elle s'approche de zéro.

$$F1 - \text{Score} = 2 \times \frac{\text{Rappel} \times \text{Précision}}{\text{Rappel} + \text{Précision}} \quad (2.26)$$

• **Le coefficient de corrélation de Matthews** est utilisé en apprentissage automatique comme mesure de la qualité des classifications binaires et multi-classes. Il prend en compte les vrais et faux positifs et négatifs et est généralement considéré comme une mesure équilibrée qui peut être utilisée même si les classes sont de tailles très différentes. La CMC est essentiellement une valeur de coefficient de corrélation comprise entre -1 et +1. Un coefficient de +1 représente une prédiction parfaite, 0 une prédiction aléatoire moyenne et -1 une prédiction inverse. Cette statistique est également connue sous le nom de coefficient phi.

$$MCC = \frac{VP \times FN - FP \times VN}{\sqrt{(VP + FP)(VP + FN)(VN + FP)(VN + FN)}} \quad (2.27)$$

2.7.2 La courbe ROC

La courbe ROC (Receiver Operating Characteristic) a été utilisée dans le traitement de signal pour faire la distinction entre le signal et le bruit. Elle est très utilisée en apprentissage automatique pour évaluer la performance des classifieurs. Il s'agit d'une courbe où on croise le taux des vrais positifs (TVP) avec celui des faux négatifs (TFN) pour tous les seuils de classification. On utilise un classifieur aléatoire comme une ligne de base. Si on classe aléatoirement nos instances, on est censé avoir cette ligne.

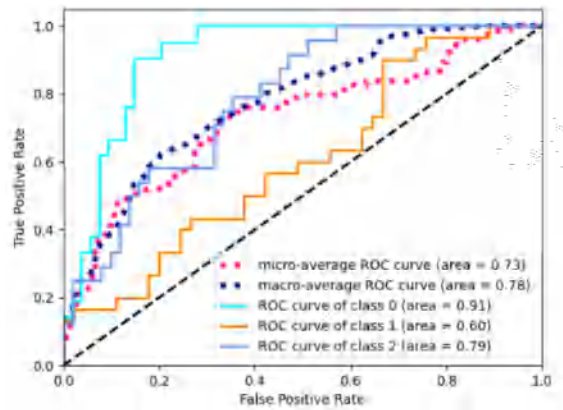


Figure 2. 16 Exemple de courbes ROC pour des classes multiples. [91]

Un modèle avec une discrimination parfaite (sans chevauchement entre les classes) passe par le coin supérieur gauche. Plus que la courbe ROC est proche du coin supérieur, meilleure est la performance de la classification.

L'aire sous la courbe ROC (AUC : Area Under the Curve) représente une mesure qui permet de quantifier numériquement la performance de nos classifieurs :

- si $AUC = 1$, Il s'agit d'un modèle qui fait une séparation parfaite entre nos classes. Il permet de classer toutes les instances positives correctement et fait la même chose avec les autres instances.
- si $AUC = 0.5$, la classification n'est pas meilleure que celle qui serait obtenue si nous générons aléatoirement nos instances. Le modèle dans ce cas, ne fait aucune distinction entre nos classes. Chaque instance a une probabilité de $1/n$ d'être bien classée en utilisant ce modèle. Où n est le nombre de classes.
- si $AUC < 0.5$, notre modèle fait pire qu'une classification aléatoire. Il vaut mieux deviner aléatoirement, qu'utiliser ce modèle.

L'AUC est très utile quand on veut faire une comparaison entre différents modèles. Le meilleur modèle est celui qui a l'AUC le plus élevé.

2.8 Conclusion

Dans ce chapitre nous avons présenté les notions de base de RNA notamment son principe de fonctionnement, ses composantes principales et également ses limitations. Ensuite, nous avons décrit une nouvelle variante de RNA qui s'appelle Deep Learning (DL). Cette technique se caractérise par son habilité de résoudre le problème de la complexité de l'entraînement de (RNA). Ainsi que les notions des réseaux de neurones convolutionnelle, et enfin on a présenté les techniques d'apprentissage par transfert et leur intérêt ainsi que les architectures CNN le

Chapitre 2 : Notions de base sur l'apprentissage profond

plus couramment utilisés. Dans le chapitre suivant on va utiliser les techniques d'apprentissage profond pour la classification des images de CE.

Chapitre 3

Classification des images WCE en utilisant l'apprentissage profond

3.1 Introduction

Généralement, WCE produit environ 86 000 images pour chaque patient [9]. Les gastro-entérologues doivent visualiser toutes ces images pour identifier les anomalies du tractus GI. Cependant, des anomalies du tractus GI peuvent être présentes seulement dans une petite fraction de la vidéo. Par conséquent, l'examen manuel est un processus fastidieux et très long. En plus, la décision peut être sujette à des erreurs humaines. Une technique automatique assistée par ordinateur est fortement exigée pour aider les médecins dans l'examen fastidieux de ces images. Dans ce chapitre on va étudier les problèmes de classification dans les images WCE. Pour cela, nous effectuons deux types de classification : 1) classification binaire (normal ou abnormal) cette classification permet de détecter automatiquement les maladies dans le tube digestif. 2) classification multiclass (stomach, small-intestine et colon) dans le cas où le médecin se concentre sur une seule partie par exemple le colon.

Dans ce contexte, nous proposons deux techniques d'apprentissage : apprentissage à partir de zéros en utilisant notre propre CNN et l'apprentissage par transfert en utilisant le réentraînement des CNN trouvés dans la littérature. Nous proposons également une classification des images WCE de différentes maladies GI. Pour cela, on effectue une méthode d'apprentissage en profondeur basée sur la combinaison de deux réseaux de neurones convolutifs (CNN) pré-entraînés, à savoir MobileNet et DenseNet169. Ensuite on va présenter notre méthode, et enfin on va discuter les résultats.

3.2 Etat de l'art sur la classification des images WCE

La détection automatique d'anomalies dans les images est récemment devenue un domaine de recherche. La prédiction automatique d'anomalies dans les images WCE permet aux médecins de faire un diagnostic précis et de commencer le traitement le plus tôt possible. En plus, elle permet d'économiser du temps et des efforts.

Dans les recherches récentes, l'apprentissage profond (DL) a obtenu des résultats remarquables dans les tâches de traitement et d'analyse des images médicales telles que la classification, la détection, la segmentation et d'autres tâches [92]. Pour la classification des images WCE, de nombreuses méthodes existantes ont été proposées dans la littérature, par exemple, Zou et al. [93] ont utilisé une méthode basée sur les réseaux de neurones convolutionnels (CNN) pour traiter le problème de la classification des organes digestifs dans les images WCE, l'architecture CNN proposé compose de trois couches de convolution suivies de deux couches entièrement connectées. Un autre CNN a également été utilisé dans [94] pour classer les anomalies dans les images WCE, ce CNN quantifie les couches convolutionnelles combinées en utilisant des

Chapitre 3 : Classification des images WCE en utilisant l'apprentissage profond

paramètres de poids entraînaables. Une autre approche basée sur le CNN pour la détection d'images anormales est proposée dans [95], où les patchs WCE ont été utilisés comme entrée du réseau. Les auteurs de [96] ont proposé un algorithme CNN pour classer cinq types de lésions dans les images WCE, et ils utilisent une stratégie de regroupement multi-échelle pour extraire plus d'informations sur les caractéristiques multi-niveaux. En raison de la faible quantité d'ensembles de données étiquetées disponibles publiquement, l'apprentissage par transfert a été choisi par de nombreux auteurs. Dans cette méthode, les CNN pré-entraînés extraient les caractéristiques de chaque image d'entrée au lieu d'entraîner un modèle à partir de zéro avec des poids initialisés de manière aléatoire. Cette méthode vise à transférer les informations d'un énorme ensemble de données (ImageNet) à un nouvel ensemble de données de petite taille. Des recherches approfondies sur les données d'imagerie médicale ont montré qu'un CNN pré-entraîné avec un réglage fin peut être plus performant qu'un CNN entraîné à partir de zéro [97]. Par conséquent, les auteurs de [98] ont proposé de combiner les caractéristiques DL produites par le modèle VGG19 et les méthodes d'analyse texturale telles que Haralick, pour la classification des maladies gastro-intestinales. Haya Alaskar et al. [99] ont utilisé deux modèles pré-entraînés, GoogLeNet et AlexNet, pour évaluer la performance de l'apprentissage par transfert pour la classification des ulcères dans les images de WCE. Caroppo et al. [100] ont proposé d'utiliser les couches intermédiaires des modèles pré-entraînés InceptionV3, ResNet50 et VGG16 pour extraire des caractéristiques. En utilisant l'approche de la pertinence maximale et de la redondance minimale, les caractéristiques ont été combinées et améliorées avant d'être introduites dans des classificateurs ML entraînés à classer les images de saignement. Rustam et al. [101] ont utilisé l'avantage de l'architecture MobileNet dans un CNN spécialement créé pour classer les images de saignement.

L'utilisation de la DL pour l'analyse des images de WCE a rencontré d'énormes difficultés en raison de la rareté des grands ensembles de données étiquetées [102], [103] et de la difficulté de la détection des images de WCE [104]. La plupart des propositions ci-dessous utilisent des ensembles de données personnels, ce que ne permet pas la comparaison avec leurs résultats. Heureusement, le jeu de données ouvert Kvasir-Capsule a été créé par des professionnels de la santé en 2020 et mis à jour en 2021 [105], qui contient 47238 images annotées et constitue la plus grande collection étiquetée d'anomalies. Cependant, l'ensemble Kvasir-Capsule n'est pas idéal et son utilisation nous confronte à certains défis. Le plus gros problème est celui du déséquilibre des données. Ainsi, certaines anomalies se produisent plus fréquemment que d'autres et certaines classes présentent un faible nombre d'images. Les créateurs de l'ensemble

considèrent le déséquilibre des données comme l'un des plus grands défis du domaine et encouragent les recherches pour éviter ses effets néfastes [105].

Récemment, de nombreux auteurs ont exploité Kvasir pour tester les modèles DL, L'auteur de [102] a sélectionné trois classes présentant des anomalies et les utilisées pour la classification binaire. Il a augmenté l'ensemble de données pour améliorer l'équilibre et a utilisé l'apprentissage par transfert avec trois modèles pré-entraînés. De même, l'utilisation de nombreuses classes a été envisagée par Smedsrud et al. [105] qui ont créé le jeu de données Kvasir-Capsule. Ils ont mené plusieurs expériences d'apprentissage par transfert, ils ont utilisé ResNet152 et DenseNet161. De plus, ils ont éliminé trois petites classes et n'ont utilisé que les onze classes restantes. De plus, L Bai et al. [106] ont proposé un réseau basé sur un transformateur qu'ils ont entraîné à partir de zéro sur le jeu de données de la capsule Kvasir. Ils ont également conservé la catégorie des découvertes lumineuses et éliminé les trois classes de la catégorie Anatomie.

3.3 Classification en utilisant l'apprentissage à partir de zéros

Dans cette méthode nous proposons une classification automatique des images WCE en utilisant un algorithme basé sur les CNNs pour les deux cas suivants :

- Classification binaire : normal ou anormal
- Classification multi classe : estomac, colon, intestin grêle

Le schéma fonctionnel de la méthodologie proposée dans cette étude est illustré sur la figure 3.1. Comme on peut le voir, la méthodologie est principalement divisée en cinq étapes :

- Collecte d'images de la WCE,
- Séparation des images en des classes spécifiques,
- Augmentation des données,
- L'entraînement du CNN proposé,
- Enfin l'évaluation du modèle.

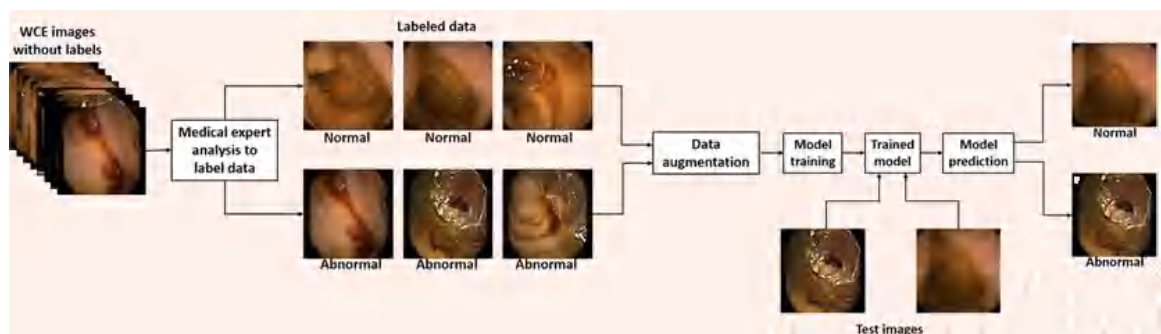


Figure 3.1 Schéma fonctionnel de la méthodologie proposée.

Chapitre 3 : Classification des images WCE en utilisant l'apprentissage profond

La structure globale de notre CNN illustré sur la figure 3.2 comporte trois couches de convolution, trois couches de mise en commun et deux couches entièrement connectées. L'entrée du réseau est de $276 \times 276 \times 3$, ce qui correspond à la taille de nos images WCE. Chaque couche de convolution contient un certain nombre de filtre (32, 64 et 128 respectivement) de taille 3×3 . Un max-pooling de 2×2 est appliqué après chaque couche de convolution. La fonction d'activation ReLU est utilisée pour toutes les couches, sauf la dernière, qui utilise la fonction Softmax qui nous permet d'obtenir les probabilités que l'entrée soit dans une classe particulière (classification). Le nombre de sorties est fixé en fonction du nombre de classes dans chacune des deux tâches de classification.

3.3.1 Base de données

Pour résoudre le problème de classification avec l'apprentissage profond, nous avons d'abord collecté des données pour entraîner les classificateurs. Nous avons pu collecter 19 courtes vidéos de différentes sources et 27 vidéos de l'hôpital Ben Badis de Constantine, Algérie. Ces données ont été annotées par un expert gastro-entérologue de ce même hôpital. Ces séquences ont été capturées avec différentes capsules PillCam, et toutes les images sont RVB et ont la même taille de 276×276 pixels.

Pour la première expérience (classification des lésions), le jeu de données comprend un total de 12024 images. Elles ont été divisées en deux classes : 4060 normales et 7964 anormales. Les images de classe anormales correspondent à différents types de lésions pathologiques (tumeur, saignement, anémie, ulcère, autre). Pour la deuxième expérience (tâche de classification anatomique), le jeu de données contient un total de 41722 images. Elles ont été divisées en trois

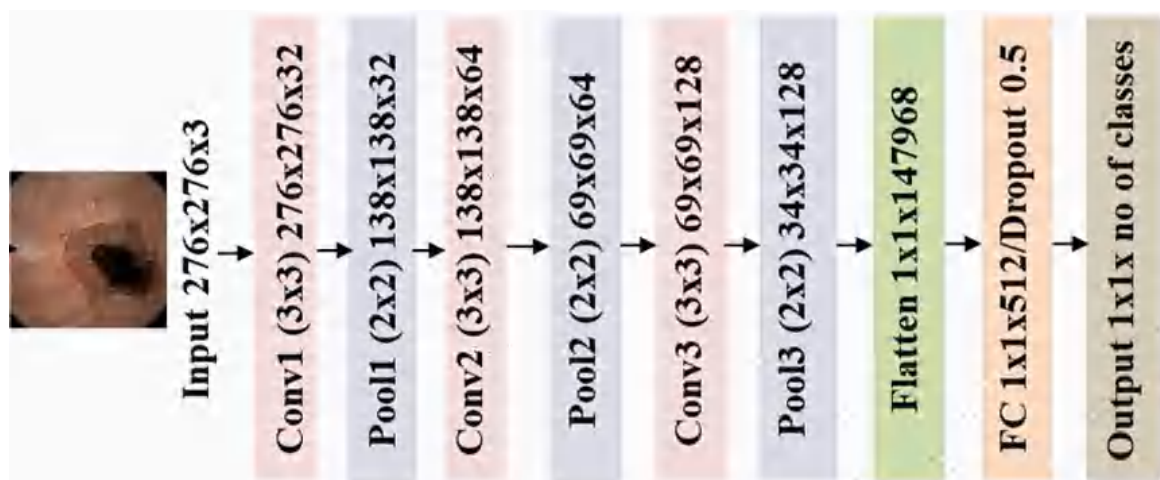


Figure 3.2 Schéma fonctionnel de la méthodologie proposée.

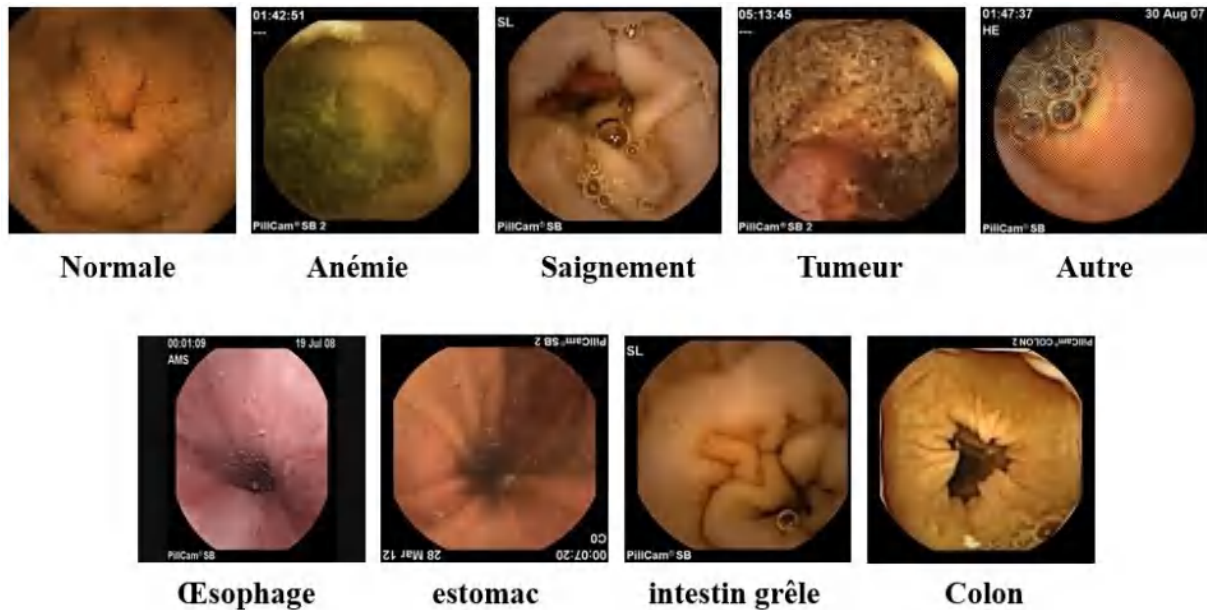


Figure 3.3 Exemple des images de notre base de données.

classes : 14640 de l'œsophage-estomac, 17317 de l'intestin grêle et 9765 du colon. Des exemples d'images de notre jeu de données sont présentés sur la figure 3.3.

3.3.2 Augmentation de données

Il est bien connu que les modèles d'apprentissage profond nécessitent une grande quantité de données d'entraînement, afin d'obtenir une bonne performance et éviter le sur-apprentissage. L'augmentation de données d'entraînement est une solution courante. Elle permet de rééquilibrer un jeu de données trop petit. Dans ce travail nous avons appliqué les techniques d'augmentation suivante :

•Rotation

Cette transformation, comme son nom l'indique, consiste à effectuer une rotation de l'image originale selon un angle souhaité. En analyse d'images médicales, l'augmentation d'un jeu de données avec des rotations aléatoires selon différents angles peut aller par exemple de 10° à 175° [107].

•Réflexions (en anglais flips)

Les réflexions d'images s'effectuent selon un axe de symétrie. L'inversion horizontale des lignes et des colonnes entières d'une image est appelée augmentation de retournement horizontal. Alors, L'inversion verticale des lignes et des colonnes entières d'une image s'appelle l'augmentation du retournement vertical.

•Contraste

Le contraste d'une image est augmenté lorsque les pixels foncés de cette image sont assombris et les pixels clairs sont éclaircis : Les variations de contraste effectuées pour l'augmentation des données à partir d'images médicales peuvent varier par exemple de 0.8 à 1.2, et sont souvent appliquées aléatoirement au jeu de données [108].

•Luminosité

La luminosité de l'image peut être augmentée pour devenir claire ou sombre en fonction de la plage de luminosité donnée, des variations de luminosité appartenant à l'intervalle $[-0.1; 0.1]$ peuvent par exemple être appliquées de manière aléatoire [108].

Un exemple d'augmentation d'une image de jeu de données est montré dans la figure 3.4. La figure 3.5 illustre le nombre des images avant et après l'augmentation.

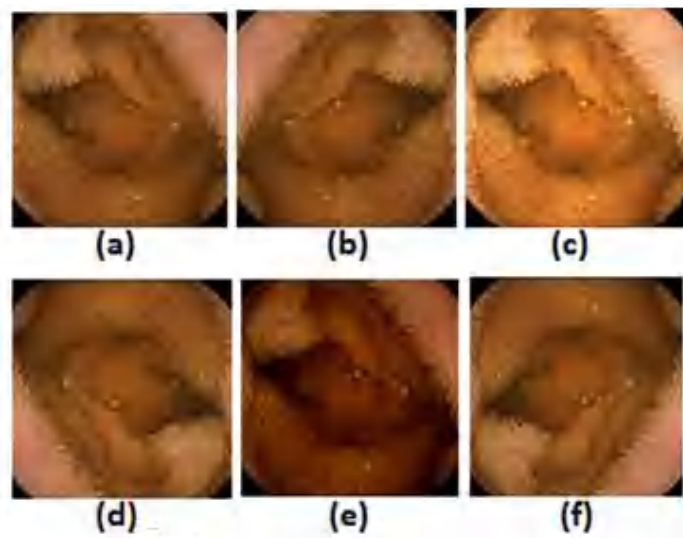


Figure 3.4 Résultat de l'application des techniques d'augmentation des données : (a) image originale, (b) retournement horizontal, (c) correction de la luminosité, (d) rotation, (e) ajustement du contraste, (f) retournement vertical.

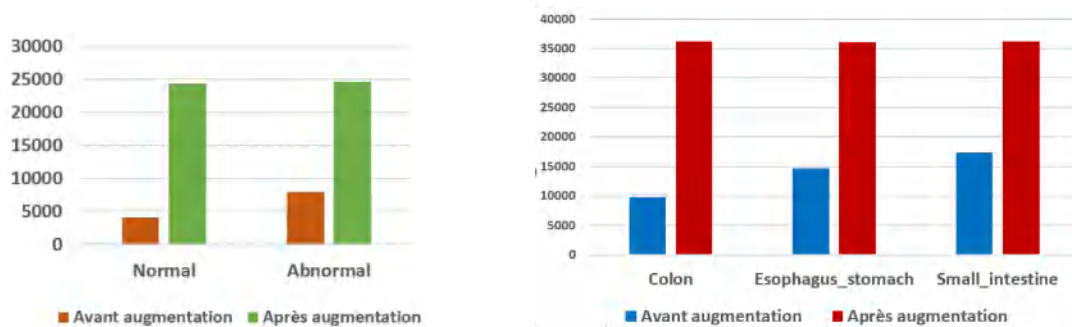


Figure 3.5 Equilibrage des données en utilisant les techniques d'augmentation des données.

3.3.3 La validation croisée

Afin d'obtenir des résultats significatifs dans nos expériences, nous avons appliqués une validation croisée à la base de données. Ce processus consiste à ne pas utiliser l'ensemble de test ou de validation, mais d'utiliser tour à tour différentes parties de l'ensemble d'apprentissage en tant qu'ensemble de test. Généralement, le choix du nombre de jeux de données dépend de la taille initiale de la base de données. Le nombre de jeux varie le plus souvent entre deux et dix. Dans notre cas on a divisé notre jeu de données en 5 folds ($K=5$), où chaque modèle est entraîné sur 4 partitions et évalué sur la partition restante. Ce processus est répété jusqu'à ce que chaque fold des cinq folds ait été utilisé comme groupe de test. Ce procédé est illustré sur la figure 3.6. La validation croisée amène un coût de calcul important, puisqu'il faut recommencer l'apprentissage chaque fois qu'une différente partie de l'ensemble d'apprentissage est utilisée comme ensemble de test.

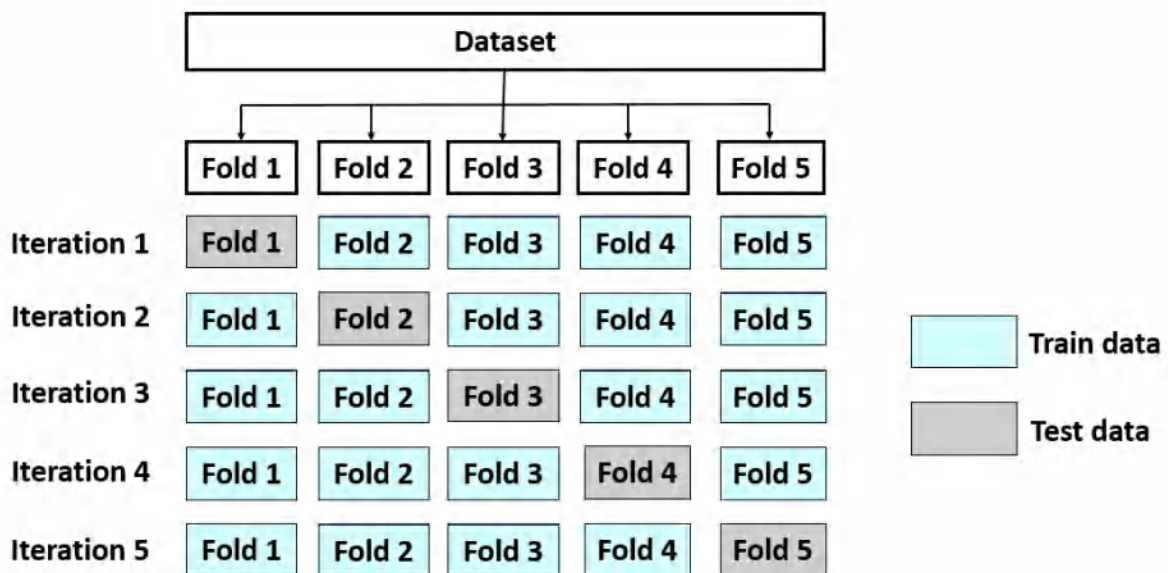


Figure 3.6 Validation croisée à 5 échantillons, dite 5-folds.

3.3.4 Les paramètres d'apprentissage utilisés :

Les poids du réseau ont été initialisés de manière aléatoire, la taille du lot est fixée à 32 images, et la couche entièrement connectée de 512 neurones est suivie d'une couche dropout avec une probabilité de 0,5 pour éviter le surapprentissage. Nous avons entraîné le modèle en utilisant l'optimiseur Adam avec un taux d'apprentissage initial fixé à 0,001. Enfin, l'apprentissage des paramètres du réseau est effectué sur 50 époques pour chaque fold. L'entraînement du réseau a été réalisé avec un ordinateur utilisant un microprocesseur Intel (R) core i7, 3,5 GHz, 64 Go de

Chapitre 3 : Classification des images WCE en utilisant l'apprentissage profond

Tableau 3.1 Résultats de la classification des lésions et des organes digestifs évalués sur 5 folds par validation croisée.

	<i>Paramètres d'évaluation</i>	<i>fold1</i>	<i>fold2</i>	<i>fold3</i>	<i>fold4</i>	<i>fold5</i>	<i>mean</i>	<i>std</i>
Classification des Lésions	<i>Accuracy(%)</i>	99.98	99.96	99.97	99.70	99.91	99.90	0.10
	<i>Précision(%)</i>	99.98	99.96	99.97	99.70	99.91	99.90	0.10
	<i>Sensitivité(%)</i>	100	99.94	100	99.97	100	99.98	0.02
	<i>Spécificité(%)</i>	99.97	99.97	99.94	99.43	99.84	99.83	0.20
	<i>MCC(%)</i>	99.97	99.92	99.94	99.41	99.82	99.81	0.20
Classification des organes digestifs	<i>Accuracy(%)</i>	99.70	99.64	99.64	99.51	99.12	99.52	0.21
	<i>Précision(%)</i>	99.70	99.65	99.64	99.52	99.12	99.52	0.21
	<i>Sensitivité(%)</i>	99.70	99.64	99.63	99.51	99.11	99.51	0.21
	<i>Spécificité(%)</i>	99.85	99.82	99.81	99.75	99.55	99.75	0.10
	<i>MCC(%)</i>	99.55	99.47	99.46	99.27	98.68	99.28	0.31

RAM et un processeur graphique (GPU) NVIDIA RTX 2080 TI. L'architecture du réseau est implémentée en Python à l'aide des packages Tensorflow et Keras.

3.3.5 Résultats et discussion

Nous avons analysé la performance de notre modèle de classification en utilisant cinq mesures d'évaluation, à savoir l'accuracy, la précision ou valeur prédictive positive (VPP), la sensibilité ou taux de vrais positifs (TVP), la spécificité ou taux de vrais négatifs (TVN) et le coefficient de corrélation de Matthews (MCC). (Voir section 2.7.1)

Les résultats de l'entraînement de chaque fold sont présentés dans le tableau 3.1.

D'après ces résultats, nous pouvons voir que notre modèle a obtenu de très bons résultats pour tous les folds avec une précision moyenne de 99.90% et 99.52% pour la classification des lésions et organes digestifs respectivement. La même observation pour toutes les mesures d'évaluation.

Pour une analyse plus approfondie des résultats obtenus, les matrices de confusion de chaque fold obtenues par le modèle proposé pour les deux classifications (lésions et anatomie) sont illustrées sur les figures 3.7 et 3.8. Les valeurs sur la diagonale principale représentent toutes les instances correctement classées.

Dans le cas du Lésions, l'ensemble de test contenait 9140 images, dont 4596 étaient abnormal et 4544 normal. L'ensemble de test des organes digestifs sont 18760 images, dont 6255 celles

Chapitre 3 : Classification des images WCE en utilisant l'apprentissage profond

du colon, 6248 du esophagus_stomach et 6257 images restantes appartiennent à la classe small_intestine.

D'après ces figures on peut voir que les FP et les FN sont relativement très faibles ce qui signifie une bonne performance du classificateur.

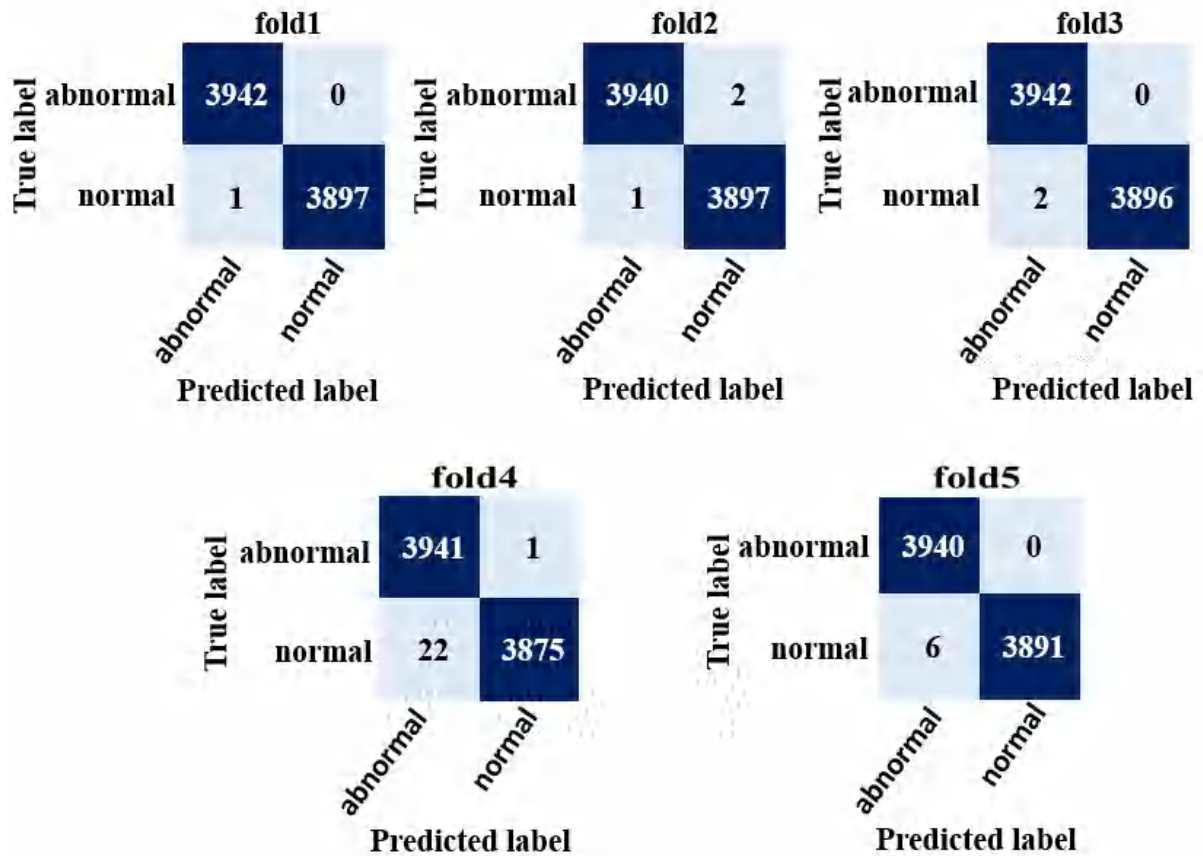


Figure 3.7 Matrice de confusion de la classification des lésions.

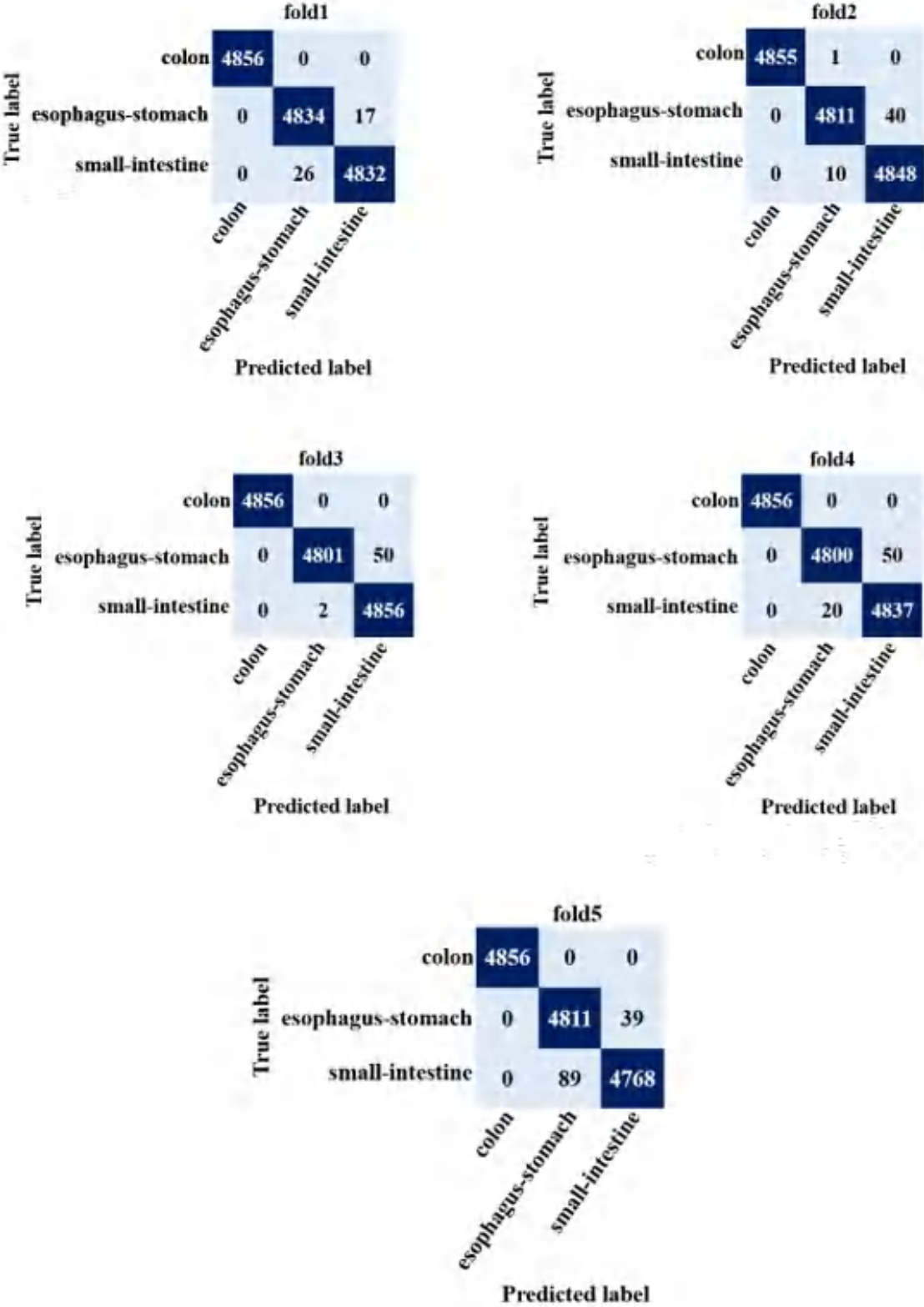


Figure 3.8 Matrice de confusion de la classification des organes digestifs.

Chapitre 3 : Classification des images WCE en utilisant l'apprentissage profond

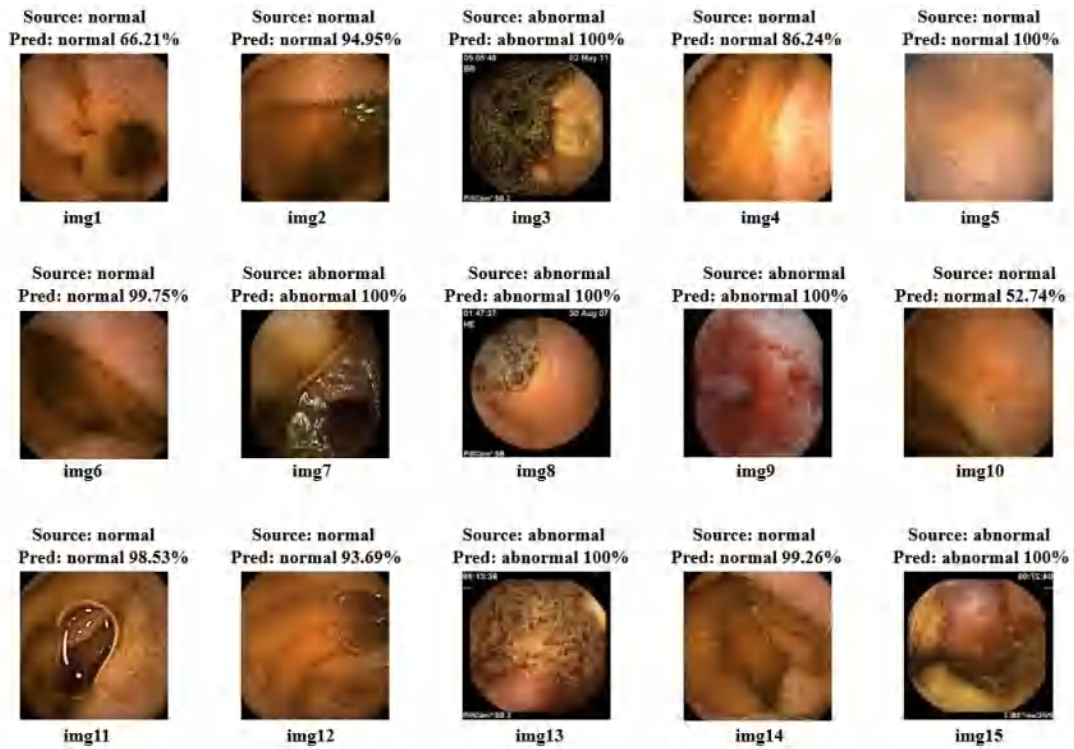


Figure 3.9 Résultats de prédiction obtenus par la classification des lésions.

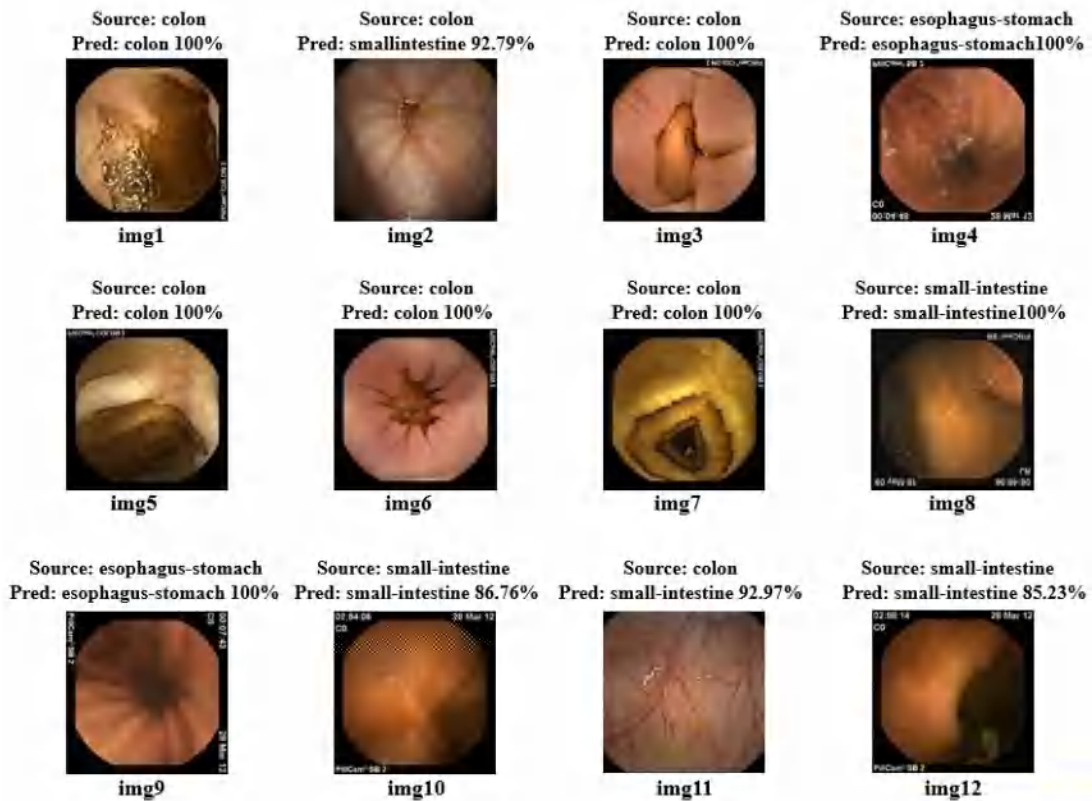


Figure 3.10 Résultats de prédiction obtenus par la classification des organes digestifs.

Chapitre 3 : Classification des images WCE en utilisant l'apprentissage profond

Les figures 3.9 et 3.10 représentent quelques exemples des résultats de prédiction obtenus par le modèle de classification proposé.

La figure 3.9 montre que la majorité des images (13 parmi 15 images) sont bien prédites avec une probabilité supérieure à 86%, ce qui signifie que ces résultats sont extrêmement positifs.

Il y a trois raisons pour une mauvaise classification pour les images WCE [109] :

- 1) l'image est sur ou sous-exposée, ou floue comme montré dans l'image 10 ;
- 2) l'existence de quelques impuretés ou bulles dans le tube digestif peut compliquer le contenu de l'image ;
- 3) S'il existe un petit nombre d'images anormales dans l'ensemble du jeu de données, telles que des images de saignement et d'ulcère, reflétant principalement la couleur, ces images sont plus susceptibles d'être mal classées.

Les résultats illustrés dans la figure 3.10 montrent que seule l'image n°11 est mal classée. Cette image est d'origine appartenant à la classe colon, mais elle a été prédite comme une image d'intestin grêle avec une probabilité de 92.79%. On peut voir que cette image n'est pas suffisamment claire, ce qui a conduit à une mauvaise classification.

3.4 Classification en utilisant la méthode d'apprentissage par transfert

Dans cette section nous utilisons la technique d'apprentissage par transfert pour les deux tâches de classification comme mentionnée précédemment dans la section 3.3.

Avec la stratégie d'apprentissage par transfert, le CNN pré-entraîné extraira les caractéristiques de chaque image d'entrée au lieu d'entraîner un modèle à partir de zéro avec des poids initialisés de façon aléatoire. Cette méthode vise à transférer les connaissances d'un grand ensemble de données (ImageNet) à un nouvel ensemble de données dans un autre domaine. Une étude approfondie sur des données d'imagerie médicale a démontré qu'un CNN pré-entraîné avec un réglage fin peut avoir des performances similaires ou supérieures à celles des CNN entraînés à partir de zéro [110].

L'apprentissage par transfert présente des avantages tels que l'accélération de la convergence du réseau, la réduction de la puissance de calcul et l'optimisation des performances du réseau [111].

Dans cette expérience, nous comparons la performance de onze réseaux profonds pré-entraînés telles que, Inceptionv3, Inceptionresnetv2, Densenet169, Densenet201, Densenet121, Vgg16, Vgg19, mobilenet, Mobilenetv2, ResNet et xception, afin d'identifier un modèle approprié pour la classification automatique des lésions et anatomique dans les images WCE. Le pipeline de l'expérience est présenté à la figure 3.11. Tout d'abord, la taille de l'ensemble de données est

augmentée en utilisant plusieurs techniques d'augmentation. Ensuite, toutes ces images sont redimensionnées selon la taille par défaut des réseaux pré-entraînés sélectionnés. Une fois les images d'entrée préparées, elles sont introduites dans la phase d'extraction des caractéristiques en utilisant les réseaux CNN pré-entraînés, les caractéristiques générées sont ensuite introduites dans un MLP pour classer chaque image dans les classes correspondantes. Nous avons également affiné tous les modèles en modifiant la dernière couche pour qu'elle produise les classes correspondantes (2 ou 3) à notre jeu de données au lieu de 1000 classes comme proposé pour ImageNet. Enfin, les performances de classification sont évaluées sur des images de test à l'aide des modèles entraînés. Dans cette expérience nous utilisons 60% de données pour l'entraînement, 20% pour la validation, tandis que les 20% restants ont été utilisés pour les tests.

3.4.1 Paramètres d'apprentissage

Tout d'abord, toutes les images sont redimensionnées selon la taille par défaut des réseaux pré-entraînés sélectionnés (224x224 pour DensNet, VggNet, mobileNet et ResNet, 299x299 pour les réseaux Inception, InceptionResnet et xception). La couche entièrement connectée a été entraînée avec la fonction d'activation (ReLU) a 512 neurones, De plus, une couche de Dropout avec une probabilité de 0,4 est appliquée, ce qui signifie que 40 % des neurones seront aléatoirement mis à zéro pendant chaque période d'apprentissage, évitant ainsi un surapprentissage de l'ensemble de données. La taille du lot est fixée à 32 images. Nous avons entraîné le modèle en utilisant l'optimiseur SGD avec un Momentum de 0.8 et un taux d'apprentissage initial fixé à 10^{-4} . Enfin, l'entraînement de tous les modèles est effectué sur 500 époques. Cette étude a utilisé 60% de données pour l'entraînement, 20% pour la validation, tandis que les 20% restants ont été utilisés pour les tests. Ce processus a été réalisé avec un ordinateur équipé d'un microprocesseur Intel (R) core i7, 3,5 GHz, 64 Go de RAM et d'un processeur graphique (GPU) NVIDIA RTX 2080 TI. L'architecture du réseau est mise en œuvre en Python à l'aide de package Keras avec backend Tensorflow.

3.4.2 Résultats et discussion

Nous avons analysé les performances de la méthode proposée en utilisant les quatre mesures d'évaluation, à savoir l'accuracy, la précision, le rappel et le f1-score (voir section 2.7.1).

Les résultats de classification des lésions sur des images de test sont résumés dans le tableau 3.2. Les résultats obtenus sur ce tableau montrent que l'architecture Densenet201 fourni les meilleurs résultats avec une précision de 99.65%. Tandis que les résultats les plus faibles sont obtenus par l'architecture Resnet50 avec seulement 89.30% de précision. Cela signifie que

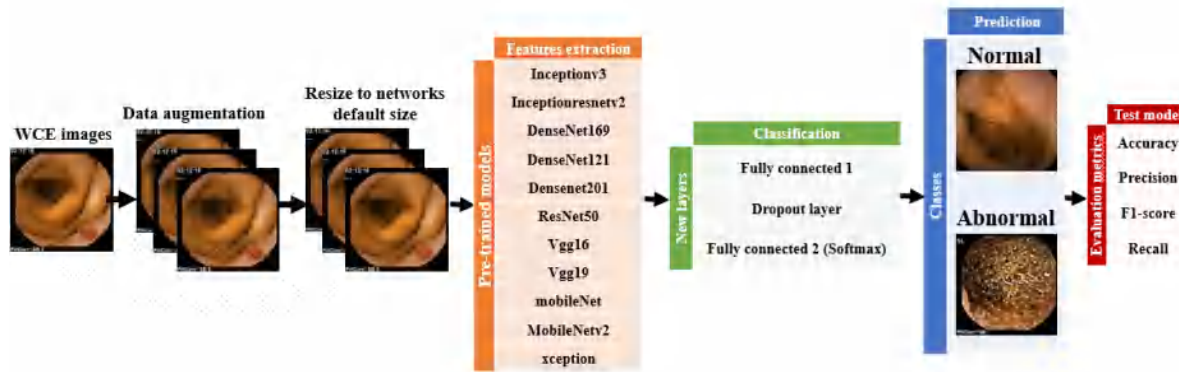


Figure 3.11 Organigramme de l'expérience globale.

l'écart de précision est de 10.35% en faveur de Densenet201 par rapport à ResNet50. On peut voir aussi que les résultats des autres modèles comme Inceptionv3, Inceptionresnetv2, Densnet169, Densenet121, mobilnet, Mobilnetv2, xception sont également très bonnes, ils ont dépassé 98.90% de précision. Si l'on compare les performances du modèle DenseNet201 à celles de VGG16 et de VGG19, l'écart de précision est de 4.73% et 6.25% respectivement. D'après ces résultats, on observe que DenseNet201 est le modèle le plus efficace pour classer les images des lésions dans le tube digestif.

Tableau 3.2 Performance de la classification de lésions obtenues par l'apprentissage par transfert.

CNNs	Performance			
	Accuracy	Précision	Rappel	F1-score
<i>Inceptionv3</i>	99.41	99.41	99.41	99.41
<i>Inceptionresnetv2</i>	99.52	99.52	99.52	99.52
<i>Densnet169</i>	99.44	99.45	99.44	99.44
<i>Densenet121</i>	99.58	99.59	99.58	99.58
<i>Densenet201</i>	99.65	99.65	99.65	99.65
<i>Resnet50</i>	89.16	89.47	89.16	89.14
<i>Vgg16</i>	94.92	95.04	94.92	94.92
<i>Vgg19</i>	93.40	93.41	93.40	93.40
<i>mobilnet</i>	99.47	99.48	99.47	99.47
<i>Mobilnetv2</i>	98.81	98.83	98.81	98.81
<i>xception</i>	99.11	99.12	99.11	99.11

Chapitre 3 : Classification des images WCE en utilisant l'apprentissage profond

En plus, les résultats obtenus pour les onze modèles sont présentés sous forme de matrices de confusion dans les figures 3.12. La ligne sous chaque matrice de confusion montre la précision atteinte et le taux d'erreur de prédictions.

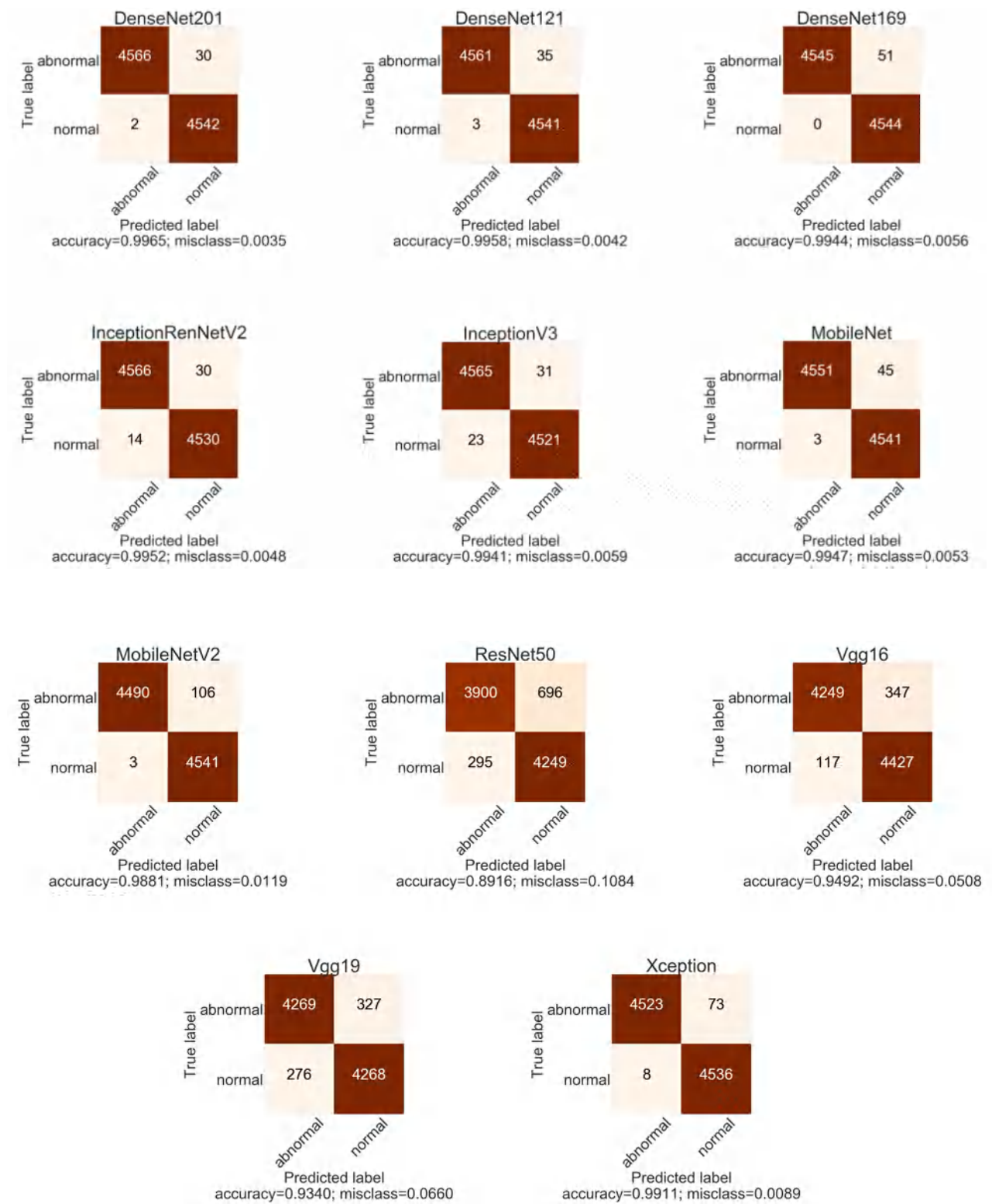


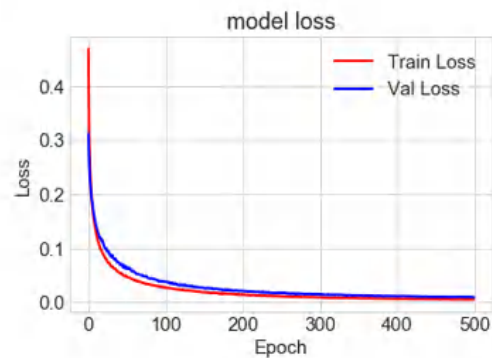
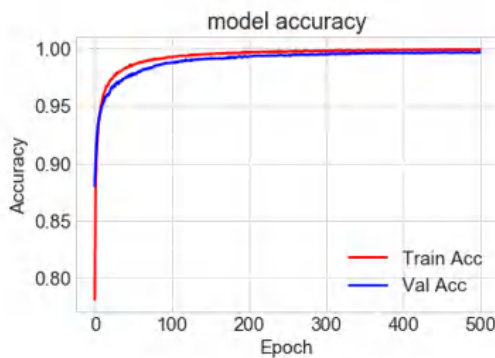
Figure 3.12 Matrice de confusion de la classification de lésions obtenues par différents modèles DCNN.

Chapitre 3 : Classification des images WCE en utilisant l'apprentissage profond

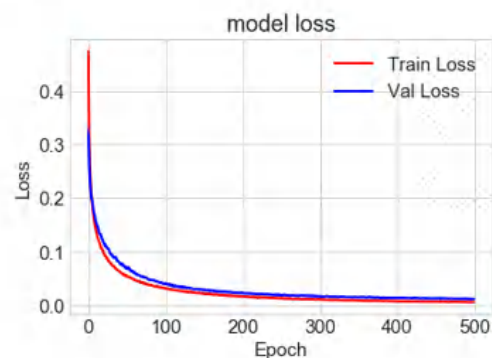
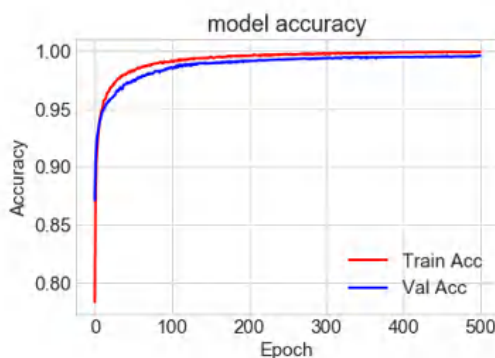
Dans le cas de classification des lésions, DenseNet201 a démontré le meilleur résultat en classifiant correctement 9108 échantillons, tandis que ResNet50 a montré une mauvaise performance en classifiant 8149 échantillons sur 9140.

La figure 3.13 montre les courbes de perte/précision d'entraînement/validation pour chaque CNN. On observe que les courbes d'apprentissage des architectures : Inceptionv3, Inceptionresnetv2, Densenet201, Densnet169, Densenet121, Mobilenet, Mobilnetv2 et Xception sont idéal. On peut voir que la perte diminue et la précision augmente progressivement. De plus, la précision d'entraînement et de validation ne s'écarte pas beaucoup l'une de l'autre dans la plupart des cas, un phénomène qui peut également être observé pour la perte d'entraînement et de validation, indiquant que les modèles ne sont pas sur-apprendre. D'autre part, le modèle Resnet50 présente non seulement une faible précision, mais leur perte de validation augmente dans certain cas, la perte de validation fluctuante signifie que le modèle est très probablement sur-apprendre. La précision et la perte d'entraînement et de validation pour les modèles vgg16 et VGG19 s'écarte à partir de l'époque 25. Ces modèles comportent un petit nombre de couches, cela peut s'expliquer par le fait que les réseaux neuronaux comportant un plus grand nombre de couches cachées avec un nombre important de données obtiennent une meilleure performance de classification.

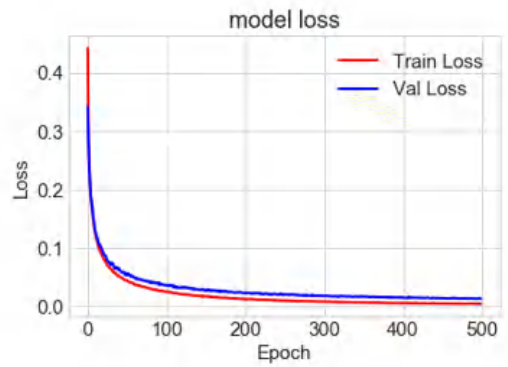
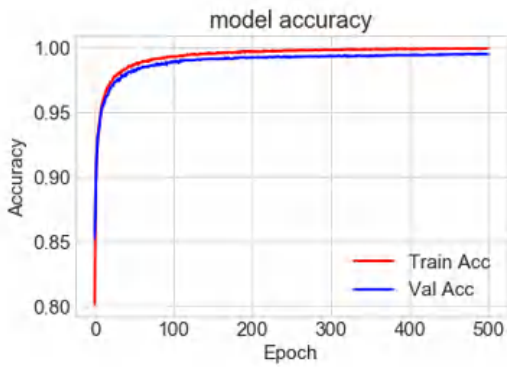
DenseNet201



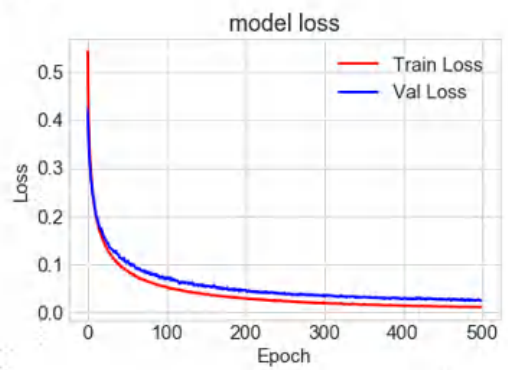
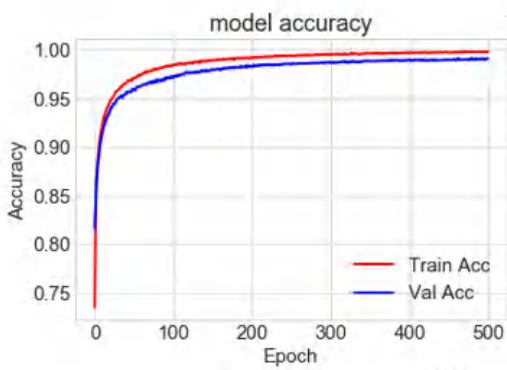
DenseNet121



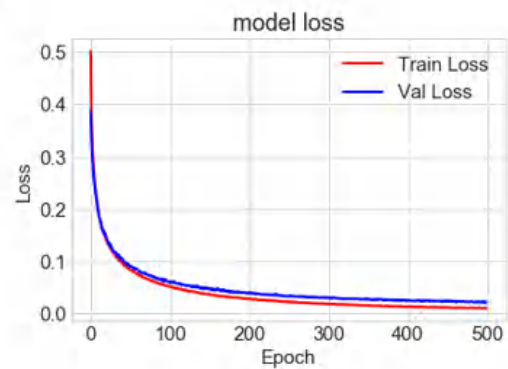
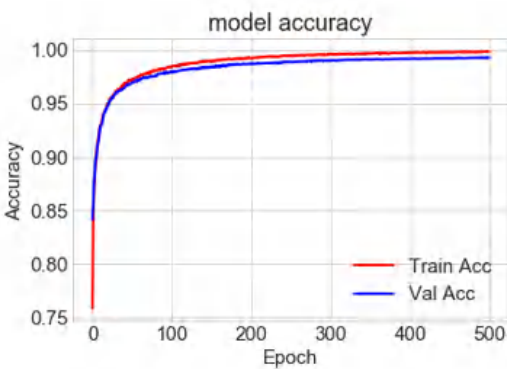
DenseNet169



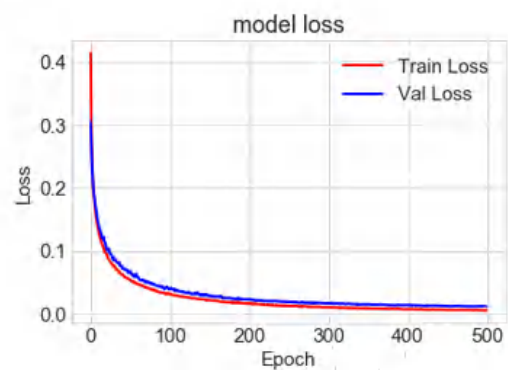
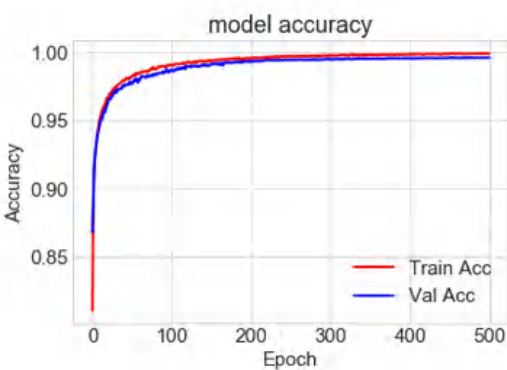
InceptionResNetV2



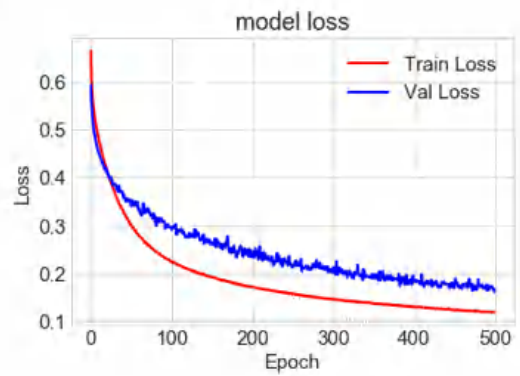
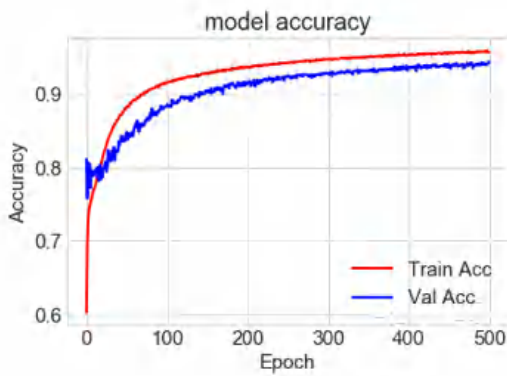
InceptionV3



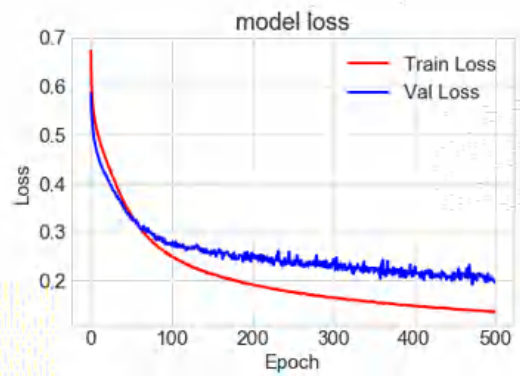
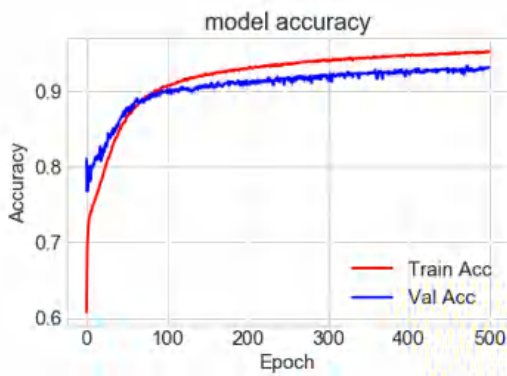
MobileNet



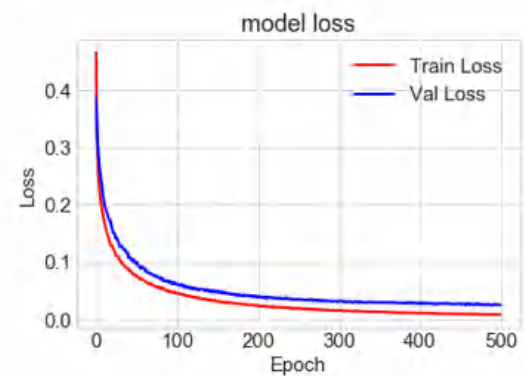
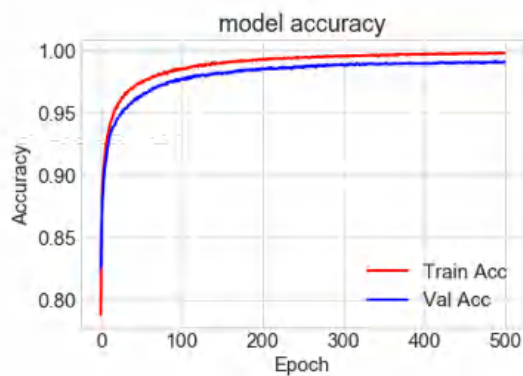
Vgg16



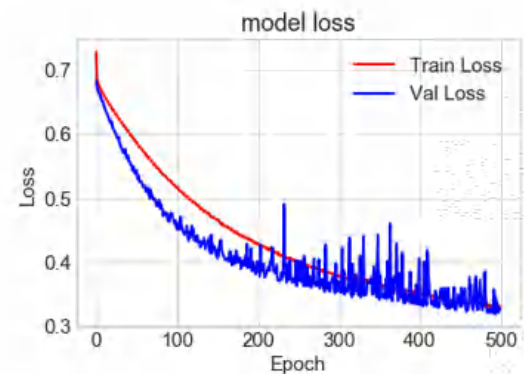
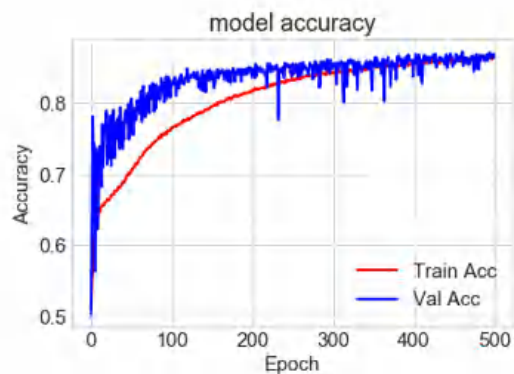
Vgg19



MobileNetV2



ResNet50



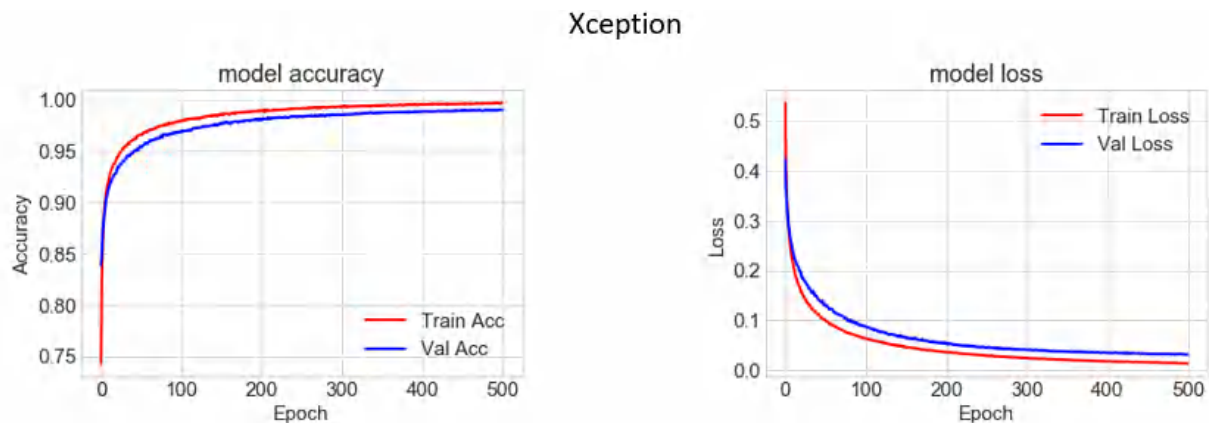


Figure 3. 13 Courbes de précision et de perte de différents modèles pour la classification des lésions.

Le tableau 3.3 montre les résultats de la classification des organes digestifs sur l'ensemble de test. On peut voir aussi que la performance des différents CNNs est élevée avec un maximum de 99.63% de précision obtenu par Dnesenet121. Des résultats similaires sont obtenus par les architectures Dnesenet169 et Dnesenet201 avec une précision de 99.62%. On observe que le réseau resnet50 représente toujours les mauvais résultats avec une précision de 84.57%.

Tableau 3.3 Performance de la classification des organes digestifs obtenus par l'apprentissage par transfert.

CNNs	Performance			
	Accuracy	Précision	Rappel	F1-score
<i>Inceptionv3</i>	99.52	99.52	99.52	99.52
<i>Inceptionresnetv2</i>	99.58	99.58	99.58	99.58
<i>Densnet169</i>	99.62	99.62	99.62	99.62
<i>Densenet121</i>	99.63	99.63	99.63	99.63
<i>Densenet201</i>	99.62	99.62	99.62	99.62
<i>Resnet50</i>	84.57	87.50	84.57	84.33
<i>Vgg16</i>	97.58	97.63	97.58	97.58
<i>Vgg19</i>	94.32	94.99	94.32	94.29
<i>mobilenet</i>	99.50	99.51	99.50	99.50
<i>Mobilenetv2</i>	99.21	99.21	99.21	99.21
<i>xception</i>	99.34	99.34	99.34	99.34

Chapitre 3 : Classification des images WCE en utilisant l'apprentissage profond

Les matrices de confusions pour chaque architecture sont illustrées sur la figure 3.14.

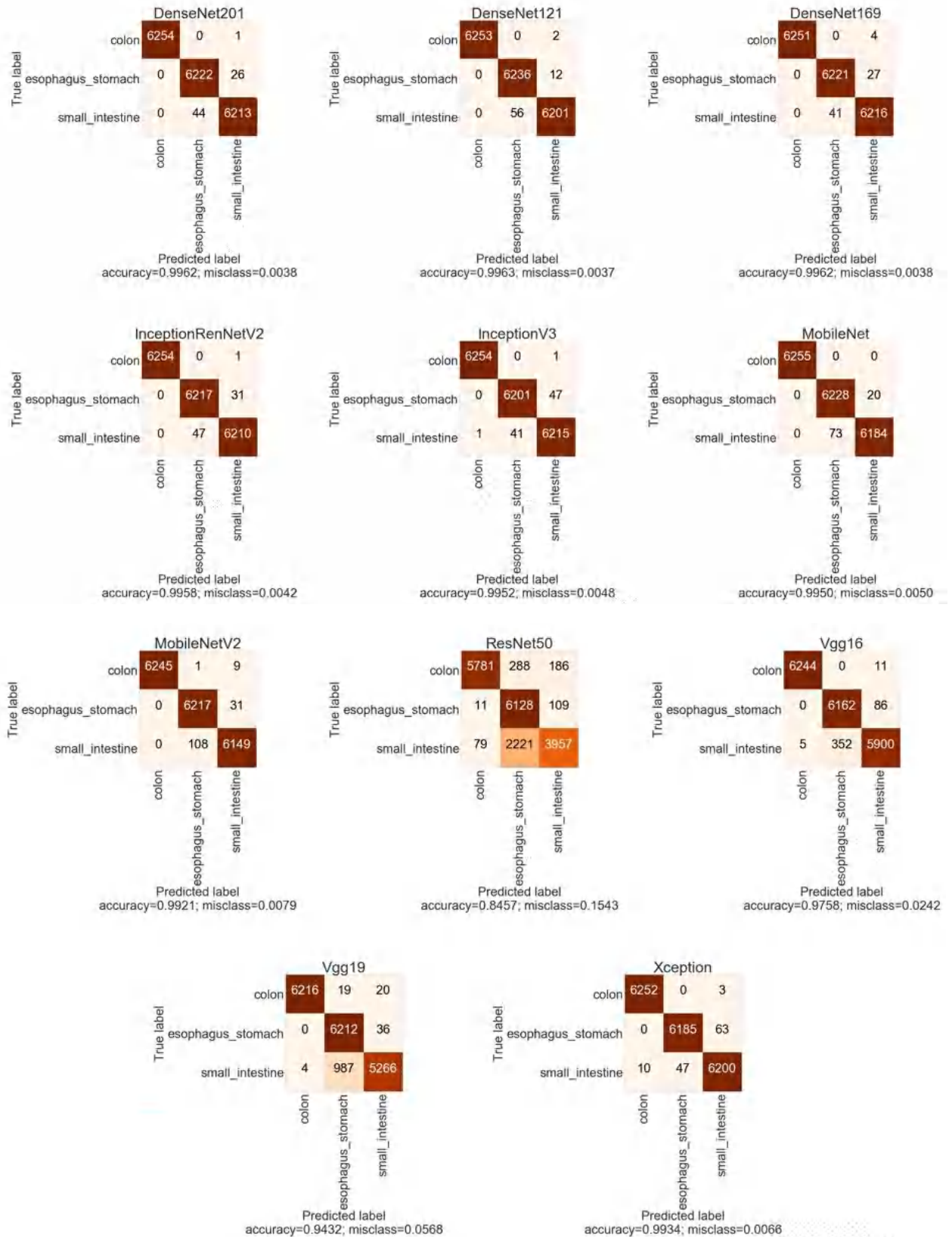


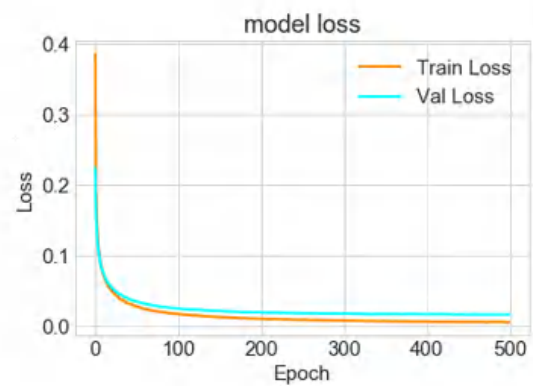
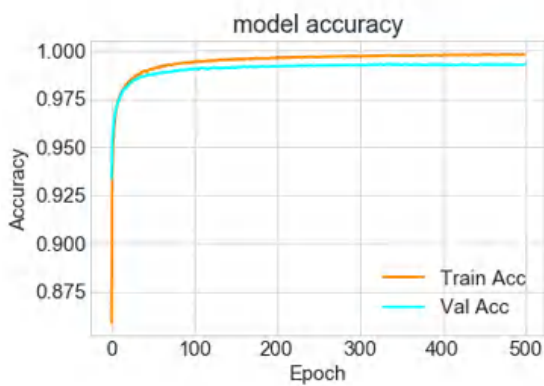
Figure 3.14 Matrice de confusion de la classification des organes digestifs obtenus par différents modèles DCNN

Chapitre 3 : Classification des images WCE en utilisant l'apprentissage profond

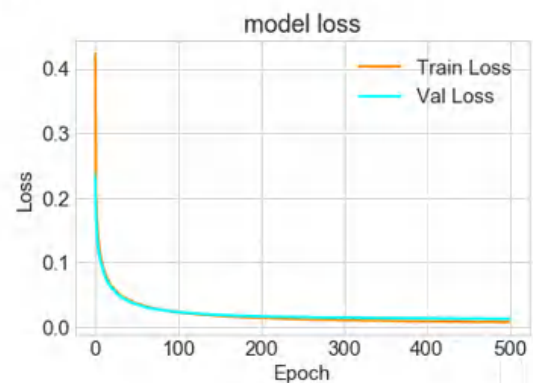
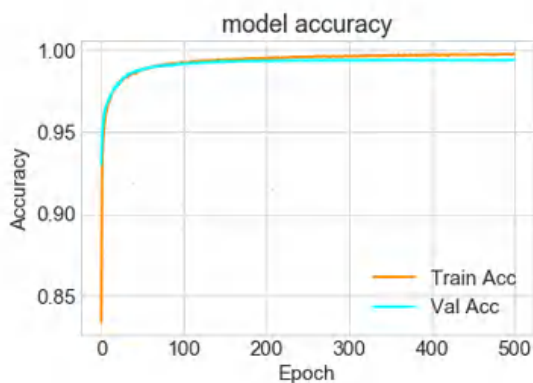
Les prédictions faussement négatives et faussement positives pour la classification des organes digestif étaient très faibles avec un taux d'erreur de 0.37%, 0.38% et 0.38% pour les architectures DenseNet121, DenseNet201 et DenseNet169, respectivement. L'architecture ResNet50 donne le mauvais résultat avec un taux d'erreur de 15.43%.

La figure 3.15 montre les courbes de perte/précision d'entraînement/validation pour chaque CNN. On observe que les courbes d'apprentissage des architectures : Inceptionv3, Inceptionresnetv2, Densenet201, Densnet169, Densenet121, Mobilenet, Mobilnetv2, Vgg16, Vgg19 et Xception sont idéals. Ces courbes sont convergés doucement et régulièrement vers des poids. Par conséquent, le phénomène de surapprentissage n'est pas apparu au cours d'apprentissage des CNNs. D'autre part, la précision et la perte d'entraînement /validation pour le modèle Resnet50 s'écarte et ne converge pas beaucoup pendant les 500 époques.

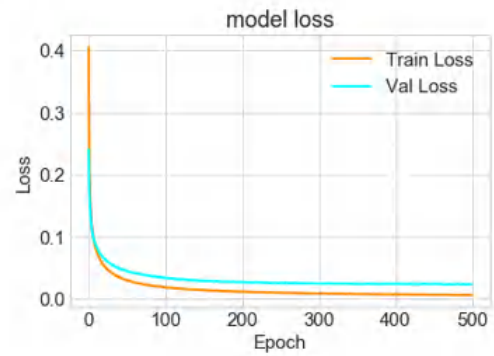
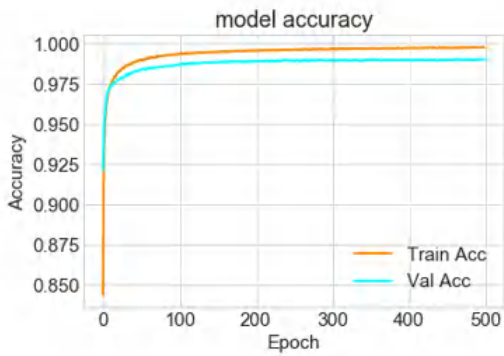
DenseNet201



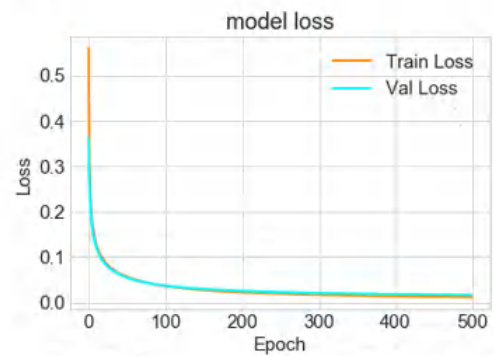
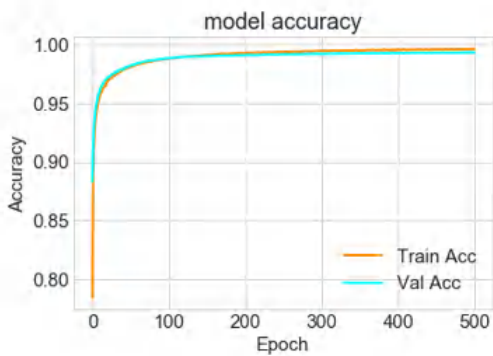
DenseNet121



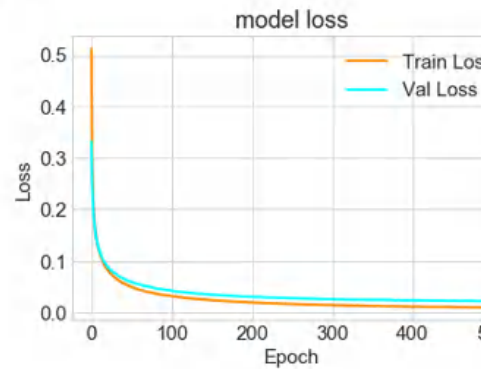
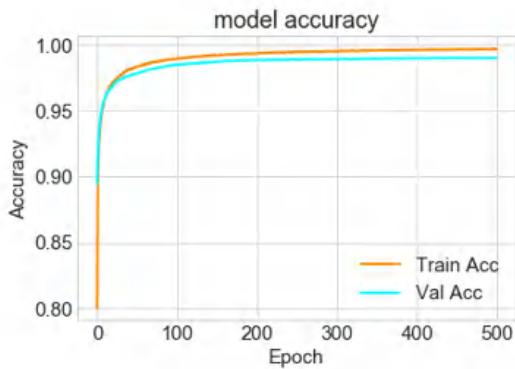
DenseNet169



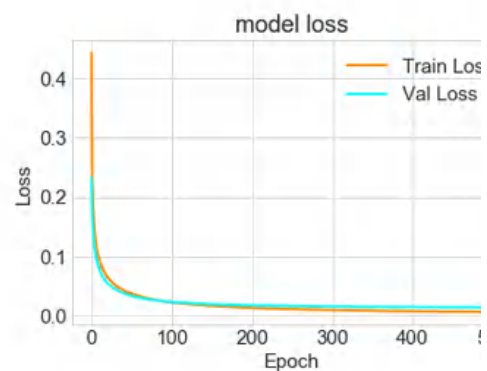
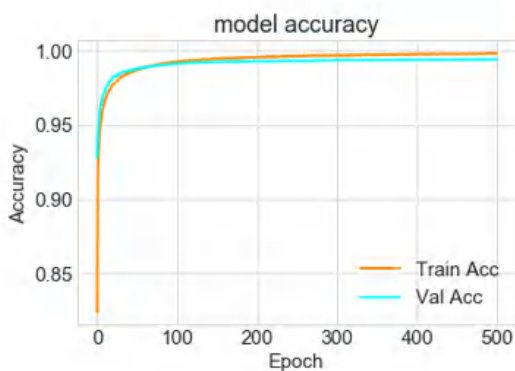
InceptionResNetV2



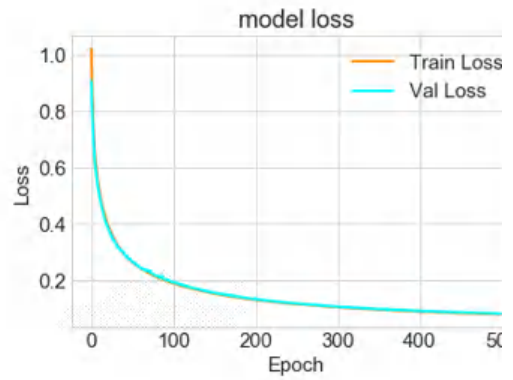
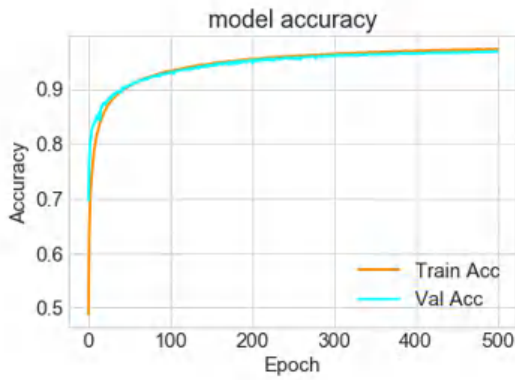
InceptionV3



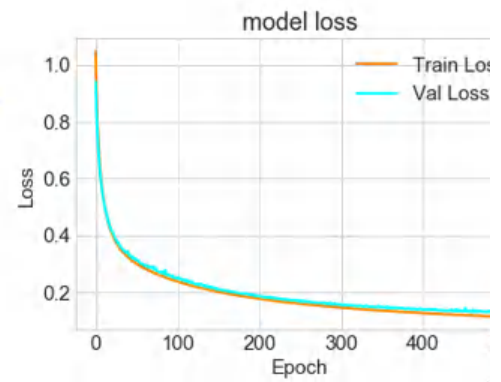
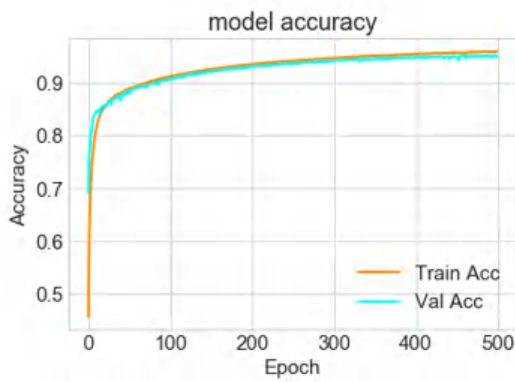
MobileNet



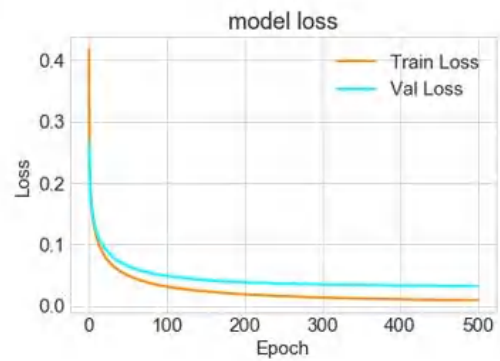
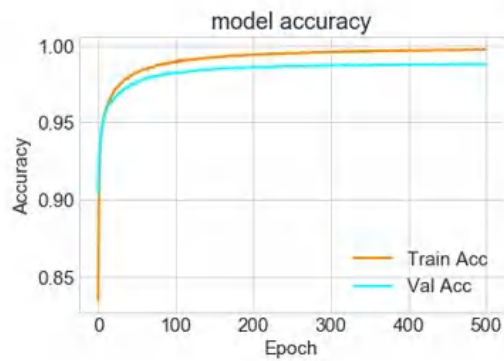
Vgg16



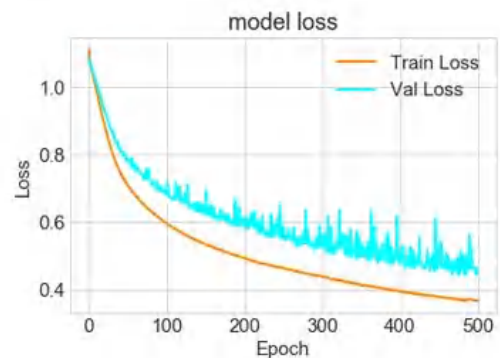
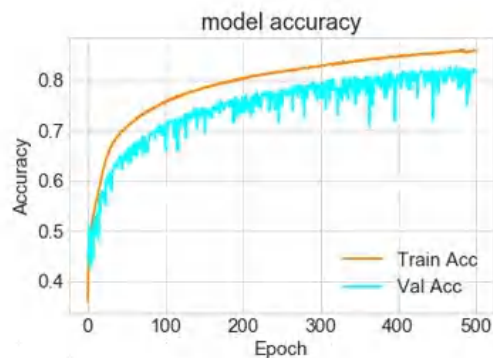
Vgg19



MobileNetV2



ResNet50



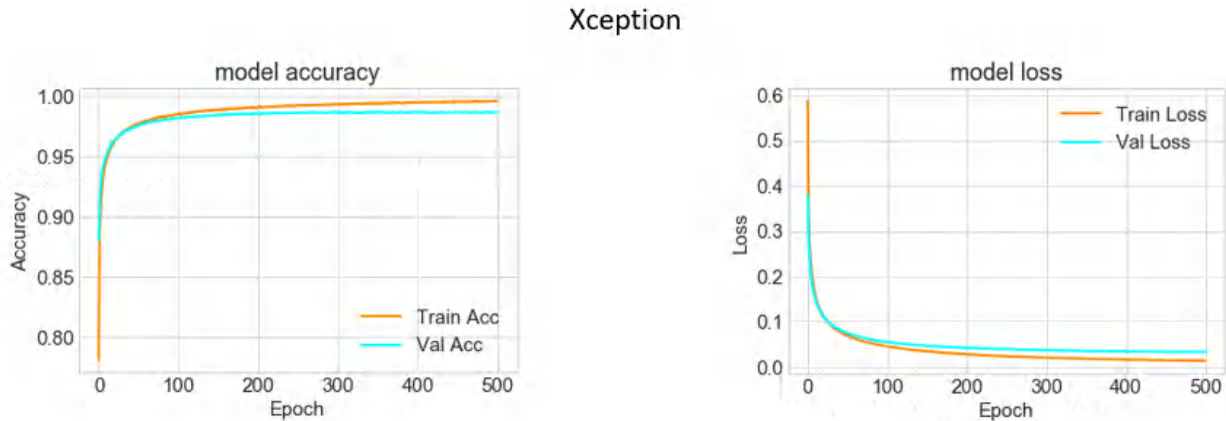


Figure 3. 15 Courbes de précision et de perte de différents modèles pour la classification des organes digestifs.

3.5 Classification en utilisant la concaténation de caractéristiques

Dans cette section nous proposons une méthode d'apprentissage en profondeur basée sur la combinaison de deux réseaux de neurones convolutifs (CNN) pré-entraînés, afin de détecter les différents types des lésions des images WCE.

Le schéma fonctionnel de la méthodologie proposée de cette étude est illustré sur la figure 3.16. Comme on peut le voir, la méthodologie est principalement divisée en cinq étapes : collecte d'images WCE, augmentation des données pour rééquilibrer le jeu de données, extraction de caractéristiques à l'aide du réseau proposé, l'entraînement du modèle et enfin l'évaluation du modèle sur des images de test.

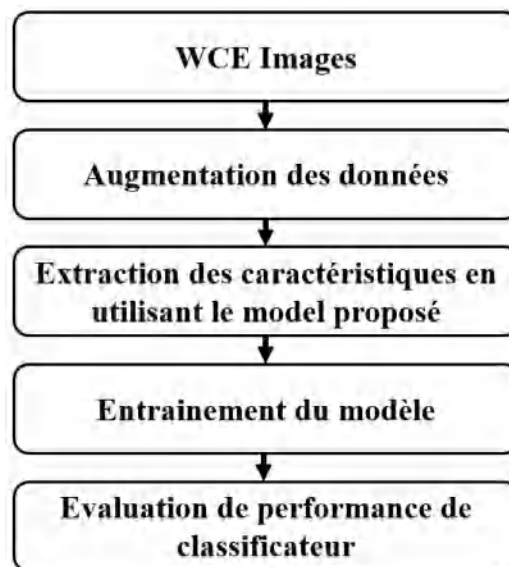


Figure 3.16 Schéma fonctionnel de la méthode proposée.

3.5.1 Description de la méthode proposée pour la classification des maladies GI

Dans cette expérience, nous avons utilisé la méthode d'apprentissage par transfert pour classer les différentes lésions des images WCE. Nous utilisons deux CNN pré-entraînés : MobileNet et DenseNet169. Pour sélectionner ces deux réseaux, nous avons testé de nombreuses combinaisons d'architectures y compris, MobileNetV2, Xception, InceptionResNetV2, DenseNet201, DenseNet121 et InceptionV3. Nous avons également testé diverses combinaisons d'hyperparamètres, tels que le taux d'apprentissage, dropout, taille de lot et optimiseurs pour obtenir les meilleurs résultats pour la classification des anomalies dans WCE. La figure 3.17 illustre l'architecture proposée. Dans la première étape, chaque image WCE est introduite dans la phase d'extraction de caractéristiques en utilisant le CNN proposés. Ensuite, les caractéristiques extraites de chaque architecture sont concaténées et aplaties ensemble pour créer le vecteur de caractéristique final. Ce vecteur de caractéristiques est utilisé comme entrée

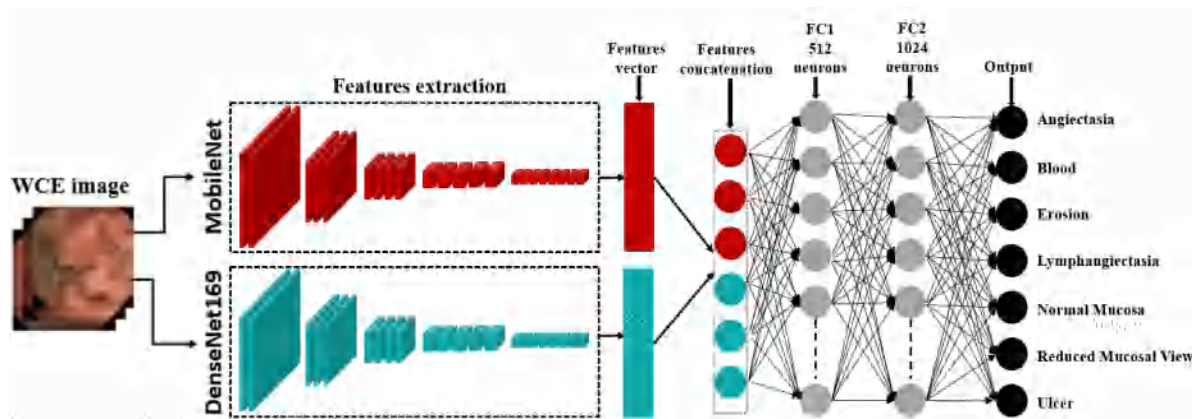


Figure 3.17 Schéma de l'architecture proposée

à un classificateur de type MLP, qui contient deux couches entièrement connectées de 512 et 1024 neurones respectivement. La combinaison des deux réseaux permet d'obtenir plus de caractéristiques. Par conséquent, des meilleurs résultats par rapport au classificateur individuel. Nous avons également affiné le modèle en changeant la dernière couche pour produire 7 classes selon notre jeu de données au lieu des 1000 classes fournies pour ImageNet.

3.5.2 Base de données

Dans cette expérience, nous avons utilisé l'ensemble de données Kvasir-capsule [5], qui est disponible publiquement. Il a été créé par des professionnels de la santé en 2020 et mis à jour en 2021. Ce jeu de données contient 47238 images étiquetées divisées en 14 classes organisé en deux catégories principales, qui sont : les images anatomiques et les images luminales. Toutes ces images sont RGB et ont une résolution de 336x336 pixels.

Chapitre 3 : Classification des images WCE en utilisant l'apprentissage profond

Le tableau 3.4 montre les différentes 14 classes représentant les images étiquetées et le nombre d'images dans chaque classe.

La catégorie Anatomie : elle contient les repères anatomiques caractérisant le tractus gastro-intestinal. Ces repères peuvent être utilisés pour l'orientation pendant les procédures endoscopiques. Cependant, pour la capsule de l'intestin grêle, leur rôle est de vérifier le passage de la capsule à travers tout l'intestin grêle pour conforter un examen complet. Smedsrud et al. [105] ont marqué trois repères anatomiques, les deux premiers délimitent respectivement l'extrémité supérieure (proximale) et inférieure (distale) de l'intestin grêle. Le **pylorus** est la jonction anatomique entre l'estomac et l'intestin grêle et est un sphincter (muscle circulaire) régulant la vidange de l'estomac dans le duodénum. La **ileocecal valve** marque la transition entre l'intestin grêle et le gros intestin et est une valve empêchant le reflux du contenu colique, les selles, vers l'intestin grêle. La troisième, **ampulla of Vater**, est la jonction entre le duodénum et le canal biliaire.

La catégorie **luminale** : Les examens endoscopiques peuvent détecter différents résultats lumineux, y compris les sous-catégories contenues de la lumière intestinale, aspect de la

Tableau 3. 4 Le nombre d'images dans les différentes classes pour les jeux de données Kvasir-Capsule.

<i>Type de catégorie</i>	<i>Classe</i>	<i>Nombre d'images</i>
<i>Anatomy</i>	<i>Pylorus</i>	1529
	<i>Ampulla of vater</i>	10
	<i>Ileocecal valve</i>	4198
<i>Luminal</i>	<i>Normal clean mucosa</i>	34338
	<i>Reduced mucosal view</i>	2906
	<i>Lymphangiectasia</i>	592
	<i>Erythema</i>	159
	<i>Angiectasia</i>	866
	<i>Blood-fresh</i>	446
	<i>Blood-hematin</i>	12
	<i>Erosion</i>	506
	<i>Ulcer</i>	854
	<i>Polyp</i>	55
	<i>Foreign body</i>	776

Chapitre 3 : Classification des images WCE en utilisant l'apprentissage profond

muqueuse et lésions muqueuses (résultats pathologiques) qui peuvent être soit grasses, surélevées ou excavées. Ces sous-catégories ne sont pas représentées dans l'ensemble de données. Normalement, l'intestin grêle ne contient qu'une certaine quantité de liquide jaune ou brun considéré comme une muqueuse propre. Cependant, de plus grandes quantités de contenu peuvent empêcher une visualisation complète de la muqueuse, cruciale pour vérifier une muqueuse normale et détecter tous les résultats pathologiques (anormaux). Pour l'évaluation du contenu de la lumière, Smedsrud et al. [105] ont défini cinq classes. **La Normal clean mucosa** et propre représente un intestin grêle propre, sans ou avec une petite quantité de fluide, et une muqueuse avec des villosités saines et sans signes pathologiques. Cette classe peut également servir de classe "normale" par rapport à la classe des anomalies lumineuses pathologiques (voir ci-dessous). La classe **reduced mucosal view** montre un contenu de l'intestin grêle réduisant la vue de la muqueuse, comme des selles ou des bulles. Cependant, les lésions du tube digestif supérieur ou de l'intestin grêle peuvent saigner, provoquant l'apparition de **blood - fresh** colorant le liquide en rouge. Dans les cas où le saignement est minime, on peut observer de petites bandes noires appelées **blood - hematin** à la surface de la muqueuse. La catégorie des **foreign body** comprend les résidus de comprimés ou les capsules retenues qui peuvent également être observés dans la lumière.

Les anomalies, appelées lésions ou résultats pathologiques, dans l'intestin grêle se manifestent par des modifications de la surface de la muqueuse. Les modifications typiques de la muqueuse couvrent parfois des segments plus importants, comme un aspect rougeâtre appelé erythema. La paroi muqueuse peut également présenter différentes lésions focales. Les classes de lésions représentées dans l'ensemble de données Kvasir-Capsule sont les **angiectasias** ; de petits vaisseaux superficiels dilatés provoquant des saignements chroniques et par la suite une anémie. Les lésions excavées érodent à des degrés divers la surface de la muqueuse. Les plus courantes sont les **erosions**, recouvertes d'une minuscule couche de fibrin, tandis que les érosions plus importantes sont appelées **ulcers**. Par exemple, la maladie de Crohn est une inflammation chronique de l'intestin grêle caractérisée par des ulcères et des érosions de la muqueuse. La **Lymphangiectasia**, qui représente des vaisseaux lymphoïdes dilatés dans la paroi muqueuse, et les **polyps**, qui peuvent être des lésions précancéreuses, sont visibles comme faisant saillie de la paroi muqueuse.

Toutes les différentes classes étiquetées sont représentées à la figure 3.18.

Puisque notre étude a pour but de détecter les différentes lésions dans le tractus gastro-intestinal, nous n'avons utilisé que les classes de la catégorie luminale, où



Figure 3.18 Exemples d'images des ensembles de données Kvasir-Capsule.

sept classes sont sélectionnées pour la tâche de classification à savoir ; Ulcer, Angiectasia, Blood, Erosion, Lymphangiectasia, Normal Mucosa et Reduced Mucosal View. La classe « foreign body » n'est pas incluse dans cette étude car elle n'est ni une pathologie ni une classe normale. Nous avons également supprimé les classes de polyp, blood-hematin, et erythema, en raison du petit nombre d'images.

3.5.3 Augmentation de données

Comme observé dans le tableau 3.4, le nombre d'images par classe n'est pas équilibré. Il s'agit d'un défi global dans le domaine médical car certaines images sont plus courantes que d'autres, on peut voir que La classe normal mucosa représente presque 78% du jeu de données global. Dans le but d'améliorer la performance et de résoudre le problème des jeux de données déséquilibrés, nous avons sélectionné seulement 3122 images aléatoires de classe normal mucosa et nous avons augmenté le nombre d'images dans les autres classes. Pour cela, nous avons appliquées plusieurs techniques d'augmentation de données, telles que la rotation aléatoire d'images, le retournement horizontal et vertical, la modification du contraste et de la luminosité, pour générer plus d'ensembles de données d'entraînement et de validation. Nous

avons utilisé ces techniques pour toutes les classes, sauf la classe normal mucosa. De plus, un nombre égal d'images (3122) de chaque classe est utilisé pour équilibrer l'ensemble de données.

3.5.4 Paramètres d'apprentissage

Tout d'abord, toutes les images sont redimensionnées à 224x224 selon la taille par défaut des réseaux pré-entraînés sélectionnés (Densenet169 et MobileNet). En outre, une couche de dropout est également ajoutée avec une probabilité de 0,25 après chaque couche entièrement connectée, pour éviter le sur-apprentissage. Nous avons utilisé l'algorithme de descente de gradient stochastique (SGD) pour entraîner le modèle, avec un taux d'apprentissage initial de 10^{-4} et une momentum de 0,8. La taille du lot est limitée à 32 images. Enfin, 300 époques sont utilisées pour entraîner le modèle. Dans cette expérience, le processus d'entraînement a été effectué en utilisant 70 % de données, la validation avec 10 % et le test a été effectué avec les 20 % restants. Ce processus a été réalisé avec un ordinateur équipé d'un microprocesseur Intel (R) core i7, 3,5 GHz, 64 Go de RAM et d'un processeur graphique (GPU) NVIDIA RTX 2080 TI. L'architecture du réseau est mise en œuvre en Python à l'aide de package Keras avec Tensorflow backend.

3.5.5 Résultats et discussion

Dans cette section, nous allons présenter les résultats obtenus par la méthode de classification proposée. Nous avons analysé les performances de la méthode proposée en utilisant quatre mesures d'évaluation, à savoir l'accuracy, la précision, le rappel et le f1-score.

Tout d'abord, nous avons analysé les résultats obtenus par la méthode proposée. Ensuite, nous avons effectué une comparaison entre les résultats obtenus du modèle d'ensemble proposé et des classificateurs CNN individuels. Enfin, la comparaison entre l'architecture d'ensemble proposée et d'autres CNN pré-entraînés est présentée.

La figure 3.19 montre les résultats de la classification sur les images de test réalisés par l'architecture d'ensemble proposée pour différentes époques.

On peut observer d'après la figure 3.19 que l'impact du nombre d'époques influe sur les résultats de classification, la précision de classification augmente progressivement avec le nombre d'époques. Cependant, l'augmentation de nombre d'époques entraîne une augmentation de temps d'entraînement. Mais ce n'est pas important car notre objectif principal est d'obtenir la plus grande précision possible.

Le tableau 3.5 montre la performance de classification de la méthode proposée pour les différentes classes. Les résultats ont montré que la méthode proposée était capable de détecter

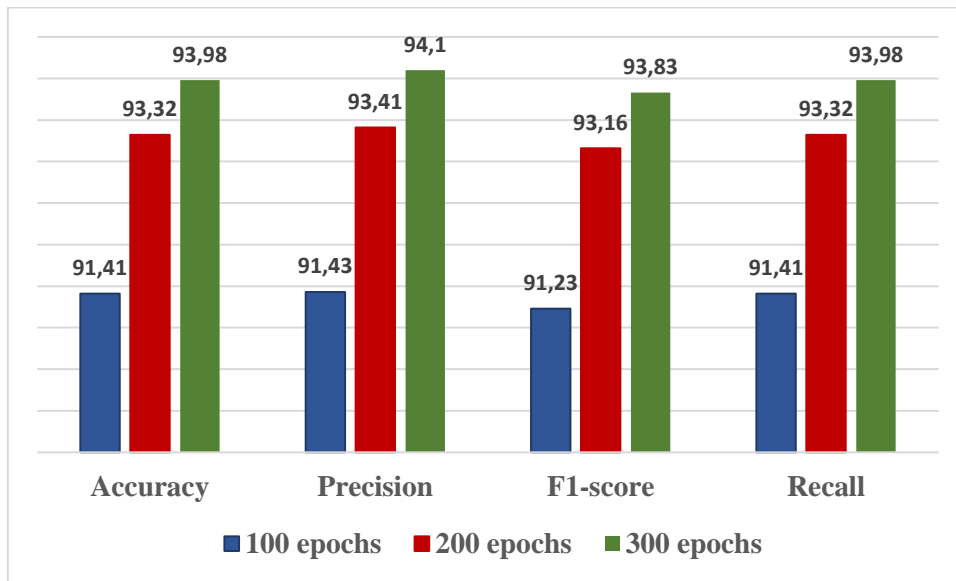


Figure 3. 19 Résultats de classification obtenus par la méthode proposée pour différentes époques.

différentes anomalies dans WCE avec une grande précision, notamment, Blood et Reduced mucosal view.

Tableau 3. 5 Performance de classification de la méthode proposé sur différentes classes.

Labels	Précision	Rappel	F1-score	Précision globale
<i>Angiectasia</i>	89.89	95.51	92.62	
<i>Blood</i>	99.20	99.20	99.20	
<i>Erosion</i>	90.54	93.59	92.04	
<i>Lymphoid-hyperphasia</i>	90.99	98.72	94.70	93.98 %
<i>Normal mucosa</i>	95.13	75.16	83.97	
<i>Reduced mucosal view</i>	98.56	98.40	98.48	
<i>Ulcer</i>	94.40	97.28	95.82	

Les cartes de caractéristiques extraites de différentes couches CNN d'une image de classe BLOOD sont présentées à la figure 3.20. Pour une meilleure démonstration, nous avons extrait des cartes de caractéristiques de certaines couches. Nous pouvons voir que les cartes de caractéristiques dans les couches CNN de bas niveau peuvent conserver les informations spatiales afin d'apprendre les paramètres de ces couches, mais elles ignorent ces informations dans les couches de haut niveau et deviennent plus abstraites.

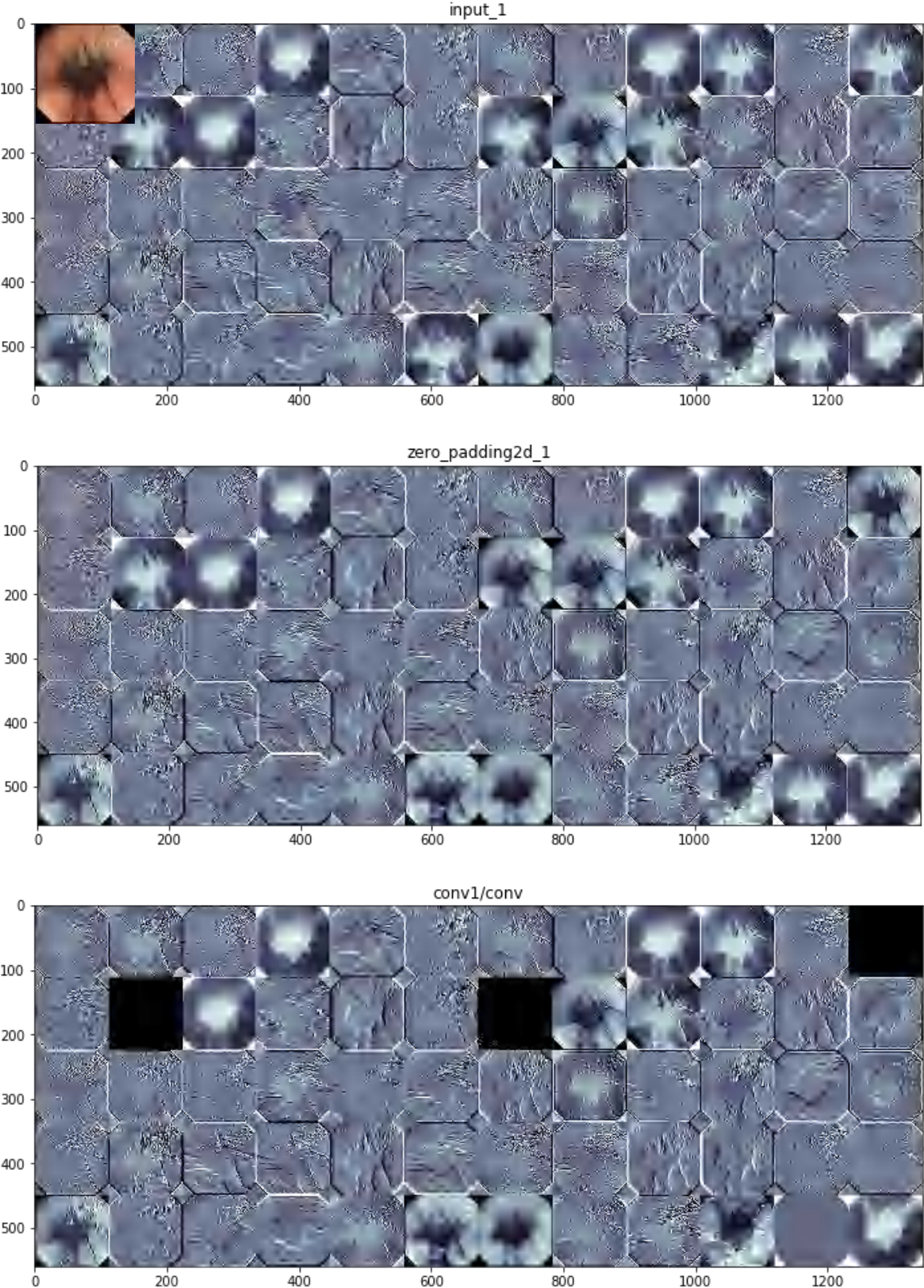


Figure 3. 20 Cartes de caractéristiques des couches convolutives apprise par notre modèle.

Chapitre 3 : Classification des images WCE en utilisant l'apprentissage profond

Pour tester l'efficacité de l'approche proposée, nous avons évalué les performances de chaque architecture individuellement. La figure 3.21 présente la comparaison des architectures individuelles mobilenet et Densenet169 avec le modèle proposé. À partir de cette figure, on peut voir que notre méthode proposée dépasse les architectures individuelles pour tous les critères d'évaluation. Notre modèle dépasse densenet169 et mobilenet avec 1.63%, 3.16% respectivement en termes de précision. De plus, densenet169 architecture (92.35%) dépasse mobilenet architecture (90.82%), Cela signifie que DenseNet169 a un avantage de précision de 1,53 %.

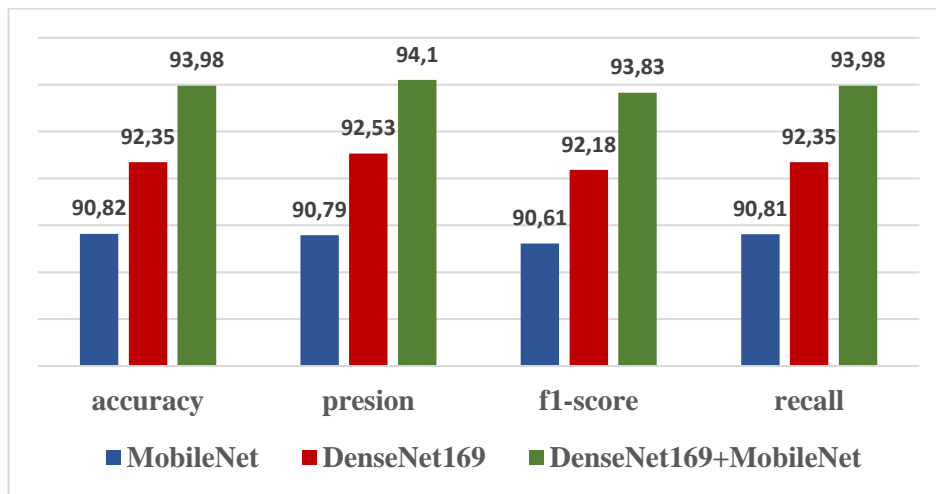


Figure 3. 21 Comparaison des résultats de classification obtenus par notre méthode et classificateurs individuelles.

La figure 3.22 montre les courbes de précision/perte pour entraînement/validation pour MobileNet, densenet169 et architecture proposée. Les courbes d'apprentissage pour toutes les expériences sont bonnes. La perte diminue et la précision augmente progressivement le long d'apprentissage. Après avoir entraîné tous les modèles avec 300 époques, nous atteignons une précision de 98.93%, 98.90% et 99,75 % avec perte de 0,042, 0,042 et 0,016 pour MobileNet, Densenet169 et l'architecture proposée respectivement.

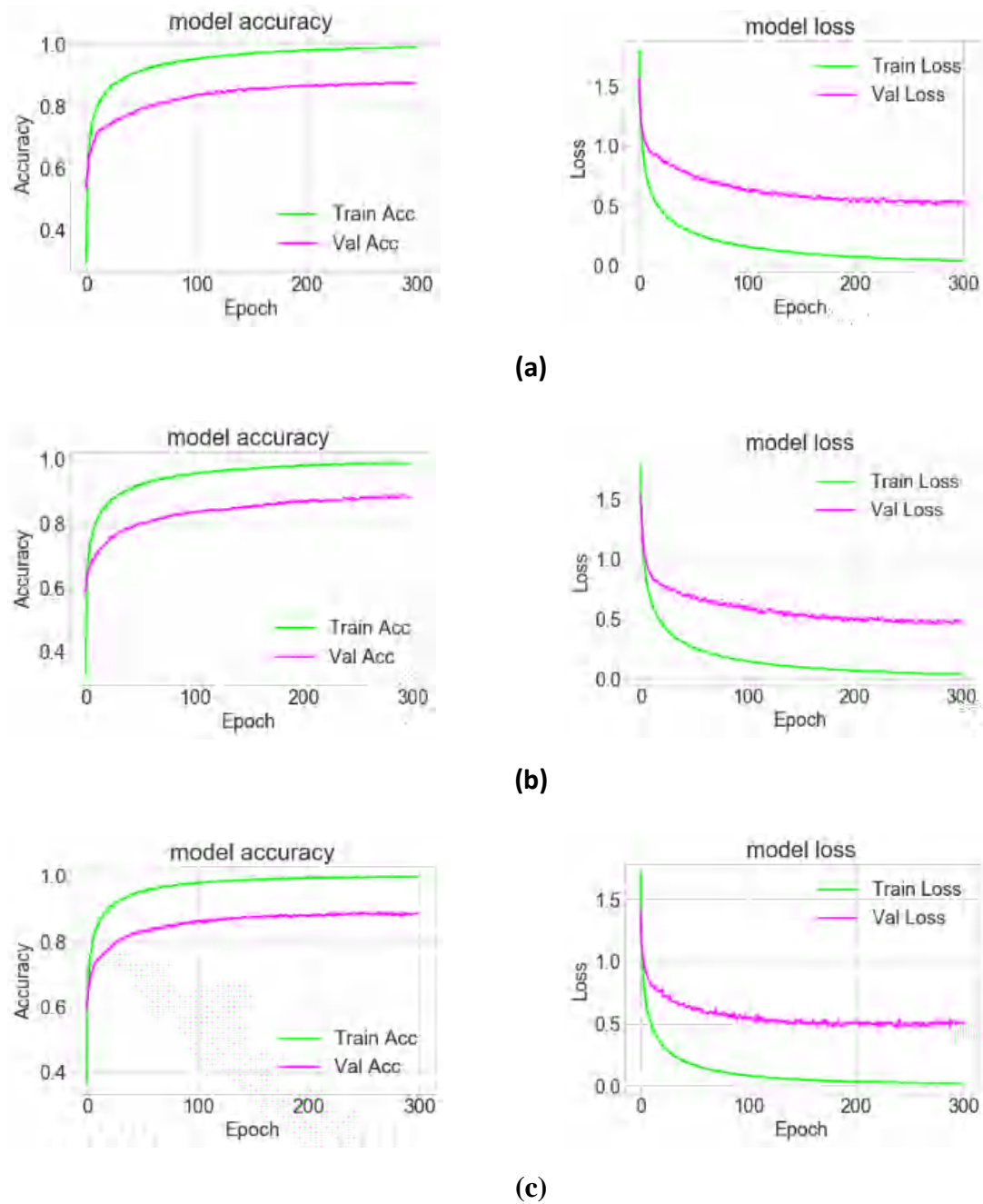
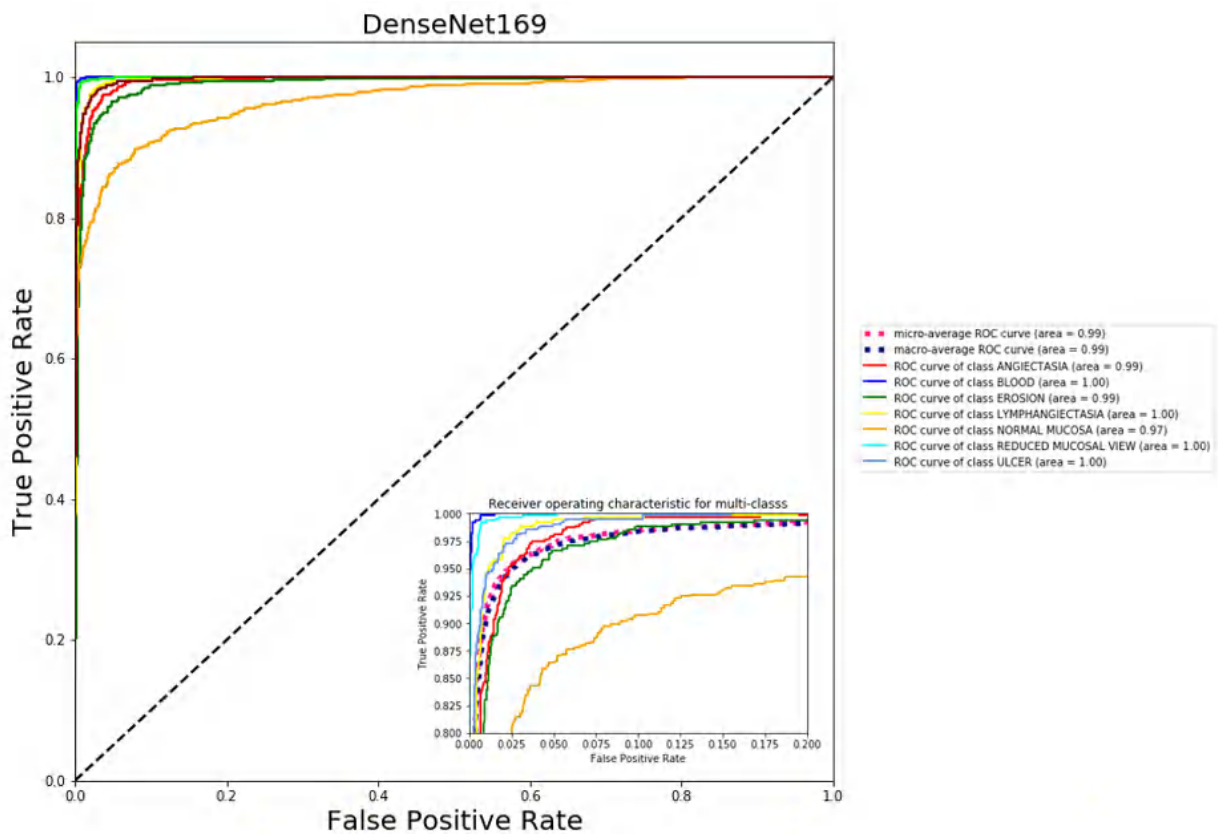
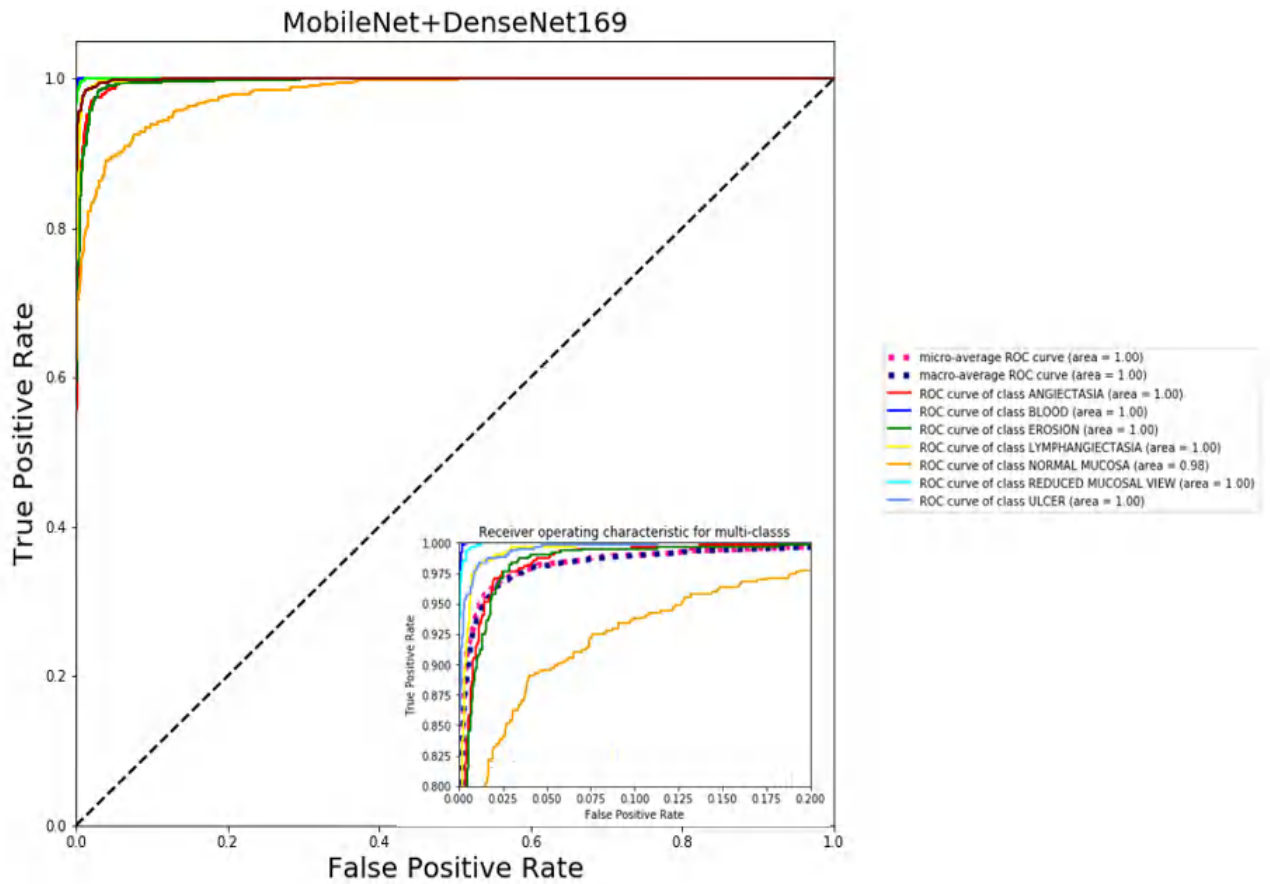


Figure 22 Courbes de précision et de perte pour (a) mobilenet, (b) densenet169 et (c) méthode proposée.

L'AUC pour chaque classe des trois architectures est indiqué dans des courbes ROC sur la figure 3.23. A partir de cette figure, on observe clairement que la méthode proposée dépasse les deux autres architectures avec le rappel de détection le plus élevé et les taux de faux positifs les plus faibles. La courbe ROC pour la méthode proposé atteint la valeur optimale 1 pour les classes Angiectasia, Blood, Erosion, Lymphoid-hyperphasia, Reduced mucosal view et Ulcer. Cela indique que notre méthode a une grande précision de détection par rapport aux autres modèles.



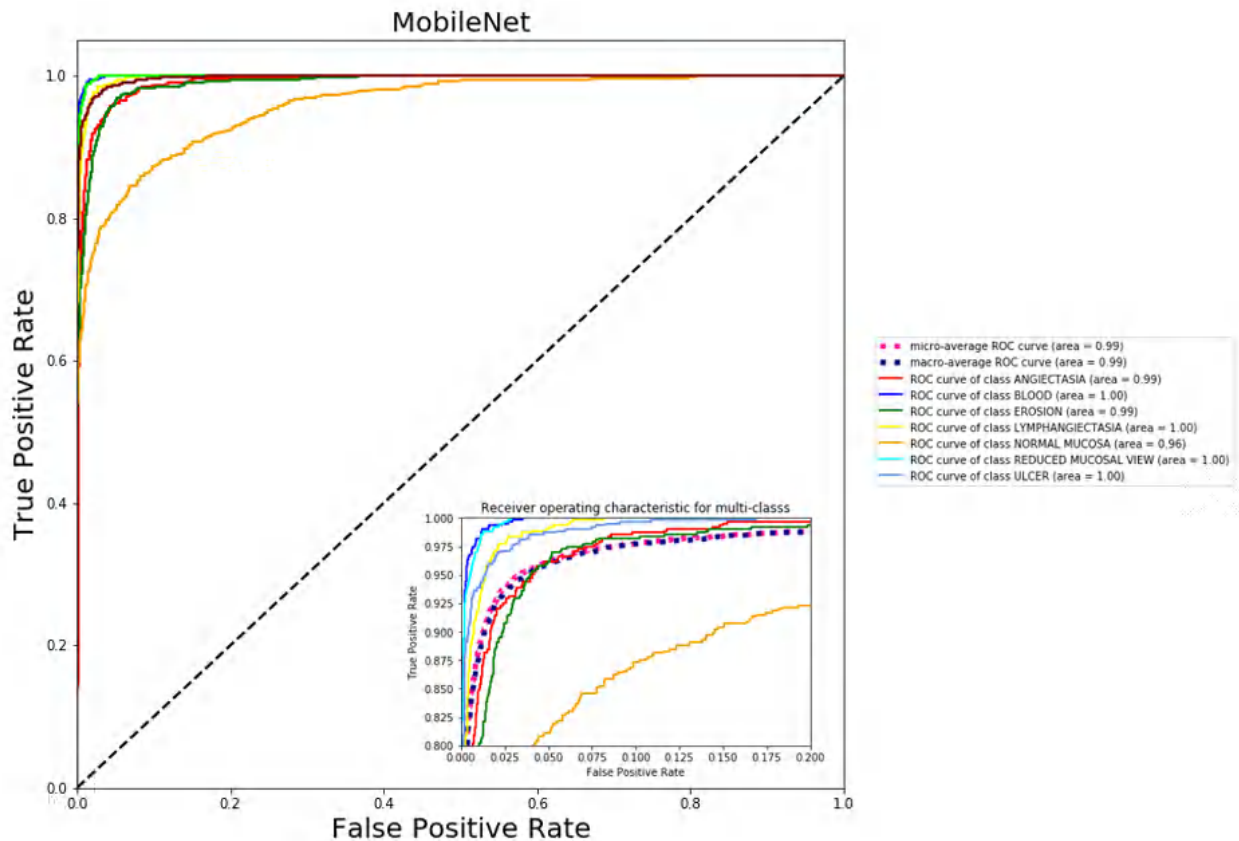


Figure 3. 23 Comparaison des courbes ROC de différents modèles.

En outre, les matrices de confusion des deux modèles (densenet169 et mobilenet) et le modèle proposé sont présentés sur la figure 3.24. Dans ce cas, notre modèle a démontré le meilleur résultat avec un taux d'erreur 6.02% (échantillons mal classés), tandis que mobilenet et densenet169 ont montrés des résultats moins performants avec des taux d'erreur de 9.18% et 7.65% respectivement.

Les résultats de classification des différentes architectures CNN pré-entraînées, y compris InceptionV3, Xception, InceptionResNetV2, MobileNetV2, Densenet121, et Densenet201, sont résumés dans le tableau 3.6. Il est clair que l'architecture proposée atteint la meilleure précision de classification avec 93.98 % par rapport aux autres architectures, tandis que les résultats les plus faibles sont obtenus par l'architecture xception avec seulement 88.80%. Les modèles Inceptionv3 et Densenet121 ont montré résultats proches avec 90,96 % et 91,12 % de précision respectivement. Le modèle Densenet201 a obtenu des résultats acceptables avec une précision de 91,92 %. La précision des autres modèles n'est pas dépassée 90,4 %.

Chapitre 3 : Classification des images WCE en utilisant l'apprentissage profond

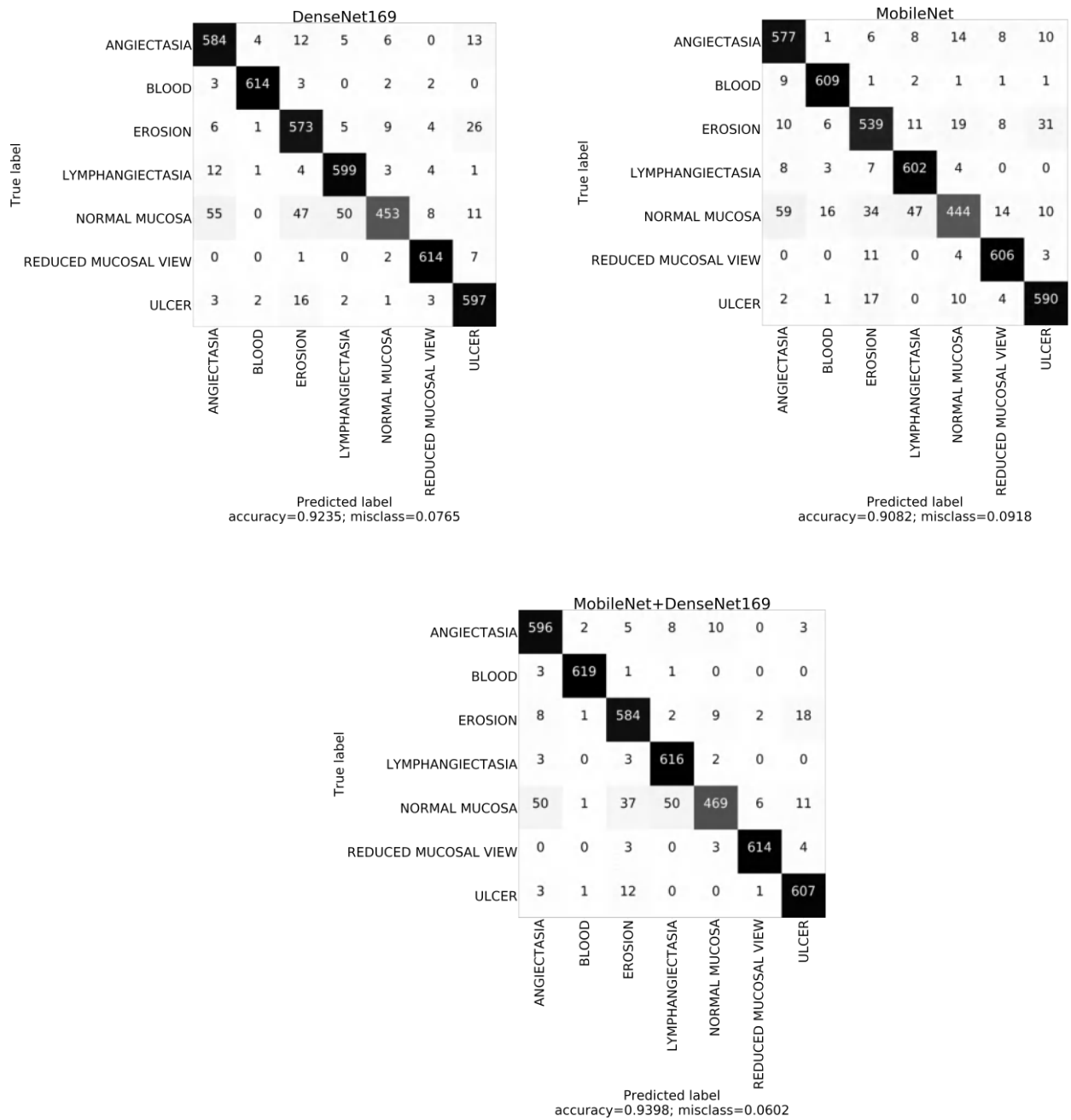


Figure 3.24 Matrice de confusion de la classification des maladies GI.

Chapitre 3 : Classification des images WCE en utilisant l'apprentissage profond

Tableau 3.6 Comparaison des résultats de classification obtenus par différents CNN pré-entraînés.

<i>CNNs</i>	<i>Accuracy (%)</i>
<i>Inceptionv3</i>	90.96
<i>Inceptionresnetv2</i>	90.36
<i>Densnet201</i>	91.92
<i>Densenet121</i>	91.12
<i>MobileNetV2</i>	89.58
<i>Xception</i>	88.80
<i>Proposed</i>	93.98

3.5 Conclusion

Dans ce chapitre, on a présenté une classification automatique des lésions et des organes digestifs des images WCE. Cette classification est réalisée en utilisant l'entraînement à partir de zéro d'un CNN proposé et un apprentissage par transfert des CNN pré-entraînés. Nous avons utilisé diverses approches d'augmentation telles que la rotation aléatoire, le retournement horizontal et vertical, le contraste et la luminosité. L'augmentation des données nous a permis d'ajuster la taille de l'ensemble d'apprentissage et de réduire le problème de sur-apprentissage. Nous avons utilisé diverses mesures d'évaluation standard telle que l'accuracy, la précision, la sensibilité, la spécificité et le MCC. Les résultats obtenus par les deux méthodes d'apprentissage sont très satisfaisants. Nous avons également présenté une approche d'apprentissage par transfert basée sur la combinaison de deux modèles CNN pré-entraînés pour détecter automatiquement les maladies dans le tube digestif. Cette combinaison est utilisée pour extraire plus de caractéristiques des images d'entrée. Les résultats obtenus montrent que la combinaison de ces diverses caractéristiques améliore les performances et donne des résultats plus précis (93.98 %) en comparaison avec les classificateurs individuels.

Chapitre 4

Systeme de compression des images

WCE propose

4.1 Introduction

Dans ce chapitre, on va présenter notre système de compression intelligent, qui répond aux problèmes de limitations énergétiques de la WCE. Le principe est d'inclure une boucle de rétroaction de classification, basée sur l'apprentissage profond, pour déterminer l'importance des images transmises. Cette classification est utilisée avec un simple algorithme de compression basé sur la prédiction pour permettre une gestion intelligente de l'énergie limitée de la capsule. Pour cela, la capsule commence par transmettre une version sous-échantillonnée de chaque image avec un faible taux. Les images seront décodées et classées, automatiquement, pour détecter d'éventuelles lésions. Suite au résultat de la classification, les images considérées comme importantes pour le diagnostic seront enrichies d'un contenu supplémentaire, tandis que les images moins importantes seront enregistrées avec une faible qualité. De cette façon, une grande quantité de bits sera économisée, sans affecter le diagnostic. L'énergie économisée peut être utilisée pour prolonger la durée de vie de la capsule ou pour augmenter la résolution et la fréquence d'images de certaines images WCE.

4.2 Etat de l'art sur la compression des images WCE

La compression des images WCE est un compromis entre le taux, la qualité et les ressources. Les techniques de compression avec perte, comme la compression d'images basée sur la transformée en cosinus discrète (DCT), ont été parmi les premières utilisées [112]–[114]. Ces algorithmes sont très populaires pour le codage des images naturelles, en raison de leur efficacité. Cependant, dans les applications à ressources limitées, ils peuvent être considérés comme complexes. Ces algorithmes, nécessitent beaucoup de stockage et de calculs. Les pixels de l'image sont traités dans des blocs de taille 4x4 ou 8x8, alors que les capteurs d'image CMOS commerciaux produisent les pixels dans un ordre ligne par ligne. Par conséquent, une mémoire tampon supplémentaire est nécessaire pour stocker les pixels et leurs coefficients DCT correspondants. Cela entraîne une augmentation de la surface du circuit et donc de la consommation d'énergie [8]. En outre, bien que les algorithmes de compression avec perte produisent une compression élevée, ils peuvent réduire la qualité de l'image. Même si une certaine distorsion des images peut être cliniquement acceptable, les algorithmes de compression sans perte sont préférés pour un diagnostic médical précis [8], [115]. La compression sans perte permet d'obtenir la qualité de l'image originale sans aucune distorsion, bien qu'elle ait un taux de compression (CR) plus faible, ce qui signifie que la transmission des données peut consommer plus d'énergie. Ainsi, de nombreux algorithmes sans perte et presque sans perte pour la compression des images WCE ont été proposés dans la littérature. Des

algorithmes de compression basés sur JPEG-LS ont été proposés pour WCE dans [6], [116], [117]. Cependant, ces algorithmes utilisent le moteur standard JPEG-LS, qui contient d'autres opérations de prétraitement, ce qui le rend complexe et difficile à mettre en œuvre. Des algorithmes de compression sans perte d'image beaucoup plus simples, basés sur une simple modulation différentielle codée par impulsions (DPCM), ont été proposés dans [118]–[123]. Les performances de la DPCM dépendent de la corrélation entre les pixels voisins. Si la différence entre les pixels adjacents est faible, de bons résultats sont obtenus, comme c'est le cas dans la plupart des images de WCE.

Pour coder les informations de couleur, on utilise généralement une transformation RVB vers YUV ou une transformation de couleur similaire. Cette opération permet de réduire les informations redondantes qui existent entre les bandes RVB des images couleur. Dans l'espace couleur YUV, l'homogénéité entre les pixels adjacents de chrominance (UV) est élevée et permet d'obtenir une plus grande compression, même avec un simple codage DPCM. Après la transformation des couleurs, chaque nouvelle composante de couleur sera codée indépendamment, avec seulement la DPCM [124] et un codage entropique simple comme un Golomb-Rice [125]. Certains chercheurs proposent d'augmenter la compression dans l'espace couleur YUV en utilisant le sous-échantillonnage de la chrominance [43], [126], car pour la perception, l'information de luminance (Y) est la plus importante. De plus, les couleurs des images WCE sont très limitées et sont dominées par le rouge. Enfin, certaines propositions, avec ou sans perte, opèrent directement sur les données provenant du capteur avec le réseau de filtres colorés (CFA), sans opérations de démosaïquage qui sont déplacées vers le décodeur [120], [122].

Après avoir enregistré toutes les images transmises, le diagnostic commence par la lecture de toutes les images par le médecin. Bien que les images soient affichées à une fréquence d'images plus élevée, cette opération reste longue et fastidieuse. Certaines études montrent que pour les capsules de l'intestin grêle, le temps de lecture peut aller de 30 à 120 minutes [127]. Par conséquent, de nombreux systèmes automatiques ont été proposés pour résumer les vidéos pour les médecins [128]–[130] ou pour détecter automatiquement les lésions. Cependant, toutes ces opérations sont effectuées après que toutes les images ont été transmises. Si la plupart des images ne sont pas visualisées, après avoir été transmises avec un coût important, cela signifie que la majeure partie de l'énergie de la capsule a été gaspillée sur des images non utiles. En général, les images importantes pour le diagnostic représentent moins de 5 % de l'ensemble des images [32]. Lorsque on sait quelles sont les images importantes et utiles, avant ou pendant leur codage, nous aurait permis d'utiliser la capsule pour capturer des images plus importantes

pour le diagnostic. Lors du codage, il a été proposé de localiser la capsule dans le corps [131] ou même de sélectionner certaines images uniquement à l'intérieur de la capsule [132]. Ces opérations peuvent être utilisées pour adapter localement le système de codage de la capsule. Cependant, la connaissance de la position seule ne permet pas de connaître exactement l'importance des images individuelles. Dans le travail de [132], une simple détection de la similarité des images (par le calcul des moments) à l'intérieur de la capsule a été proposée. De cette manière, il est possible d'empêcher l'envoi de trames similaires. Cependant, cette opération doit être très simple pour être mise en œuvre à l'intérieur du WCE. Dans [133], les mêmes auteurs ont proposé de détecter les polypes à l'intérieur de WCE. Ils ont utilisé deux caractéristiques simples de filtres de contraste et de bords pour détecter les images de polypes. Dans les deux cas, il est nécessaire de stocker les images entières et de les traiter. La capsule doit également traiter chaque image pour déterminer leur importance, tout en ne transmettant que celles qui sont importantes. Cela peut réduire la consommation d'énergie, si la fraction des images transmises est faible. Cependant, ces systèmes ne peuvent pas utiliser des opérations de détection intelligentes et leur taux de détection sera faible. En outre, personne ne saura quelles images importantes, pour le diagnostic, ont été manquées.

L' utilisation d'une gestion de l'énergie par la compression, pour chaque image de WCE, a été proposée dans [134]. Cependant, les auteurs supposent que l'importance des images est connue. Dans [135], les auteurs ont proposé d'utiliser l'apprentissage profond, au niveau du décodeur, pour contrôler la taille et la fréquence d'images des images WCE, mais aucun système pratique n'a été proposé. Dans cette thèse, nous proposons de combiner ces deux idées pour produire un système de compression intelligent. Pour cela, nous allons inclure une classification basée sur l'apprentissage profond dans la boucle de rétroaction d'un codage prédictif de WCE. Ce système suivra la compression de chaque image, afin de s'assurer que l'énergie limitée a été utilisée de manière optimale pour permettre le meilleur diagnostic. L'humain peut revoir la même image plusieurs fois et modifier ses décisions, mais il ne peut pas le faire à grande vitesse. Nous ne pouvons pas mettre un humain dans la boucle car la collecte des images prendra plus de 10 heures et elle est effectuée à une vitesse de 2-4 images par seconde. Ainsi, cette opération ne peut être réalisée que par un système automatique intelligent. La classification sera également effectuée par le décodeur, car elle nécessite un système complexe. En fonction des résultats de la classification, le système modifiera la qualité des images codées. Cette opération permettra d'économiser de l'énergie qui pourra être utilisée pour changer dynamiquement le nombre ou la résolution des images, ou simplement pour prolonger la durée de vie de la capsule. Cette classification ne remplacera pas le médecin dans le diagnostic, mais effectue juste un premier

tri qui doit être fait en temps réel, lors de la transmission de chaque image. Cela signifie que même pour les images non importantes, nous devons maintenir une qualité acceptable, au cas où la classification serait incorrecte. D'un autre côté, les médecins préfèrent perdre de l'énergie plutôt que de manquer une lésion importante qui pourrait affecter le diagnostic médical. Par conséquent, nous proposons de ne réduire la qualité que des images moins importantes, et de ne pas les ignorer du tout. Par conséquent, notre système codera toutes les images, mais allouera plus d'énergie aux images les plus importantes, où les lésions potentielles sont détectées. Enfin, toute amélioration future des systèmes de diagnostic intelligents rendra automatiquement la technique proposée plus efficace et plus fiable. Nous pouvons changer le réseau de classification par un réseau plus efficace, lorsqu'il est disponible, sans modifier le reste du système. En outre, lorsque la classification est plus efficace, nous pouvons réduire davantage la qualité des images non importantes. Cela permettra une économie supplémentaire de bits ou d'énergie, sans craindre qu'elle soit basée sur des décisions incorrectes.

Actuellement, notre objectif principal de la classification est de réduire uniquement l'énergie consommée par les images qui ne sont pas importantes, pour le diagnostic, en réduisant leur coût en bits. Nous avons également proposé de classer les images des organes GI (estomac, colon, intestin grêle, etc.). Nous utilisons cette classification pour contrôler la qualité et le débit d'images dans les cas où les médecins se concentrent sur des zones spécifiques. Lorsque le système détecte que la WCE a atteint la région ciblée, le décodeur envoie un signal de commande à la capsule pour augmenter temporairement la qualité de l'image et la fréquence d'images. Si les images ne proviennent pas de la région ciblée, la capsule est invitée à réduire sa fréquence d'images au taux faible de 2 à 4 images par seconde.

4.3 La méthode proposée

Cette section présente notre méthode de compression intelligente qui vise à optimiser la gestion énergétique du système WCE, par un codage adaptatif. Nous utilisons une classification basée sur l'apprentissage profond pour déterminer les images importantes, pour le diagnostic, et celles qui sont moins importantes. Cette classification est effectuée au niveau du décodeur qui contrôle le codeur via un canal de retour. Nous contrôlons la qualité des images en appliquant un sous-échantillonnage seulement sur les composantes de chrominance (422, 420,...) des images considérées comme non importantes. Le codage avec perte de l'information de chrominance

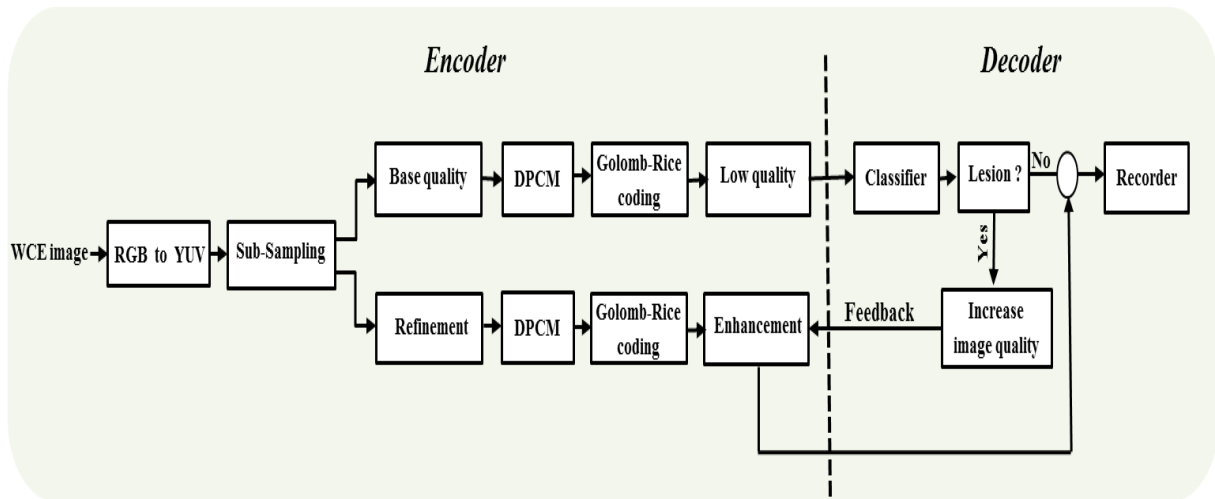


Figure 4. 1 Schéma de principe de la méthode de codage proposée

réduira le coût des images tout en permettant une bonne qualité visuelle, pour effectuer une classification précise.

Le schéma de la méthode proposée est présenté dans la Figure 4.1. Tout d'abord, une image couleur RVB est transformée en un espace couleur YUV en utilisant la transformation RVB-YUV. L'image résultante est divisée en deux parties : une partie codée avec perte produite par sous-échantillonnage (qualité de base) et une partie raffinée qui contient les pixels écartés et qui, si elle est ajoutée, permettra le décodage sans perte. Pour réduire le taux de transmission de certaines images, nous ne transmettons que la qualité de base. Les deux parties de l'image, de qualité de base et de raffinement, sont codées en utilisant le codage DPCM et les valeurs résiduelles résultantes sont codées en utilisant le codage Golomb-Rice. Il s'agit du même système proposé par [43], et le codage est effectué ligne par ligne. Le flux binaire compressé résultant de la qualité de base entière est d'abord transmise au décodeur. Ensuite, l'image de qualité de base décodée est classifiée automatiquement, afin de détecter si elle représente une région normale ou pathologique du tube GI. Si le résultat de la classification indique une image de classe normale, le décodage est terminé et l'enregistreur stocke l'image. En revanche, si l'image est classée comme importante (classe pathologique), le décodeur demandera à l'encodeur la transmission de la partie d'amélioration, en utilisant le canal de retour. En ajoutant les pixels manquants, dans ce cas, le décodeur produira une image sans perte avant de la stocker dans l'enregistreur. Cette opération permet de réduire le nombre de bits transmis et donc d'économiser la quantité d'énergie qui aurait été utilisée pour les transmettre. La quantité de bits économisés dépend du sous-échantillonnage utilisé et du nombre d'images qui sont classées comme non importantes. Les bits économisés peuvent être utilisés pour coder des images supplémentaires, ce qui prolongera la durée de vie de la capsule, sans affecter le diagnostic. De

même, nous pouvons conserver la même durée de vie, mais nous augmentons la taille de toutes les images capturées. Dans un autre scénario, nous pouvons également utiliser cette approche pour augmenter la qualité et la fréquence d'images WCE dans certaines régions anatomiques spécifiques uniquement. Dans ce cas, l'algorithme de classification doit déterminer si l'image appartient à la région d'intérêt ou non. Ainsi, le décodeur demandera la transmission de plus d'images ou d'images plus grandes, alors que la capsule traverse uniquement la région d'intérêt. Dans ce cas, la demande du décodeur peut être faite uniquement au début et à la fin de la région d'intérêt. Plus de détails sur la méthode proposée sont discutés dans les sections suivantes.

4.3.1 Le codeur

4.3.1.1 Conversion de RVB en YUV

Tout d'abord, nous convertissons les images WCE de l'espace couleur RVB à YUV, en utilisant les équations suivantes :

$$Y = ((66 \times R + 129 \times V + 25 \times B + 128) / 256) + 16 \quad (4.1)$$

$$U = ((-38 \times R - 74 \times V + 122 \times B + 128) / 256) + 128 \quad (4.2)$$

$$V = ((112 \times R - 94 \times V - 18 \times B + 128) / 256) + 128 \quad (4.3)$$

Où R, V, B sont les composantes rouge, verte et bleue respectivement.

Y est la composante de luminance et U (chrominance bleue) et V (chrominance rouge) sont les composantes de chrominance.

Figure 4.2 représente un exemple des composantes Y, U et V d'une image endoscopique.

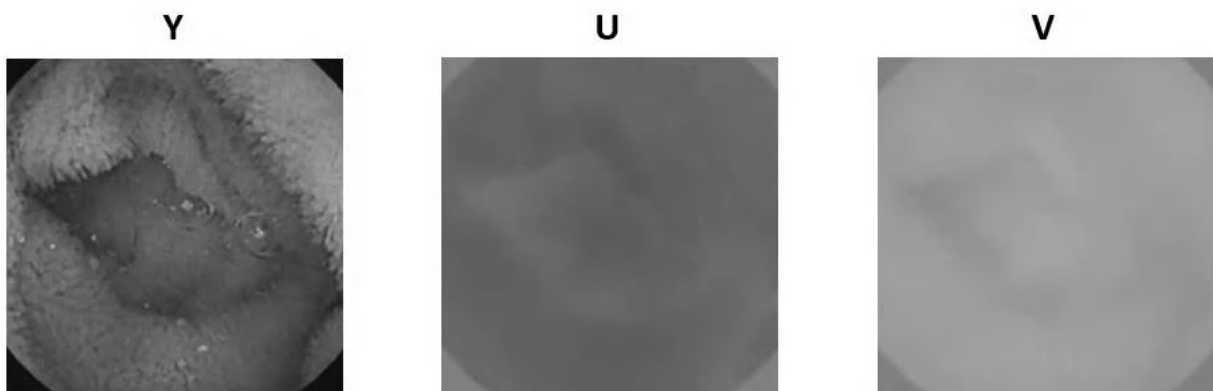


Figure 4.2 Les différentes composantes YUV d'une image endoscopique

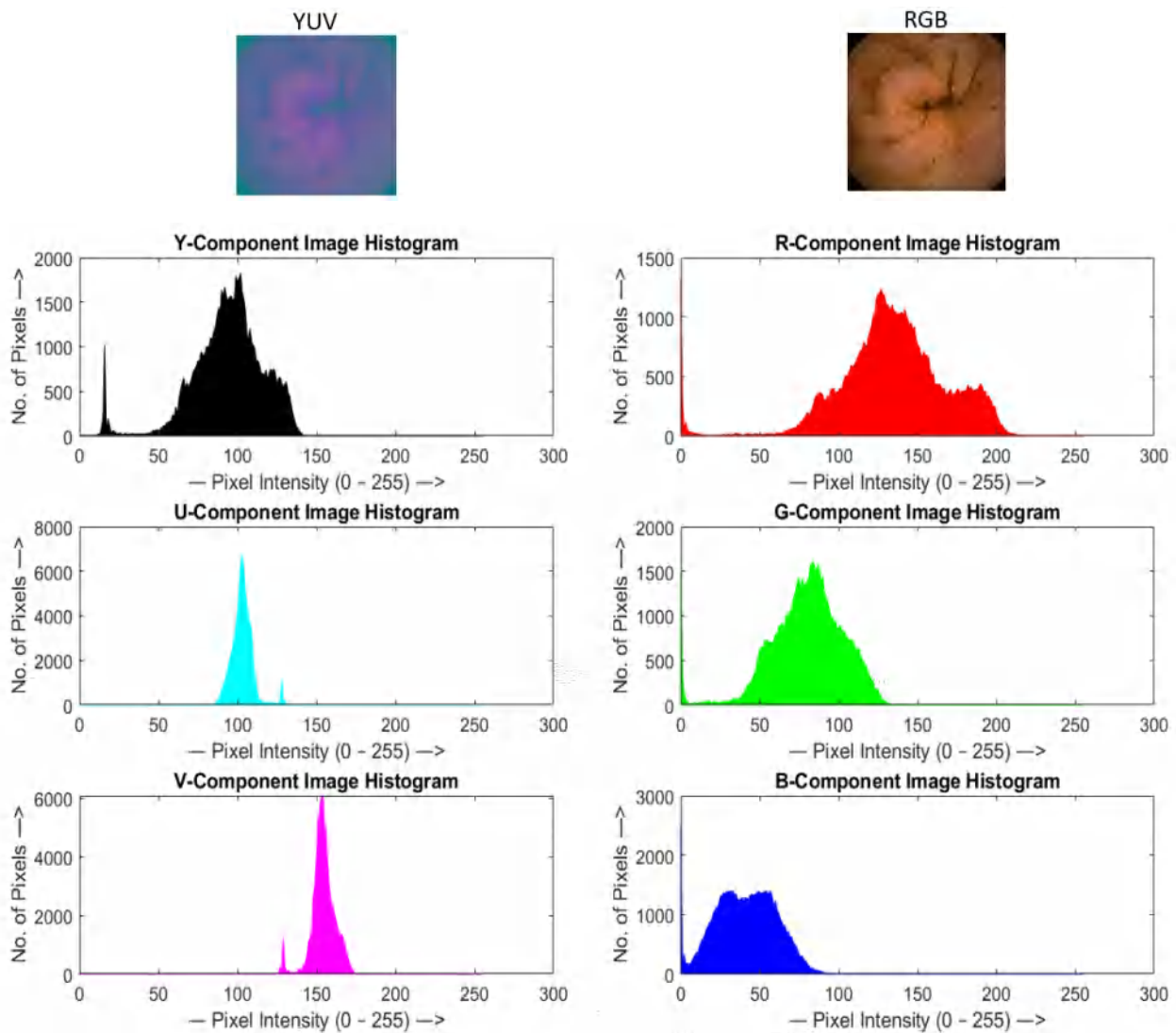


Figure 4.3 Histogramme des composantes RVB et YUV d'une image endoscopique.

La figure 4.3 représente la distribution d'intensité des composantes RVB et YUV respectivement.

D'après cette figure, on observe que la distribution d'intensité de RVB est large par rapport aux composantes YUV, dans laquelle on observe que les composantes U et V ont moins d'informations, ce qui conduit à une compression élevée des données.

La figure 4.3 montre aussi que les composantes de couleur U et V ne varient généralement pas trop, ce qui est dû à l'homogénéité des couleurs dans les images de la WCE. Il est également à noter que cette transformation ne dégrade pas théoriquement la qualité de l'image car aucune information n'est perdue [136].

Bien que nous utilisons YUV, ce codage peut exploiter toute autre transformation de couleur comme YEF ou YEN ou même la CFA [8].

Comme le codage est effectué ligne par ligne, cette transformation est effectuée pour chaque pixel codé, mais chaque composante résultante est codée séparément. Cette opération permet de réduire la redondance spectrale et de séparer la luminance (Y) importante de la chrominance moins importante [137].

4.3.1.2 Sous échantillonnage

Dans notre système, nous avons utilisé les schémas de sous-échantillonnage de la chrominance 4 :2 :2 et 4 :2 :0. La figure 4.5 montre la différence entre les schémas YUV444, YUV422 et YUV420.

D’après la figure 4.5, le sous-échantillonnage 422 et 444 est effectué sur les composantes de chrominance (U et V) uniquement. Y n’est pas sous-échantillonné du tout. Le schéma 444 montre une image de haute qualité, sans aucun sous-échantillonnage de chrominance. Dans le schéma 422, le sous-échantillonnage est effectué uniquement dans la direction horizontale. Dans ce cas, la taille totale de l’image sera réduite à 2/3 de sa taille originale. Tandis que le sous échantillonnage 420 réduit la résolution verticale de la chrominance ainsi que la résolution horizontale. Alors, la taille totale d’image sera réduite à 1/2 de sa taille.

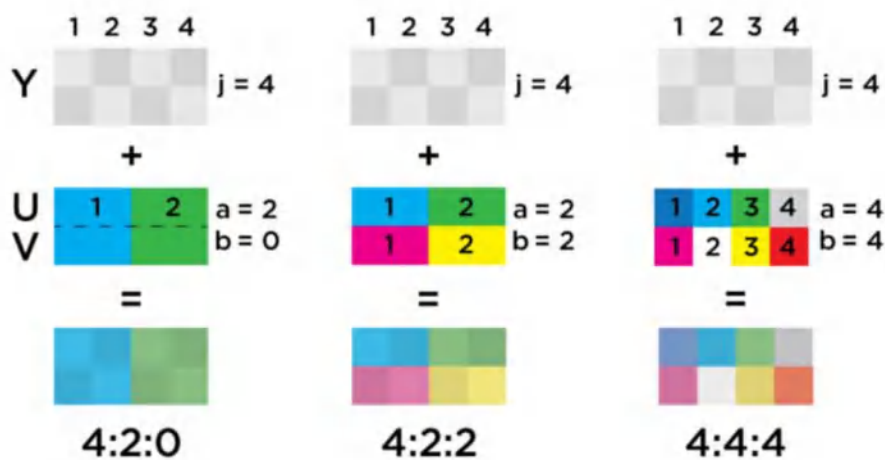


Figure 4.4 Schéma de sous-échantillonnage YUV444, YUV422 et YUV420. [138]

Si l’on suppose que P est le nombre d’images non importantes, on peut s’attendre à ce que le nombre de pixels sauvegardés soit suffisant pour coder un certain nombre d’images supplémentaires (444). Ce nombre peut être calculé à l’aide de l’équation suivante :

$$N_g = \begin{cases} P/3 & \text{for subsampling 422} \\ P/2 & \text{for subsampling 420} \end{cases} \quad (4.4)$$

Où :

N_g : nombre d'images gagnées (444), avant compression.

P : nombre d'images non importantes.

4.3.1.3 Codage prédictif DPCM

Après l'étape de sous-échantillonnage, un algorithme de codage DPCM est appliqué pour produire la différence entre les valeurs de pixels consécutifs en utilisant l'équation suivante :

$$e(i) = x(i) - x(i - 1) \quad (4.5)$$

Où x est la valeur actuelle du pixel, qui peut être Y, U ou V, $x(i-1)$ est la valeur du pixel adjacent précédent, et $e(i)$ est la valeur résiduelle de la prédiction.

Nous utilisons le codage DPCM car il est très simple, sans perte, et la différence entre les pixels consécutifs dans les images endoscopiques est faible, en raison de l'homogénéité entre les pixels [43]. La figure 4.6 montre la variation des pixels pour les images endoscopiques RVB, où l'on observe que la variation des pixels est plus importante. La figure 4.7 montre la variation des pixels pour les images endoscopiques YUV, où l'on observe que dans les composantes U et V, la variation des pixels est moins importante en raison de l'absence d'information dans la composante chroma.

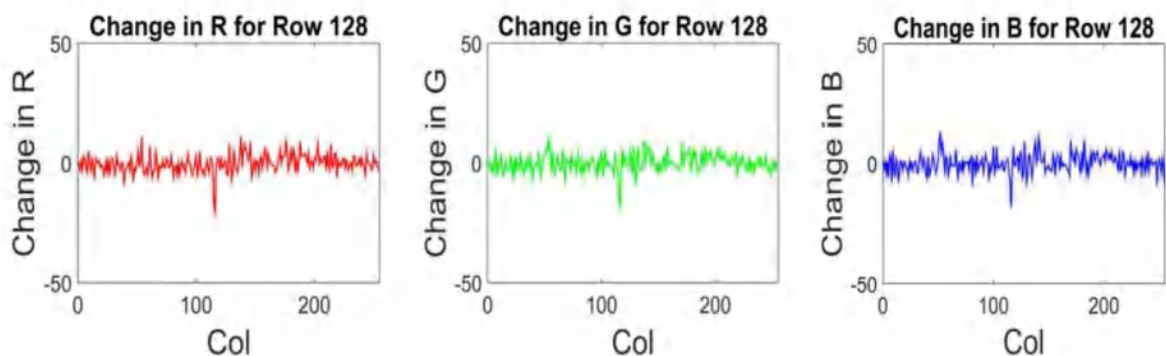


Figure 4.5 Valeurs de la différence $e(i)$ des composantes R, V et B.

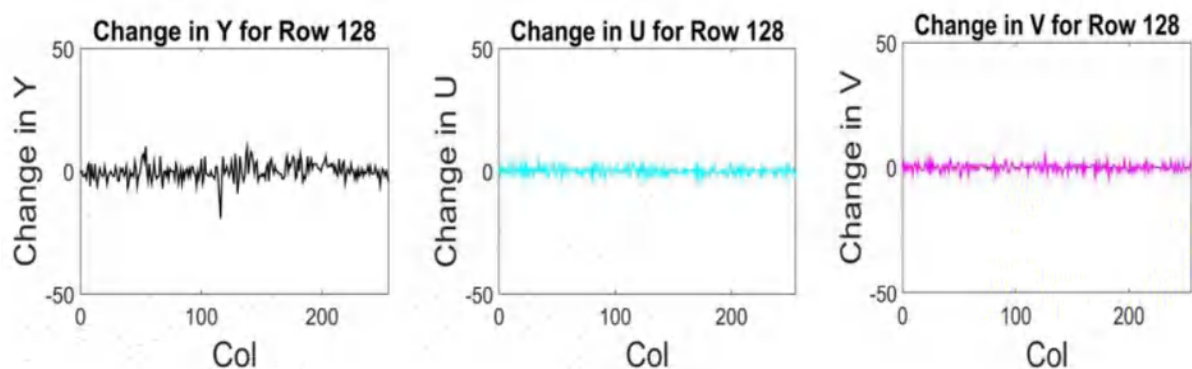


Figure 4.6 Valeurs de la différence $e(i)$ des composantes Y, U et V.

4.3.1.4 Codage Golomb-Rice

Enfin, les résidus de prédiction sont codés par le codeur de Golomb-Rice [125]. Les flux binaires de code résultants sont ensuite transmis au décodeur. On peut utiliser n'importe quel codage entropique, mais le codage de Golomb-Rice est largement utilisé dans les petits capteurs, car il est simple, facile à implémenter en matériel, et donne un bon taux de compression [8]. Ce codeur utilise des valeurs entières positives, mais comme les valeurs résiduelles prédictives peuvent être positives ou négatives, nous les transformons en valeurs positives en utilisant l'équation (4.6).

$$m_{dE} = \begin{cases} 2|e(i)| - 1, & \text{when } e(i) < 0 \\ 2e(i), & \text{when } e(i) \geq 0 \end{cases} \quad (4.6)$$

Le codage DPCM (section 4.3.1.3) est appliqué directement aux pixels Y de luminance. Pour les composantes de chrominance, cependant, le système doit être légèrement modifié pour produire un codage adaptatif. Comme le codeur ne sait pas si l'image codée est importante ou non, il codera à la fois les pixels transmis et les pixels rejetés comme des images distinctes. Les pixels de l'image de faible qualité seront toujours transmis, tandis que le reste des valeurs prédictives seront stockées temporairement jusqu'à ce que le décodeur les demande. Par conséquent, nous devons modifier le codage DPCM pour l'adapter à cette situation. Par exemple, en cas d'utilisation du sous-échantillonnage 4:2:2, la prédiction de la qualité de base des composantes U et V est calculée comme suit :

$$e(i) = x(i) - x(i - 2) \quad (4.7)$$

Cela permettra le décodage de l'image de qualité de base indépendamment du type de l'image. Alors que pour la prédiction des pixels d'amélioration, les erreurs résiduelles sont calculées à l'aide de l'équation (4.5). Les deux opérations de DPCM sont effectuées en même temps. Chaque ligne n'est balayée qu'une seule fois et le résidu de chaque pixel codé est soit stocké localement en tant que raffinement, soit transmis en tant que qualité de base.

En utilisant le codage adaptatif, nous pouvons contrôler efficacement la qualité de chaque image codée, mais cela posera deux nouveaux problèmes. Le premier problème est que le codage d'images séparées entraînera un coût de bits supplémentaire. L'équation (4.7) signifie que le codage DPCM utilisera des pixels de référence plus éloignés, par rapport au cas non adaptatif. Cependant, lorsque nous ajoutons la partie d'amélioration, la perte de prédiction ne sera pas compensée. L'utilisation de l'équation (4.5) pour coder les pixels d'amélioration garantira que la perte de compression n'affectera que les images de faible qualité de U et V. Heureusement,

les pixels de chrominance sont largement corrélés et la perte de prédiction devrait être très faible. En outre, le codage avec perte du nombre beaucoup plus important d'images compensera tout coût additionnel dans le codage des images sans perte. Au cours de son parcours, la WCE peut produire plus de 86 000 images [115]. Si nous ne ciblons que l'intestin grêle, une petite fraction seulement de ces images est importante. En plus, il est peu probable que l'ensemble du tube digestif contienne des lésions. De même, de nombreux articles de l'état de l'art ont abordé le problème de la réduction du temps de lecture des images de la WCE par le clinicien. Ainsi, ils ont proposé de résumer ces séquences d'images en analysant les images automatiquement (avec un CNN ou autre) et de ne laisser que les images importantes. Ceci est fait par exemple par [109], [139] et [140]. Tous partent de l'observation importante que "la plupart des images ne sont pas importantes". Enfin, le mode de lecture intégré Quick-View (de PillCam), qui est utilisé pour réduire le temps de lecture, sélectionne seulement environ 10% des images les plus pertinentes pour créer un aperçu rapide [141].

Le deuxième problème du codeur est que les bits d'amélioration doivent être stockés localement pendant le codage d'une image, jusqu'à ce que le décodeur décide l'importance de l'image. Cette opération nécessite une mémoire et un traitement supplémentaires. Il existe un compromis entre l'énergie de transmission, l'énergie de stockage et l'énergie de traitement. Le stockage des pixels d'amélioration sous une forme comprimée réduit l'espace de stockage, mais nécessite l'exécution des opérations de compression. Cependant, nous pouvons décider de ne stocker que les pixels d'amélioration nécessaires, après leur avoir appliqué le codage DPCM mais non pas le codage Golomb-Rice, jusqu'à ce qu'il soit nécessaire. En moyenne, cela réduira l'énergie de traitement puisque, généralement, la plupart des images WCE sont de faible importance. Cela permettra également de disposer d'un espace de stockage connu et fixe, mais important, au niveau du codeur.

Dans le cas du codage d'une région d'intérêt, l'information sur la qualité de codage nécessaire peut être déterminée au début de la région d'intérêt et à sa fin seulement. Dans ce cas, on peut utiliser un seul mode de codage qui utilise ou non le sous-échantillonnage. Cela signifie qu'un seul type de DPCM est utilisé et qu'il n'y a pas de coût de bits supplémentaires ou une nécessité de stocker les bits d'amélioration.

Comme pour le sous-échantillonnage, la compression permet également d'économiser des bits. Ceci est possible lorsque le sous-échantillonnage soit utilisé ou non. Par conséquent, le gain réel de bits de notre système, par rapport à un codage sans perte de 444 images de toutes les images, ne peut être déterminé qu'après la compression. Le nombre d'images gagnées peut être calculé à l'aide de l'équation suivante :

$$G_i = \frac{G_b \times N_i}{N_b} \quad (4.8)$$

Où :

G_i : nombre d'images gagnées, après compression.

G_b : nombre moyen de bits gagnés par image.

N_i : nombre d'images de la séquence.

N_b : nombre moyen de bits par image, après compression.

Comme les images sont codées avec des taux de compression différents et ont un contenu différent, nous devons utiliser les taux de compression moyens pour calculer le nombre moyen de bits des images codées. La moyenne des bits gagnés par image est la différence entre le nombre moyen de bits des 444 images compressées et le nombre moyen de bits de notre système.

4.3.2 Le décodeur

Le décodeur reçoit d'abord les bits de chaque image de basse qualité. Il décode les codes Golomb-Rice et effectue un décodage DPCM. Ensuite, il interpole l'image YUV sous-échantillonnée décodée et la transforme dans l'espace couleur RVB. L'image résultante sera classée par le CNN pour déterminer son importance. Lorsque le classificateur détecte une image non importante (classe d'image normale), cette image sera enregistrée directement par l'enregistreur. En revanche, lorsque le classificateur détecte une image de classe pathologique, le décodeur demande à l'encodeur de transmettre les bits d'image d'amélioration.

Puisque l'opération de codage DPCM est très simple, le codage et le décodage peuvent être effectués à une vitesse très élevée avec des ressources matérielles très faibles. Ainsi, les principaux problèmes du décodeur seront de produire une classification rapide et précise. Ceci est nécessaire pour fournir un retour d'information en temps réel à la capsule. Comme les images sont codées à une fréquence d'images donnée, la classification d'une image doit être terminée avant que l'image suivante soit codée. Par exemple, avec 2 images par seconde, la classification d'une image doit être terminée en moins de 0,5 seconde, si l'on considère le temps nécessaire au codage DPCM comme négligeable.

Ces dernières années, certains réseaux CNN profonds populaires ont montré d'excellentes performances dans diverses tâches de classification. Cependant, ces réseaux ont des millions de paramètres et nécessitent des ressources informatiques importantes. Le temps nécessaire à la classification correspond au temps de traitement direct (inférence) des images par le CNN. Si les ressources sont disponibles, l'entraînement de réseaux complexes, ou l'utilisation de

l'apprentissage par transfert, donnera des résultats plus précis. De nombreux exemples de benchmarks montrent que même un matériel simple comme les cartes Jetson Nano peut effectuer une classification de 11 fps avec l'inception V4 et une inférence beaucoup plus rapide avec d'autres réseaux [142]. Actuellement, cela est largement suffisant pour une WCE de 2 fps. Cependant, cela peut nécessiter plus d'énergie que ce que nous pouvons fournir pour un décodeur portable. Pour surmonter cette limitation, nous avons utilisé le même réseau de classification présenté dans la section 3.3 du chapitre 3.

4.4 Résultats et discussion

4.4.1 Effet de sous-échantillonnage sur la classification

Afin de voir l'effet de sous-échantillonnage de la chrominance sur les résultats de la classification, nous avons également effectué un autre test avec de nouvelles images. Ces images proviennent de 20% de l'ensemble des images des lésions et anatomiques, qui n'ont pas été utilisées précédemment. Les résultats obtenus sont résumés dans le tableau 4.1.

La notation 444 indique que les résultats sont pour des images de haute qualité, sans aucun sous-échantillonnage de chrominance. Cette notation n'est utilisée qu'à titre de référence, puisque notre système de codage effectuera toujours la classification sur les images sous-échantillonnées de qualité de base. Selon le tableau 4.1, nous pouvons voir que les résultats de la classification sont proches. Cela signifie que l'utilisation du sous-échantillonnage de la chrominance n'affecte pas beaucoup la précision de la classification.

De plus, la Figure 4.9 montre la matrice de confusion de la classification. Comme on peut l'observer, seules 10 images de la classe normale sont mal classées dans la classe anormale. Dans ce test, bien que nous ayons considéré certaines images non importantes comme importantes, nous n'avons manqué aucune image importante. Cependant, dans la pratique, il est possible de manquer certaines images importantes. Le coût d'une mauvaise classification peut être plus élevé pour les images importantes manquées, puisque la WCE transmettra une qualité faible pour une image importante. Il n'y a cependant pas de gros problème de faux positifs, car nous pouvons accepter de coder certaines images non importantes avec une qualité élevée.

Heureusement, deux caractéristiques de notre système de codage peuvent réduire les effets de ces problèmes. Premièrement, notre CNN ne donnera qu'un diagnostic initial, mais le médecin, qui prend la décision finale, si la classification automatique n'est pas correcte, le médecin peut décider de prendre sa propre décision. Pour cela, notre système s'assure qu'en sous-échantillonnant uniquement l'information de chrominance, il produira une bonne qualité même pour les images qui ont été classées comme non importantes. Deuxièmement, nous pouvons

Chapitre 4 : Système de compression des images WCE proposé

modifier le seuil de classification pour réduire les erreurs de classification des images importantes. Le classificateur calcule la probabilité qu'une image appartienne à une classe particulière en utilisant une valeur seuil au-dessus de laquelle l'échantillon est classé dans une classe positive.

Les résultats du tableau 4.1 sont obtenus avec un seuil de 0.5. Nous pouvons diminuer cette valeur et rechercher les résultats qui servent le mieux notre objectif.

Tableau 4. 1 Résultats de la précision avec 95% CI du réseau de classification proposé sur les ensembles de données de lésions et d'anatomie.

	<i>Sous-échantillonnage</i>	<i>Précision sur les images de test (%) (CI 95 %)</i>
<i>Lésions</i>	YUV444 vers RGB	99.89 (99.83 , 99.96)
	YUV422 vers RGB	99.81 (99.73 , 99.90)
	YUV420 vers RGB	99.69 (99.58 , 99.80)
<i>Anatomique</i>	YUV444 vers RGB	99 .06 (98.93 , 99.20)
	YUV422 vers RGB	99.14 (99.01 , 99.27)
	YUV420 vers RGB	99.13 (99.00 , 99.26)

Enfin, la Figure 4.9 montre également les résultats de la classification anatomique. Dans ce cas , les résultats indiquent une bonne classification. Nous pouvons voir qu'il n'y a que 175 images qui ont été incorrectement classées, sur 18760 images. De plus, comme cette classification sera utilisée pour déterminer une région d'intérêt, le fait de mal classer une seule image n'est pas un problème. Notre système peut utiliser les résultats d'un certain nombre d'images consécutives pour déterminer le début ou la fin de cette région. Cela permettra une décision robuste. En outre, le système intelligent peut utiliser des informations supplémentaires en plus de la classification pour prendre sa décision. Par exemple, une image unique d'une classe différente ne peut pas être présente au milieu de plusieurs images d'une autre classe. De plus, le tractus GI a un ordre strict et connu des images et nous ne pouvons pas trouver une image d'intestin avant les images d'estomac.

4.4.2 Analyse des performances de la compression proposée

Dans cette section, nous testons l'efficacité du système de compression intelligent proposé. Cela signifie combien de bits peuvent être gagnés en codant les images non importantes par sous-échantillonnage. Pour cela, nous avons utilisé notre système pour coder six séquences courtes d'images WCE. Dans cette analyse, toutes les séquences de test sont automatiquement classées,

Chapitre 4 : Système de compression des images WCE proposé

pour détecter les régions normales et anormales, puis codées. Cela permettra de montrer comment la compression peut être améliorée, combien de bits et d'images peuvent être gagnés, et comment la fréquence d'images peut être augmentée. Nous avons mesuré la valeur moyenne du taux de compression [8], qui est calculé à partir de la taille de l'image originale et de celle du flux binaire compressé en utilisant la formule suivante [143] :

$$CR = \left(1 - \frac{\text{Total Bits After Compression}}{\text{Total Bits Before Compression}}\right) \times 100 \quad (4.9)$$

Le PSNR est calculé à l'aide de l'équation (4.15) suivante :

$$PSNR = 20 \times \log_{10} \times \frac{255}{\sqrt{\left(1/I \times J\right) \sum_{i=1}^I \sum_{j=1}^J (a_{i,j} - a'_{i,j})^2}} \quad (4.10)$$

Où I et J sont la hauteur et la largeur de l'image, a et a' sont les valeurs des composantes originales et reconstruites.

Le tableau 4.2 présente les résultats de la classification pour chaque séquence de test. Dans ce tableau, nous pouvons voir le nombre d'images normales et anormales, pour chaque séquence de test, tel que déterminé par notre classificateur. Comme prévu, nous avons observé que le nombre d'images importantes dépend du contenu de la séquence. Parfois, le nombre d'images importantes est plus grand que le nombre d'images non-importantes. En fait, nos séquences de test sont courtes, par rapport au nombre d'images que nous pouvons capturer pour l'ensemble

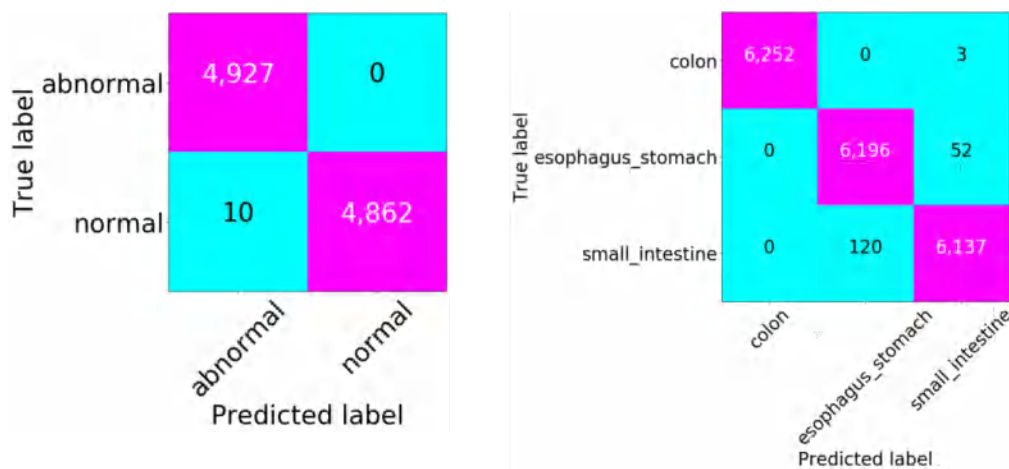


Figure 4. 7 Matrice de confusion de la classification des lésions et de l'anatomie.

Chapitre 4 : Système de compression des images WCE proposé

Tableau 4. 2 Nombre d'images normales et anormales, pour chaque séquence de test, tel que déterminé par notre classificateur.

<i>Sequences</i>	<i>Nombre d'images</i>		
	<i>Classe normal</i>	<i>class Abnormal</i>	<i>Total</i>
<i>Test 1</i>	2219 (88.97%)	275 (11.02%)	2494
<i>Test 2</i>	1306(56.14%)	1020(43.85%)	2326
<i>Test 3</i>	1533(84.83%)	274 (15.16%)	1807
<i>Test 4</i>	287 (48.64%)	303 (51.35%)	590
<i>Test 5</i>	76(23.25%)	244(76.25%)	320
<i>Test 6</i>	257(86.38%)	49(13.61%)	306

du tractus GI, et correspondent à des cas de problèmes médicaux. C'est pourquoi la fraction d'images anormales est élevée. Pour la séquence GI entière, le pourcentage d'images anormales peut être très faible. En outre, les images anormales ne sont pas nécessairement consécutives. Cela signifie que le nombre d'images non importantes dans les régions contenant des lésions peut être faible, et nous pouvons alors trouver de longues séquences d'images non importantes. Dans notre test, chaque image de ces séquences a été codée en fonction de son résultat de classification. Les images pathologiques importantes ont été codées avec une bonne qualité (444), alors que les images non pathologiques ont été sous-échantillonnées en 422 ou 420. La notation (444, 422) signifie que, pour une séquence, les images anormales sont codées 444 et les images normales sont codées 422. Dans ce cas, le codage avec perte n'affecte que la partie UV (Chroma). La composante Y est la plus importante pour la perception. C'est pourquoi nous avons codé Y sans perte. Les résultats obtenus en termes de taux de compression moyen et de qualité (PSNR) sont présentés dans le tableau 4.3. Le PSNR moyen est calculé pour les composantes UV des images non importantes uniquement, puisque les composantes UV des images anormales et toutes les composantes Y sont codés sans perte. Nous pouvons voir que pour le sous-échantillonnage 422, la qualité est toujours très élevée. Pour 420, le PSNR diminue largement, mais on peut considérer que la qualité visuelle restera bonne, puisqu'elle repose principalement sur Y, qui est sans perte. En fait, la qualité dépend aussi de l'opération d'interpolation utilisée pour récupérer la taille originale. Dans notre cas, l'interpolation bilinéaire est utilisée. Dans la littérature, d'autres schémas de sous-échantillonnage ont été proposés. Il existe même des propositions visant à supprimer entièrement la partie chrominance et à la récupérer au niveau du décodeur [144]. Même dans ce cas, on obtient une bonne qualité.

Chapitre 4 : Système de compression des images WCE proposé

Tableau 4. 3 Taux de compression et PSNR obtenus par notre système pour les séquences de test WCE

Séquences	(444,422)			(444,420)		
	CR (%)	PSNR (Y)	PSNR(UV) dB	CR (%)	PSNR (Y)	PSNR(UV) dB
Test 1	73.81	∞	55.29	78.69	∞	31.51
Test 2	68.37	∞	55.64	71.59	∞	31.54
Test 3	71.49	∞	48.49	76.43	∞	31.23
Test 4	70.59	∞	57.71	73.30	∞	31.50
Test 5	68.16	∞	52.25	70.00	∞	31.52
Test 6	71.85	∞	55.25	76.35	∞	31.51

Le CR est le CR moyen de Y, U et V qui est calculé sur toutes les images de la séquence de test. La compression la plus élevée est obtenue pour les composantes U et V, car les pixels de chrominance sont plus corrélés. Ainsi, le gain en CR obtenu dépendra principalement du contenu en chrominance des images, du sous-échantillonnage utilisé et du nombre d'images non importantes de la séquence.

Nous pouvons également voir dans le tableau 4.3 que le schéma (444, 422) fournit un faible CR et un bon PSNR pour toutes les séquences de test. Tandis que, le CR le plus élevé et le PSNR le plus faible sont obtenus par le schéma (444, 420). Ainsi, il est clair que l'augmentation du sous-échantillonnage entraîne toujours une augmentation de la compression. En outre, les résultats de notre système montrent que le taux de compression dépend du nombre d'images importantes. Plus le nombre d'images importantes est élevé, plus la valeur du taux de compression diminue et vice versa. Dans notre test, la meilleure compression est obtenue pour la séquence de test 1, car elle contient un petit nombre d'images anormales. Les résultats indiquent que 2219 sur 2494 images ont été classées comme normales. Cela signifie que 88,97% des images ont été codées avec la qualité 420 (faible énergie), et donc que le taux de compression (CR) est de 78,69%. Cependant, pour la séquence de test 5, la différence de compression moyenne est faible, car le sous-échantillonnage n'est utilisé que pour 23,25% des images.

En conclusion, le meilleur schéma à utiliser dépend du gain recherché et de l'efficacité de la classification et du diagnostic. Ainsi, si la classification basée sur les images 4:2:0 est considérée comme acceptable, l'utilisation du schéma (444,420), pour gagner le maximum d'énergie, est le meilleur choix. Cependant, si nous voulons garantir la meilleure qualité pour toutes les images, l'utilisation du schéma (444,422) est préférable. Dans les deux cas, il y a toujours un gain.

Chapitre 4 : Système de compression des images WCE proposé

Un autre facteur qui affecte le CR est que si le sous-échantillonnage réduit le nombre de pixels, il peut entraîner une compression réduite des pixels restants. Cet effet est principalement observé dans les images anormales codées sans perte, où le codage adaptatif réduit sa compression. Nous avons objecté dans la section 4.3.1 que, en moyenne, la compression de la séquence entière est augmentée.

Pour vérifier l'efficacité du gain de compression du système proposé, nous avons effectué un autre test dans lequel nous codons toutes les images (normales et anormales) avec le même sous-échantillonnage, et nous avons comparé les résultats à notre technique de codage adaptatif.

Tableau 4. 4 Une comparaison du CR et de la qualité des différents schémas de sous-échantillonnage

	444		422		420		Proposé (444,422)	
	CR (%)	PSNR(UV) (db)	CR (%)	PSNR(UV) (db)	CR (%)	PSNR(UV) (db)	CR (%)	PSNR(UV) (db)
Test 1	66.30	∞	74.89	55.26	80.32	31.51	73.81	55.29
Test 2	63.86	∞	72.46	55.77	77.87	31.54	68.37	55.64
Test 3	64.10	∞	73.10	48.46	78.86	31.23	71.49	48.49
Test 4	66.91	∞	75.35	57.72	80.47	31.50	70.59	57.71
Test 5	67.18	∞	75.65	52.00	81.25	31.57	68.16	52.25
Test 6	64.78	∞	73.33	55.24	78.63	31.51	71.85	55.25

Les résultats sont présentés dans le tableau 4.4.

Les techniques de l'état de l'art utilisent soit un codage fixe sans perte, soit un codage avec perte [8], [118], [126]. Dans le tableau 4.4, les résultats notés 444 donneront les mêmes résultats que le système de codage sans perte trouvés dans [118], tandis que les résultats 422 et 420 sont pour un codage avec perte. Les schémas fixes donneront une compression similaire à celle de notre système lorsqu'il considère que toutes les images sont importantes, ou inversement toutes les images sont non importantes.

Les résultats montrent que notre système obtient un gain de CR, dans tous les tests, par rapport à la compression sans perte 444, qui a le CR le plus faible. La compression avec perte 420 donne les meilleurs résultats en termes de CR, mais elle présente le PSNR (UV) le plus faible dans tous les tests. La compression avec sous-échantillonnage 422 donne des résultats élevés en termes de PSNR par rapport à la compression 420. Comme prévu, les schémas qui codent uniquement 422 images ou uniquement 420 images fournissent un bon CR par rapport à notre système. C'est normal puisque toutes les images sont sous-échantillonnées en 420 ou 422.

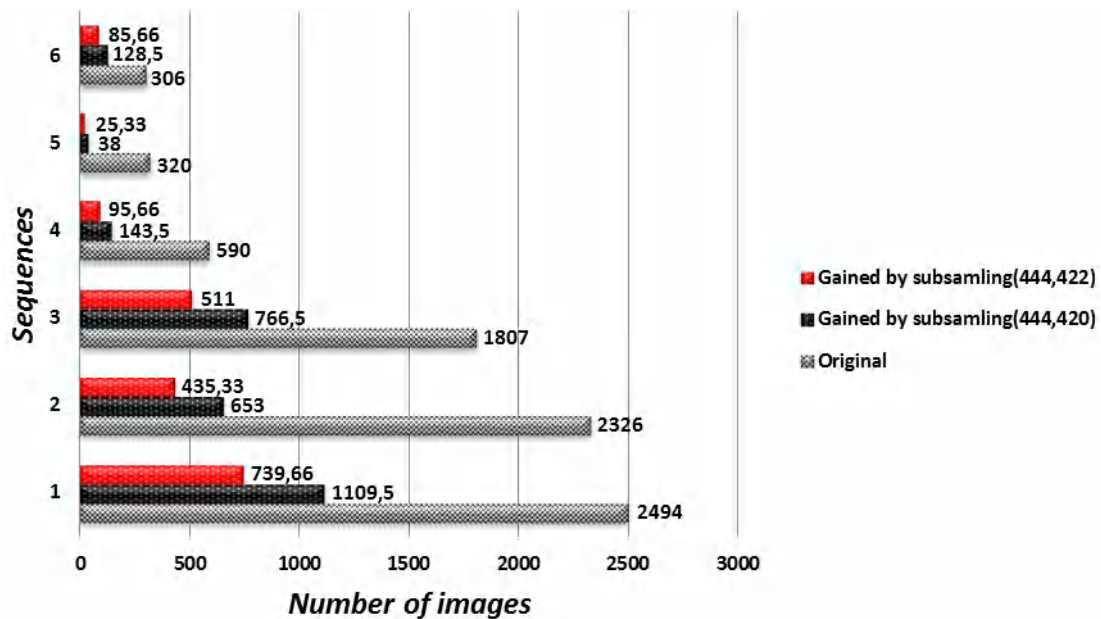


Figure 4.8 Nombre d'images 444 gagnées qui sont obtenues par sous-échantillonnage (444,422) et (444,420), avant compression pour chaque séquence de test.

Cependant, les images décodées avec perte (importantes) sont de qualité inférieure et peuvent donc affecter la précision du diagnostic, alors que notre système augmente la compression tout en conservant une qualité élevée pour les images importantes.

Le pire des cas de notre système de sous-échantillonnage adaptatif intelligent est celui où toutes les images sont importantes. Dans ce cas, le meilleur choix aurait été d'utiliser un simple codage 444 DPCM (66,30%), en une seule étape (non adaptatif). Par exemple, dans le cas de la séquence 1, notre codage adaptatif en deux étapes produira un CR de 65,20%. Ce taux est légèrement inférieur à celui du codage DPCM 444 simple, qui est de 66,30%. Cette perte de 1,1% en compression est le coût pour utiliser une compression à deux couches. En pratique, nous savons que la plupart des images ne sont pas importantes et que le pire scénario ne se produira jamais. Ainsi, quel que soit le nombre d'images importantes, nous pouvons être sûrs que notre système ne risque pas d'augmenter le nombre de bits.

Dans le test final, nous allons calculer le nombre d'images gagnées et comment nous pouvons les utiliser pour augmenter la fréquence d'images. Si nous considérons que la consommation d'énergie est proportionnelle au nombre de bits transmis, le pourcentage de bits économisés est le même pourcentage d'énergie économisée de la WCE. Puisque notre objectif est d'utiliser les bits sauvegardés pour coder des images supplémentaires, notre gain peut également être mesuré en fonction du nombre d'images que nous avons ajoutées. L'utilisation d'un schéma de sous-échantillonnage fixe, avant la compression, nous permettra de gagner 33% ou même 50%

Chapitre 4 : Système de compression des images WCE proposé

d'images supplémentaires pour le sous-échantillonnage 422 et 420, respectivement. Par exemple, dans la séquence 1, cela signifie que les pixels de chrominance rejetés sont égaux à ceux de 831,33 ou 1247 images, respectivement. Dans notre système de test, ce nombre varie et dépend du nombre d'images non importantes calculé par l'équation (4.4).

La figure 4.10 montre le nombre d'images gagnées pour chaque séquence de test, avec différents sous-échantillonnages. Ici, nous considérons que les pixels écartés par notre sous-échantillonnage adaptatif sont utilisés pour coder des images supplémentaires (444). Par exemple, dans la séquence 1, nous avons à l'origine 2494 images et après sous-échantillonnage, nous pouvons utiliser les bits ou pixels gagnés pour coder 1109,5 images supplémentaires de la même taille que les 444 images. Toutefois, ces résultats ne sont corrects que si la transmission est effectuée sans compression. En outre, les images gagnées elles-mêmes ne seront pas nécessairement codées comme 444 et un sous-échantillonnage supplémentaire sera effectué. Par conséquent, nous avons calculé le nombre moyen d'images gagnées après compression.

Le tableau 4.5 montre le nombre d'images gagnées qui sont réellement sauvegardées par notre système pour les six séquences de test. Pour la séquence 1, le nombre des images gagnées après compression pour le schéma (444,420) sont 1450.1 contre 1109.5 avant compression. Comme nous pouvons le constater, il y a un gain supplémentaire dû à la compression. La différence provient uniquement du codage de la chrominance, car la luminance (Y) est la même après la compression. Comme les pixels de la chrominance sont largement corrélés, la prédiction peut les coder efficacement même après sous-échantillonnage. Cependant, pour la compression 444, si la plupart des valeurs résiduelles de prédiction sont petites, elles seront codées avec quelques bits par rapport aux zéro bit avec le sous-échantillonnage.

Tableau 4. 5 Bits gagnés par le système de compression proposé

<i>WCE test</i>	<i>Images gagnées après compression</i>				
	<i>444</i>	<i>422</i>	<i>420</i>	<i>Proposé (444,422)</i>	<i>Proposé (444,420)</i>
<i>Test 1</i>	0	853.18	1776.7	715.15	1450.1
<i>Test 2</i>	0	726.34	1472.5	331.65	632.87
<i>Test 3</i>	0	604.57	1261.7	468.38	945.28
<i>Test 4</i>	0	202.01	409.64	73.82	141.20
<i>Test 5</i>	0	111.31	240.12	9.84	30.08
<i>Test 6</i>	0	98.09	198.32	76.85	159.54

Chapitre 4 : Système de compression des images WCE proposé

Dans l'expérience finale, nous voulons montrer comment nous pouvons améliorer la précision du diagnostic d'une région spécifique, en augmentant la qualité de l'image. A titre d'exemple, nous avons analysé une longue séquence et calculé comment nous pouvons augmenter le taux d'images dans la région d'intérêt. Dans notre test, nous avons utilisé une séquence de test qui contient 7525 images. Les images de cette séquence ont été automatiquement classées, lors du codage, pour détecter la région d'intérêt. Nous avons sélectionné la région du colon comme région d'intérêt. La Figure 4.11 montre la position des images en fonction de leur classification. Bien qu'il y ait trois fausses décisions de la classification montrant quelques images isolées qui ont été considérées comme importantes, la région d'intérêt est clairement identifiée. Cela indique également que l'identification d'une région donnée doit être basée sur plusieurs images et non sur une seule.

Les résultats de la classification indiquent que 2539 images consécutives sont classées comme colon, 520 images d'estomac et 4466 images d'intestin grêle. Sur la base de ces résultats, nous avons codé les images de la région d'intérêt de haute qualité (444), tandis que les parties qui

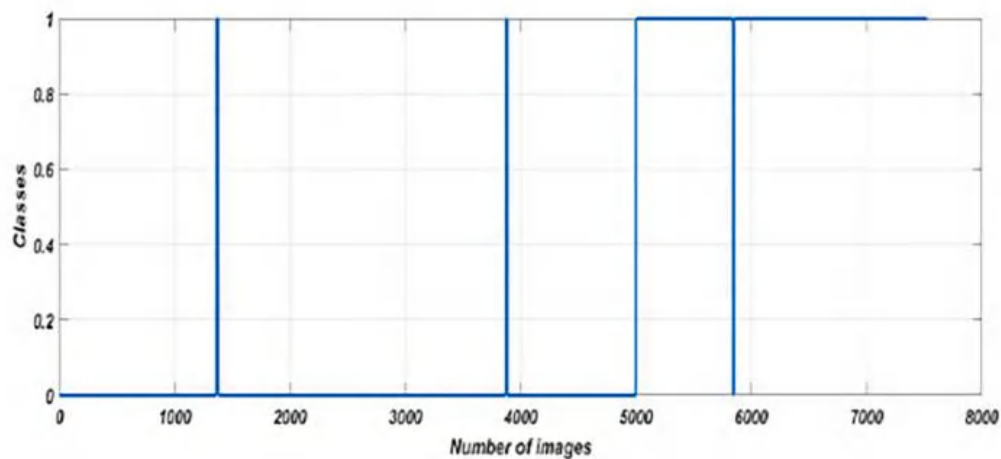


Figure 4. 9 Un exemple des positions des images classées, pour la séquence anatomique test. La classe 1 correspond à la région d'intérêt (côlon) et la classe 0 aux autres régions (estomac, intestin grêle).

ne présentent pas d'intérêt ont été codées avec une qualité faible (422 ou 420). Sur la base des résultats de la compression, nous avons calculé le nombre d'images gagnées, comme ci-dessus. Dans ce cas, ces images supplémentaires ont été utilisées pour augmenter le taux de trame des images, dans la région d'intérêt. Dans le test, la région d'intérêt contient 2539 images et nous avons codé un total de 7525 images. Le sous-échantillonnage des 4986 autres images (codées en 420) nous a permis de gagner 2493 images supplémentaires. Ces 2493 images peuvent être ajoutées aux 2539 pour avoir 5032 images dans la région d'intérêt. Cela

signifie que la fréquence d'images dans la région d'intérêt peut être augmentée de 2 fps à 3,96 ou environ 4 fps.

En pratique, la taille exacte de la région d'intérêt ne sera pas connue par notre système, jusqu'à ce qu'il atteigne sa fin. Par conséquent, la taille de la région et la nouvelle fréquence d'images correspondante doivent être fixées à partir des données statistiques de l'anatomie humaine. Le rôle de notre système est uniquement de déterminer le début et la fin de la région d'intérêt pendant l'opération de codage.

4.5 Conclusion

Dans ce chapitre, nous avons proposé une nouvelle approche de la compression d'images spécialement conçue pour la WCE. Nous avons utilisé un algorithme de compression peu compliqué qui est combiné avec une classification intelligente. En fonction des résultats de la classification, les images non importantes sont codées avec une faible qualité, tandis que les images importantes sont enregistrées avec leur taille et leur qualité d'origine. La classification est effectuée par le décodeur et utilisée pour contrôler le codeur en utilisant le canal de retour. Le but d'un tel système est d'augmenter la durée de vie de la batterie ou d'obtenir des images de haute qualité dans des zones spécifiques. D'après les résultats obtenus, nous concluons que notre système proposé est efficace et offre une bonne optimisation énergétique pour WCE.

Conclusion générale

Le but de cette thèse est de développer des algorithmes de traitement automatique des images WCE, pour aider les médecins dans l'examen fastidieux de ces images. Ainsi que pour la résolution des problèmes de la limitation énergétique de la WCE. Pour cela on a proposé dans la première contribution de classer automatiquement les images WCE en deux classes (normal ou abnormal) dans le but de détecter automatiquement les lésions GI, et aussi en trois classes (stomach, small-intestine et colon) dans le cas où les médecins sont concentrés sur une seule partie du tube digestif. Cette classification a été réalisée en utilisant deux techniques d'apprentissage : la première est l'apprentissage à partir de zéros, dans ce cas on a proposé un réseau CNN composé de trois couches de convolution, trois couches de mise en commun et deux couches entièrement connectées. La deuxième technique concerne l'apprentissage par transfert par le réentraînement des réseaux CNN standards. Les résultats obtenus par les deux méthodes de classification sont très satisfaisants.

On a aussi proposé une autre méthode pour la classification des différentes maladies GI dans les images WCE, il s'agit d'un algorithme d'apprentissage en profondeur basé sur la combinaison de deux CNN pré-entraînés, à savoir MobileNet et DenseNet169. Les caractéristiques extraites produites par la méthode proposée sont ensuite introduites dans un classificateur MLP. Cette combinaison permet d'obtenir davantage de caractéristiques et par conséquent des meilleurs résultats par rapport au classificateur individuel.

D'autre part, afin d'augmenter la durée de vie de la CE, il faut assurer une bonne gestion de l'énergie disponible. Un des moyens efficaces pour la gestion d'énergie est l'utilisation des algorithmes de traitement et de communication efficaces, car une grande partie de l'énergie de la batterie est consommée par la transmission d'images de la capsule au récepteur. L'énergie de transmission dépend du nombre de bits transmis. Par conséquent, la réduction de la quantité de bits transmis (par compression) réduira l'énergie consommée. Alors, dans la deuxième contribution, on a présenté un système de compression intelligent, qui répond aux problèmes de limitations énergétiques de la WCE. Le principe est d'inclure une boucle de rétroaction de classification, basé sur l'apprentissage profond, pour déterminer l'importance des images transmises. Cette classification est utilisée avec un algorithme de compression peu compliqué, basée sur la prédiction pour permettre une gestion intelligente de l'énergie limitée de la capsule. Pour cela, la capsule commence par transmettre une version sous-échantillonnée de chaque image avec un faible taux. Les images seront décodées et classées, automatiquement, pour

Conclusion générale

détecter d'éventuelles lésions. Suite au résultat de la classification, les images considérées comme importantes pour le diagnostic seront enrichies d'un contenu supplémentaire, tandis que les images moins importantes seront enregistrées avec une faible qualité. De cette façon, une grande quantité de bits a été économisée, sans affecter le diagnostic. L'énergie économisée peut être utilisée pour prolonger la durée de vie de la capsule ou pour augmenter la résolution et la fréquence d'images de certaines images WCE.

Nous avons également proposé d'utiliser la classification des images des organes GI (estomac, colon, intestin grêle, etc.) pour contrôler la qualité et le débit d'images dans les cas où les médecins se concentrent sur des zones spécifiques. Lorsque le système détecte que la capsule a atteint la région ciblée, le décodeur envoie un signal de commande à la capsule pour augmenter temporairement la qualité de l'image et la fréquence d'images. Si les images ne proviennent pas de la région ciblée, la capsule va réduire sa fréquence d'images au taux faible de 2 à 4 images par seconde. D'après les résultats obtenus, nous concluons que notre système proposé est efficace et offre une bonne optimisation énergétique pour WCE.

Enfin, l'utilisation d'un système d'imagerie intelligent permet d'effectuer un pré-diagnostic ou une pré-sélection pour tous les systèmes d'imagerie médicale. Il y a toujours des contraintes de ressources qui ne permettent pas de prendre un grand nombre d'images. Il y a un compromis entre l'imagerie, le diagnostic et les images recueillies ne sont pas nécessairement les bonnes ou les meilleures. En outre, l'analyse des images après leur prise prend beaucoup de temps aux médecins.

Dans les travaux futurs, on essaiera de proposer des nouveaux systèmes pour améliorer la résolution des images WCE, pour un diagnostic plus précis, cela se fait par la proposition des nouvelles techniques d'apprentissage automatique.

Bibliographies

- [1] R. Eliakim, A. Suissa, K. Yassin, D. Katz, et D. Fischer, « Wireless capsule video endoscopy compared to barium follow-through and computerised tomography in patients with suspected Crohn's disease—final report », *Digestive and Liver Disease*, vol. 36, n° 8, p. 519-522, août 2004, doi: 10.1016/j.dld.2004.03.011.
- [2] « IntroMedic ». <http://www.intromedic.com/eng/main/>.
- [3] G. Iddan, G. Meron, A. Glukhovsky, et P. Swain, « Wireless capsule endoscopy », *Nature*, vol. 405, n° 6785, Art. n° 6785, mai 2000, doi: 10.1038/35013140.
- [4] « Welcome to Olympus », *Olympus*. <https://www.olympus-europa.com/>.
- [5] « RF - The Next Generation Capsule Endoscope Sayaka ». <http://www.rfsystemlab.com/en/sayaka/index.html>.
- [6] X. Chen *et al.*, « A Wireless Capsule Endoscope System With Low-Power Controlling and Processing ASIC », *IEEE Transactions on Biomedical Circuits and Systems*, vol. 3, n° 1, p. 11-22, févr. 2009, doi: 10.1109/TBCAS.2008.2006493.
- [7] C. M. Sadler et M. Martonosi, « Data compression algorithms for energy-constrained devices in delay tolerant networks », in *Proceedings of the 4th international conference on Embedded networked sensor systems*, New York, NY, USA, oct. 2006, p. 265-278. doi: 10.1145/1182807.1182834.
- [8] M. W. Alam, Md. M. Hasan, S. K. Mohammed, F. Deeba, et K. A. Wahid, « Are Current Advances of Compression Algorithms for Capsule Endoscopy Enough? A Technical Review », *IEEE Reviews in Biomedical Engineering*, vol. 10, p. 26-43, 2017, doi: 10.1109/RBME.2017.2757013.
- [9] Y.-G. Lee et G. Yoon, « Bleeding Detection Algorithm for Capsule Endoscopy », *International Journal of Biomedical and Biological Engineering*, vol. 5, n° 9, p. 388-393, sept. 2011.
- [10] A. S. Vemuri, « Survey of Computer Vision and Machine Learning in Gastrointestinal Endoscopy », *arXiv:1904.13307 [physics]*, avr. 2019. [En ligne]. Disponible sur: <http://arxiv.org/abs/1904.13307>
- [11] L. Shen *et al.*, « Management of gastric cancer in Asia: resource-stratified guidelines », *The Lancet Oncology*, vol. 14, n° 12, p. e535-e547, nov. 2013, doi: 10.1016/S1470-2045(13)70436-4.
- [12] G. Iddan, G. Meron, A. Glukhovsky, et P. Swain, « Wireless capsule endoscopy », *Nature*, vol. 405, n° 6785, p. 417-417, mai 2000, doi: 10.1038/35013140.

Bibliographies

- [13] K. M. Reavis et W. S. Melvin, « Advanced endoscopic technologies », *Surg Endosc*, vol. 22, n° 6, p. 1533-1546, juin 2008, doi: 10.1007/s00464-008-9831-1.
- [14] G. J. Iddan et C. P. Swain, « History and development of capsule endoscopy », *Gastrointestinal Endoscopy Clinics*, vol. 14, n° 1, p. 1-9, janv. 2004, doi: 10.1016/j.giec.2003.10.022.
- [15] A. Loeve, P. Breedveld, et J. Dankelman, « Scopes Too Flexible...and Too Stiff », *IEEE Pulse*, vol. 1, n° 3, p. 26-41, nov. 2010, doi: 10.1109/MPUL.2010.939176.
- [16] « Wireless capsule endoscopy: from diagnostic devices to multipurpose robotic systems | SpringerLink ». <https://link.springer.com/article/10.1007/s10544-006-9025-3>.
- [17] P. Valdastri, R. J. Webster, C. Quaglia, M. Quirini, A. Menciassi, et P. Dario, « A New Mechanism for Mesoscale Legged Locomotion in Compliant Tubular Environments », *IEEE Transactions on Robotics*, vol. 25, n° 5, p. 1047-1057, oct. 2009, doi: 10.1109/TRO.2009.2014127.
- [18] R. Eliakim et V. K. Sharma, « Esophageal Capsule Endoscopy », in *Video Capsule Endoscopy: A Reference Guide and Atlas*, M. Keuchel, F. Hagenmüller, et H. Tajiri, Éd. Berlin, Heidelberg: Springer, 2014, p. 437-445. doi: 10.1007/978-3-662-44062-9_42.
- [19] T. Rahim, M. A. Usman, et S. Y. Shin, « A survey on contemporary computer-aided tumor, polyp, and ulcer detection methods in wireless capsule endoscopy imaging », *Computerized Medical Imaging and Graphics*, vol. 85, p. 101767, oct. 2020, doi: 10.1016/j.compmedimag.2020.101767.
- [20] P. Swain, « At a watershed? Technical developments in wireless capsule endoscopy », *Journal of Digestive Diseases*, vol. 11, n° 5, p. 259-265, 2010, doi: 10.1111/j.1751-2980.2010.00448.x.
- [21] M. Keuchel, F. Hagenmüller, et H. Tajiri, Éd., *Video Capsule Endoscopy: A Reference Guide and Atlas*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014. doi: 10.1007/978-3-662-44062-9.
- [22] « First clinical trial of the “MiRo” capsule endoscope by using a novel transmission technology: electric-field propagation - ScienceDirect ». <https://www.sciencedirect.com/science/article/abs/pii/S0016510708017781>.
- [23] M. W. Alam, M. H. A. Sohag, A. H. Khan, T. Sultana, et K. A. Wahid, « IoT-Based Intelligent Capsule Endoscopy System: A Technical Review », in *Intelligent Data Analysis for Biomedical Applications*, Elsevier, 2019, p. 1-20. doi: 10.1016/B978-0-12-815553-0.00001-X.

Bibliographies

- [24] T. Nakamura et A. Terano, « Capsule endoscopy: past, present, and future », *J Gastroenterol*, vol. 43, n° 2, p. 93-99, févr. 2008, doi: 10.1007/s00535-007-2153-6.
- [25] E. Ben-Soussan, G. Savoye, M. Antonietti, S. Ramirez, P. Ducrotté, et E. Lerebours, « Is a 2-Liter PEG Preparation Useful Before Capsule Endoscopy? », *Journal of Clinical Gastroenterology*, vol. 39, n° 5, p. 381-384, juin 2005, doi: 10.1097/01.mcg.0000159271.43233.45.
- [26] N. Dai, C. Gubler, P. Hengstler, C. Meyenberger, et P. Bauerfeind, « Improved capsule endoscopy after bowel preparation », *Gastrointestinal Endoscopy*, vol. 61, n° 1, p. 28-31, janv. 2005, doi: 10.1016/S0016-5107(04)02444-7.
- [27] V. K. Sharma, R. Eliakim, P. Sharma, et D. Faigel, « ICCE Consensus for Esophageal Capsule Endoscopy », *Endoscopy*, vol. 37, n° 10, p. 1060-1064, oct. 2005, doi: 10.1055/s-2005-870311.
- [28] C. Spada *et al.*, « Meta-analysis Shows Colon Capsule Endoscopy Is Effective in Detecting Colorectal Polyps », *Clinical Gastroenterology and Hepatology*, vol. 8, n° 6, p. 516-522.e8, juin 2010, doi: 10.1016/j.cgh.2010.02.018.
- [29] C. Spada *et al.*, « Second-generation colon capsule endoscopy compared with colonoscopy », *Gastrointestinal Endoscopy*, vol. 74, n° 3, p. 581-589.e1, sept. 2011, doi: 10.1016/j.gie.2011.03.1125.
- [30] R. Eliakim *et al.*, « Prospective multicenter performance evaluation of the second-generation colon capsule compared with colonoscopy », *Endoscopy*, vol. 41, n° 12, p. 1026-1031, déc. 2009, doi: 10.1055/s-0029-1215360.
- [31] J. Park, Y. K. Cho, et J. H. Kim, « Current and Future Use of Esophageal Capsule Endoscopy », *Clin Endosc*, vol. 51, n° 4, p. 317-322, juill. 2018, doi: 10.5946/ce.2018.101.
- [32] Y. Yuan, B. Li, et M. Q.-H. Meng, « WCE Abnormality Detection Based on Saliency and Adaptive Locality-Constrained Linear Coding », *IEEE Transactions on Automation Science and Engineering*, vol. 14, n° 1, p. 149-159, janv. 2017, doi: 10.1109/TASE.2016.2610579.
- [33] L. Ke et M. W. Marcellin, « Near-lossless image compression: minimum-entropy, constrained-error DPCM », *IEEE Transactions on Image Processing*, vol. 7, n° 2, p. 225-228, févr. 1998, doi: 10.1109/83.660999.
- [34] W. B. Pennebaker et J. L. Mitchell, *JPEG: Still Image Data Compression Standard*. Springer Science & Business Media, 1992.

Bibliographies

- [35] D. Turgis et R. Puers, « Image compression in video radio transmission for capsule endoscopy », *Sensors and Actuators A: Physical*, vol. 123-124, p. 129-136, sept. 2005, doi: 10.1016/j.sna.2005.05.016.
- [36] D. A. Clunie, « Lossless compression of grayscale medical images: effectiveness of traditional and state-of-the-art approaches », in *Medical Imaging 2000: PACS Design and Evaluation: Engineering and Clinical Issues*, mai 2000, vol. 3980, p. 74-84. doi: 10.1117/12.386389.
- [37] S.-L. Chen, T.-Y. Liu, C.-W. Shen, et M.-C. Tuan, « VLSI Implementation of a Cost-Efficient Near-Lossless CFA Image Compressor for Wireless Capsule Endoscopy », *IEEE Access*, vol. 4, p. 10235-10245, 2016, doi: 10.1109/ACCESS.2016.2638475.
- [38] G. Liu, G. Yan, B. Zhu, et L. Lu, « Design of a video capsule endoscopy system with low-power ASIC for monitoring gastrointestinal tract », *Med Biol Eng Comput*, vol. 54, n° 11, p. 1779-1791, nov. 2016, doi: 10.1007/s11517-016-1472-2.
- [39] M. J. Weinberger, G. Seroussi, et G. Sapiro, « LOCO-I: a low complexity, context-based, lossless image compression algorithm », in *Proceedings of Data Compression Conference - DCC '96*, mars 1996, p. 140-149. doi: 10.1109/DCC.1996.488319.
- [40] M. J. Weinberger, G. Seroussi, et G. Sapiro, « The LOCO-I lossless image compression algorithm: principles and standardization into JPEG-LS », *IEEE Transactions on Image Processing*, vol. 9, n° 8, p. 1309-1324, août 2000, doi: 10.1109/83.855427.
- [41] C. C. Cutler, « Differential quantization of communication signals », US2605361A, 29 juillet 1952. [En ligne]. Disponible sur: <https://patents.google.com/patent/US2605361A/en>
- [42] « DPCM ». <https://einstein.informatik.uni-oldenburg.de/rechnernetze/dpcm.htm>.
- [43] T. H. Khan et K. A. Wahid, « Low Power and Low Complexity Compressor for Video Capsule Endoscopy », *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, n° 10, p. 1534-1546, oct. 2011, doi: 10.1109/TCSVT.2011.2163985.
- [44] T. H. Khan et K. A. Wahid, « Design of a Lossless Image Compression System for Video Capsule Endoscopy and Its Performance in In-Vivo Trials », *Sensors*, vol. 14, n° 11, Art. n° 11, nov. 2014, doi: 10.3390/s141120779.
- [45] C. Hu, M. Q.-H. Meng, L. Liu, Y. Pan, et Z. Liu, « Image representation and compression for capsule endoscope robot », in *2009 International Conference on Information and Automation*, juin 2009, p. 506-511. doi: 10.1109/ICINFA.2009.5204976.

Bibliographies

- [46] N. Ahmed, T. Natarajan, et K. R. Rao, « Discrete Cosine Transform », *IEEE Transactions on Computers*, vol. C-23, n° 1, p. 90-93, janv. 1974, doi: 10.1109/T-C.1974.223784.
- [47] A. M. Raid, W. M. Khedr, M. A. El-dosuky, et W. Ahmed, « Jpeg Image Compression Using Discrete Cosine Transform - A Survey », *arXiv:1405.6147 [cs]*, mai 2014. [En ligne]. Disponible sur: <http://arxiv.org/abs/1405.6147>
- [48] G. K. Wallace, « The JPEG still picture compression standard », *IEEE Transactions on Consumer Electronics*, vol. 38, n° 1, p. xviii-xxxiv, févr. 1992, doi: 10.1109/30.125072.
- [49] M.-C. Lin et L.-R. Dung, « A Subsample-Based Low-Power Image Compressor for Capsule Gastrointestinal Endoscopy », *EURASIP Journal on Advances in Signal Processing*, vol. 1, n° 2011, p. 1-15, 2011, doi: 10.1155/2011/257095.
- [50] K. Wahid, S.-B. Ko, et D. Teng, « Efficient hardware implementation of an image compressor for wireless capsule endoscopy applications », in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, juin 2008, p. 2761-2765. doi: 10.1109/IJCNN.2008.4634186.
- [51] M.-C. Lin, L.-R. Dung, et P.-K. Weng, « An ultra-low-power image compressor for capsule endoscope », *BioMedical Engineering OnLine*, vol. 5, n° 1, p. 14, févr. 2006, doi: 10.1186/1475-925X-5-14.
- [52] M.-C. Lin, L.-R. Dung, et P.-K. Weng, « A cardinal image compressor for capsule endoscope », in *2006 IEEE Biomedical Circuits and Systems Conference*, nov. 2006, p. 146-149. doi: 10.1109/BIOCAS.2006.4600329.
- [53] P. Turcza et M. Duplaga, « Hardware-Efficient Low-Power Image Processing System for Wireless Capsule Endoscopy », *IEEE Journal of Biomedical and Health Informatics*, vol. 17, n° 6, p. 1046-1056, nov. 2013, doi: 10.1109/JBHI.2013.2266101.
- [54] H. A. Peterson, H. Peng, J. H. Morgan, et W. B. Pennebaker, « Quantization of color image components in the DCT domain », in *Human Vision, Visual Processing, and Digital Display II*, juin 1991, vol. 1453, p. 210-222. doi: 10.1117/12.44357.
- [55] S. Aisyah et J. S. Mandeep, « Image processing using DCT and wavelet transform », *Optoelectronics and Advanced Materials - Rapid Communications*, vol. 6, n° January-February 2012, p. 29-35, févr. 2012.
- [56] A. Mostafa, K. Wahid, et S.-B. Ko, « An efficient YCgCo-based image compression algorithm for capsule endoscopy », in *14th International Conference on Computer and Information Technology (ICCIT 2011)*, déc. 2011, p. 219-222. doi: 10.1109/ICCITechn.2011.6164787.

Bibliographies

- [57] P. Turcza et M. Duplaga, « Low-Power Image Compression for Wireless Capsule Endoscopy », in *2007 IEEE International Workshop on Imaging Systems and Techniques*, mai 2007, p. 1-4. doi: 10.1109/IST.2007.379586.
- [58] M. Elhaji, N. Jarray, et A. Zitouni, « An optimized DCT compressor based on Cordic-Loeffler approach for wireless endoscopic capsule », *International Journal of ADVANCED AND APPLIED SCIENCES*, vol. 4, p. 59-65, juill. 2017.
- [59] A. L. Samuel, « Some studies in machine learning using the game of checkers », *IBM Journal of Research and Development*, vol. 44, n° 1.2, p. 206-226, janv. 2000, doi: 10.1147/rd.441.0206.
- [60] R. S. Michalski, J. G. Carbonell, et T. M. Mitchell, *Machine Learning: An Artificial Intelligence Approach*. Springer Science & Business Media, 2013.
- [61] « Introduction à la régression linéaire simple ». <http://www.josephazar.net/blog/mlsreg.php>.
- [62] E. Hatipoglu, « Machine Learning — Classification — Logistic Regression — Part 8 », *Medium*, 11 juillet 2018. <https://medium.com/@ekrem.hatipoglu/machine-learning-classification-logistic-regression-part-8-b77d2a61aae1>.
- [63] O. M. Parkhi, A. Vedaldi, et A. Zisserman, « Deep Face Recognition », in *Proceedings of the British Machine Vision Conference 2015*, Swansea, 2015, p. 41.1-41.12. doi: 10.5244/C.29.41.
- [64] K. He, X. Zhang, S. Ren, et J. Sun, « Deep Residual Learning for Image Recognition », présenté à *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, p. 770-778. [En ligne]. Disponible sur: https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html
- [65] C. Szegedy *et al.*, « Going Deeper With Convolutions », présenté à *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, p. 1-9. [En ligne]. Disponible sur: https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Szegedy_Going_Deeper_With_2015_CVPR_paper.html
- [66] R. Collobert et J. Weston, « A unified architecture for natural language processing: deep neural networks with multitask learning », in *Proceedings of the 25th international conference on Machine learning*, New York, NY, USA, juill. 2008, p. 160-167. doi: 10.1145/1390156.1390177.

Bibliographies

- [67] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, et K. J. Lang, « Phoneme recognition using time-delay neural networks », *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, n° 3, p. 328-339, mars 1989, doi: 10.1109/29.21701.
- [68] D. Amodei *et al.*, « Deep Speech 2 : End-to-End Speech Recognition in English and Mandarin », in *Proceedings of The 33rd International Conference on Machine Learning*, juin 2016, p. 173-182. [En ligne]. Disponible sur: <https://proceedings.mlr.press/v48/amodei16.html>
- [69] D. Silver *et al.*, « Mastering the game of Go without human knowledge », *Nature*, vol. 550, n° 7676, p. 354-359, oct. 2017, doi: 10.1038/nature24270.
- [70] Y. Djeriri, « Les Réseaux de Neurones Artificiels », 2017.
- [71] R. Delassus, « Apprentissage automatique pour la détection d'anomalies dans les données ouvertes: application à la cartographie », p. 158.
- [72] C. Touzet, « LES RESEAUX DE NEURONES ARTIFICIELS, INTRODUCTION AU CONNEXIONNISME », p. 130.
- [73] A. Krenker, J. Bester, et A. Kos, « Introduction to the Artificial Neural Networks », in *Artificial Neural Networks - Methodological Advances and Biomedical Applications*, K. Suzuki, Éd. InTech, 2011. doi: 10.5772/15751.
- [74] F. Sur, « Introduction à l'apprentissage automatique », p. 172.
- [75] J. Duchi, E. Hazan, et Y. Singer, « Adaptive Subgradient Methods for Online Learning and Stochastic Optimization », p. 39.
- [76] M. D. Zeiler, « ADADELTA: An Adaptive Learning Rate Method ». arXiv, 22 décembre 2012. doi: 10.48550/arXiv.1212.5701.
- [77] D. P. Kingma et J. Ba, « Adam: A Method for Stochastic Optimization ». arXiv, 29 janvier 2017. doi: 10.48550/arXiv.1412.6980.
- [78] C. Hardy, « Contribution au développement de l'apprentissage profond dans les systèmes distribués », p. 152.
- [79] « Underfitting vs. Overfitting », *scikit-learn*. https://scikit-learn/stable/auto_examples/model_selection/plot_underfitting_overfitting.html.
- [80] A. Dertat, « Applied Deep Learning - Part 4: Convolutional Neural Networks », *Medium*, 13 novembre 2017. <https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2>.
- [81] O. Russakovsky *et al.*, « ImageNet Large Scale Visual Recognition Challenge », *Int J Comput Vis*, vol. 115, n° 3, p. 211-252, déc. 2015, doi: 10.1007/s11263-015-0816-y.

Bibliographies

- [82] Y. Lecun, L. Bottou, Y. Bengio, et P. Haffner, « Gradient-based learning applied to document recognition », *Proceedings of the IEEE*, vol. 86, n° 11, p. 2278-2324, nov. 1998, doi: 10.1109/5.726791.
- [83] « Evolution of CNN Architectures: LeNet, AlexNet, ZFNet, GoogleNet, VGG and ResNet », *OpenGenus IQ: Computing Expertise & Legacy*, 21 janvier 2019. <https://iq.opengenus.org/evolution-of-cnn-architectures/>.
- [84] D. Cornelisse, « An intuitive guide to Convolutional Neural Networks », *We've moved to freeCodeCamp.org/news*, 24 avril 2018. <https://medium.com/free-code-camp/an-intuitive-guide-to-convolutional-neural-networks-260c2de0a050>.
- [85] S. Saha, « A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way », *Medium*, 17 décembre 2018. <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>.
- [86] A. Krizhevsky, I. Sutskever, et G. E. Hinton, « ImageNet Classification with Deep Convolutional Neural Networks », in *Advances in Neural Information Processing Systems*, 2012, vol. 25. [En ligne]. Disponible sur: <https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html>
- [87] K. Simonyan et A. Zisserman, « Very Deep Convolutional Networks for Large-Scale Image Recognition », *arXiv:1409.1556 [cs]*, avr. 2015. [En ligne]. Disponible sur: <http://arxiv.org/abs/1409.1556>
- [88] A. G. Howard *et al.*, « MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications ». *arXiv*, 16 avril 2017. doi: 10.48550/arXiv.1704.04861.
- [89] G. Huang, Z. Liu, L. van der Maaten, et K. Q. Weinberger, « Densely Connected Convolutional Networks », présenté à Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, p. 4700-4708. [En ligne]. Disponible sur: https://openaccess.thecvf.com/content_cvpr_2017/html/Huang_Densely_Connected_Convolutional_CVPR_2017_paper.html
- [90] R. Mifdal, « Application des techniques d'apprentissage automatique pour la prédiction de la tendance des titres financiers », p. 196.
- [91] « 3.3. Metrics and scoring: quantifying the quality of predictions », *scikit-learn*. https://scikit-learn/stable/modules/model_evaluation.html.
- [92] G. Litjens *et al.*, « A survey on deep learning in medical image analysis », *Medical Image Analysis*, vol. 42, p. 60-88, déc. 2017, doi: 10.1016/j.media.2017.07.005.

Bibliographies

- [93] Y. Zou, L. Li, Y. Wang, J. Yu, Y. Li, et W. J. Deng, « Classifying digestive organs in wireless capsule endoscopy images based on deep convolutional neural network », in *2015 IEEE International Conference on Digital Signal Processing (DSP)*, juill. 2015, p. 1274-1278. doi: 10.1109/ICDSP.2015.7252086.
- [94] Y. Cao, W. Yang, K. Chen, Y. Ren, et Q. Liao, « Capsule Endoscopy Image Classification with Deep Convolutional Neural Networks », in *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, déc. 2018, p. 1584-1588. doi: 10.1109/CompComm.2018.8780859.
- [95] V. S. Sadasivan et C. S. Seelamantula, « High Accuracy Patch-Level Classification of Wireless Capsule Endoscopy Images Using a Convolutional Neural Network », in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, avr. 2019, p. 96-99. doi: 10.1109/ISBI.2019.8759324.
- [96] W. Yang, Y. Cao, Q. Zhao, Y. Ren, et Q. Liao, « Lesion Classification of Wireless Capsule Endoscopy Images », in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, avr. 2019, p. 1238-1242. doi: 10.1109/ISBI.2019.8759577.
- [97] N. Tajbakhsh *et al.*, « Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? », *IEEE Transactions on Medical Imaging*, vol. 35, n° 5, p. 1299-1312, mai 2016, doi: 10.1109/TMI.2016.2535302.
- [98] S. Nadeem, M. A. Tahir, S. S. A. Naqvi, et M. Zaid, « Ensemble of Texture and Deep Learning Features for Finding Abnormalities in the Gastro-Intestinal Tract », in *Computational Collective Intelligence*, Cham, 2018, p. 469-478. doi: 10.1007/978-3-319-98446-9_44.
- [99] H. Alaskar, A. Hussain, N. Al-Aseem, P. Liatsis, et D. Al-Jumeily, « Application of Convolutional Neural Networks for Automated Ulcer Detection in Wireless Capsule Endoscopy Images », *Sensors*, vol. 19, n° 6, Art. n° 6, janv. 2019, doi: 10.3390/s19061265.
- [100] A. Caroppo, A. Leone, et P. Siciliano, « Deep transfer learning approaches for bleeding detection in endoscopy images », *Computerized Medical Imaging and Graphics*, vol. 88, p. 101852, mars 2021, doi: 10.1016/j.compmedimag.2020.101852.
- [101] F. Rustam *et al.*, « Wireless Capsule Endoscopy Bleeding Images Classification Using CNN Based Model », *IEEE Access*, vol. 9, p. 33675-33688, 2021, doi: 10.1109/ACCESS.2021.3061592.

Bibliographies

- [102] F. Fonseca, B. Nunes, M. Salgado, et A. Cunha, « Abnormality classification in small datasets of capsule endoscopy images », *Procedia Computer Science*, vol. 196, p. 469-476, janv. 2022, doi: 10.1016/j.procs.2021.12.038.
- [103] D. Jha *et al.*, « A comprehensive analysis of classification methods in gastrointestinal endoscopy imaging », *Medical Image Analysis*, vol. 70, p. 102007, mai 2021, doi: 10.1016/j.media.2021.102007.
- [104] V. B. S. Prasath, « Polyp Detection and Segmentation from Video Capsule Endoscopy: A Review », *Journal of Imaging*, vol. 3, n° 1, Art. n° 1, mars 2017, doi: 10.3390/jimaging3010001.
- [105] P. H. Smedsrud *et al.*, « Kvasir-Capsule, a video capsule endoscopy dataset », *Sci Data*, vol. 8, n° 1, Art. n° 1, mai 2021, doi: 10.1038/s41597-021-00920-z.
- [106] L. Bai, L. Wang, T. Chen, Y. Zhao, et H. Ren, « Transformer-Based Disease Identification for Small-Scale Imbalanced Capsule Endoscopy Dataset », *Electronics*, vol. 11, n° 17, Art. n° 17, janv. 2022, doi: 10.3390/electronics11172747.
- [107] Z. Hussain, F. Gimenez, D. Yi, et D. Rubin, « Differential Data Augmentation Techniques for Medical Imaging Classification Tasks », *AMIA Annu Symp Proc*, vol. 2017, p. 979-984, avr. 2018.
- [108] K. Chaitanya *et al.*, « Semi-supervised task-driven data augmentation for medical image segmentation », *Medical Image Analysis*, vol. 68, p. 101934, févr. 2021, doi: 10.1016/j.media.2020.101934.
- [109] Y. Zou, L. Li, Y. Wang, J. Yu, Y. Li, et W. J. Deng, « Classifying digestive organs in wireless capsule endoscopy images based on deep convolutional neural network », in *2015 IEEE International Conference on Digital Signal Processing (DSP)*, juill. 2015, p. 1274-1278. doi: 10.1109/ICDSP.2015.7252086.
- [110] N. Tajbakhsh *et al.*, « Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? », *IEEE Transactions on Medical Imaging*, vol. 35, n° 5, p. 1299-1312, mai 2016, doi: 10.1109/TMI.2016.2535302.
- [111] S. Lu, Z. Lu, et Y.-D. Zhang, « Pathological brain detection based on AlexNet and transfer learning », *Journal of Computational Science*, vol. 30, p. 41-47, janv. 2019, doi: 10.1016/j.jocs.2018.11.008.
- [112] P. Turcza et M. Duplaga, « Hardware-Efficient Low-Power Image Processing System for Wireless Capsule Endoscopy », *IEEE Journal of Biomedical and Health Informatics*, vol. 17, n° 6, p. 1046-1056, nov. 2013, doi: 10.1109/JBHI.2013.2266101.

Bibliographies

- [113] Lan-Rong Dung, Yin-Yi Wu, Hsin-Cheng Lai, et Ping-Kuo Weng, « A modified H.264 intra-frame video encoder for capsule endoscope », in *2008 IEEE Biomedical Circuits and Systems Conference*, nov. 2008, p. 61-64. doi: 10.1109/BIOCAS.2008.4696874.
- [114] A. Shabani et S. Timarchi, « Low-power DCT-based compressor for wireless capsule endoscopy », *Signal Processing: Image Communication*, vol. 59, p. 83-95, nov. 2017, doi: 10.1016/j.image.2017.03.003.
- [115] Q. Al-Shebani, P. Premaratne, P. J. Vial, et D. J. McAndrew, « The development of a clinically tested visually lossless Image compression system for capsule endoscopy », *Signal Processing: Image Communication*, vol. 76, p. 135-150, août 2019, doi: 10.1016/j.image.2019.04.008.
- [116] X. Xie, G. Li, X. Chen, X. Li, et Z. Wang, « A Low-Power Digital IC Design Inside the Wireless Endoscopic Capsule », *IEEE Journal of Solid-State Circuits*, vol. 41, n° 11, p. 2390-2400, nov. 2006, doi: 10.1109/JSSC.2006.882884.
- [117] G. Liu, G. Yan, B. Zhu, et L. Lu, « Design of a video capsule endoscopy system with low-power ASIC for monitoring gastrointestinal tract », *Med Biol Eng Comput*, vol. 54, n° 11, p. 1779-1791, nov. 2016, doi: 10.1007/s11517-016-1472-2.
- [118] T. H. Khan et K. A. Wahid, « Lossless and Low-Power Image Compressor for Wireless Capsule Endoscopy », *VLSI Design*, 9 juin 2011. <https://www.hindawi.com/journals/vlsi/2011/343787/>.
- [119] K. A. Fante, B. Bhaumik, et S. Chatterjee, « Design and Implementation of Computationally Efficient Image Compressor for Wireless Capsule Endoscopy », *Circuits Syst Signal Process*, vol. 35, n° 5, p. 1677-1703, mai 2016, doi: 10.1007/s00034-015-0136-z.
- [120] S. K. Mohammed, K. M. M. Rahman, et K. A. Wahid, « Lossless Compression in Bayer Color Filter Array for Capsule Endoscopy », *IEEE Access*, vol. 5, p. 13823-13834, 2017, doi: 10.1109/ACCESS.2017.2726997.
- [121] T. H. Khan et K. A. Wahid, « Design of a Lossless Image Compression System for Video Capsule Endoscopy and Its Performance in In-Vivo Trials », *Sensors*, vol. 14, n° 11, Art. n° 11, nov. 2014, doi: 10.3390/s141120779.
- [122] P. Turcza et M. Duplaga, « Near-lossless energy-efficient image compression algorithm for wireless capsule endoscopy », *Biomedical Signal Processing and Control*, vol. 38, p. 1-8, sept. 2017, doi: 10.1016/j.bspc.2017.04.006.

Bibliographies

- [123] T. H. Khan et K. A. Wahid, « White and narrow band image compressor based on a new color space for capsule endoscopy », *Signal Processing: Image Communication*, vol. 29, n° 3, p. 345-360, mars 2014, doi: 10.1016/j.image.2013.12.001.
- [124] D. Salomon, *Data Compression: The Complete Reference*. Springer Science & Business Media, 2004.
- [125] S. Golomb, « Run-length encodings », *IEEE Transactions on Information Theory*, p. pp.399-401.
- [126] T. H. Khan et K. A. Wahid, « Subsample-based image compression for capsule endoscopy », *J Real-Time Image Proc*, vol. 8, n° 1, p. 5-19, mars 2013, doi: 10.1007/s11554-011-0208-7.
- [127] B. S. Lewis et P. Swain, « Capsule endoscopy in the evaluation of patients with suspected small intestinal bleeding: Results of a pilot study », *Gastrointestinal Endoscopy*, vol. 56, n° 3, p. 349-353, sept. 2002, doi: 10.1016/S0016-5107(02)70037-0.
- [128] R. Hamza, K. Muhammad, Z. Lv, et F. Titouna, « Secure video summarization framework for personalized wireless capsule endoscopy », *Pervasive and Mobile Computing*, vol. 41, p. 436-450, oct. 2017, doi: 10.1016/j.pmcj.2017.03.011.
- [129] X. Chu *et al.*, « Epitomized Summarization of Wireless Capsule Endoscopic Videos for Efficient Visualization », in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*, Berlin, Heidelberg, 2010, p. 522-529. doi: 10.1007/978-3-642-15745-5_64.
- [130] N. Sudha et N. M. Santhiyakumari, « Summarizing wireless capsule endoscopy video frame based on inter-frame structural similarity index », *Advances in Natural and Applied Sciences*, vol. 10, n° 6 SE, p. 199-208, mai 2016.
- [131] D. Fischer, R. Schreiber, D. Levi, et R. Eliakim, « Capsule endoscopy: the localization system », *Gastrointestinal Endoscopy Clinics*, vol. 14, n° 1, p. 25-31, janv. 2004, doi: 10.1016/j.giec.2003.10.020.
- [132] M. A. Khorsandi *et al.*, « Hardware image assessment for wireless endoscopy capsules », in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, août 2016, p. 2050-2053. doi: 10.1109/EMBC.2016.7591130.
- [133] M. A. Khorsandi, N. Karimi, et S. Samavi, « Polyp detection inside the capsule endoscopy: an approach for power consumption reduction », *arXiv:1712.10164 [cs, eess]*, déc. 2017. [En ligne]. Disponible sur: <http://arxiv.org/abs/1712.10164>
- [134] D. E. Boudechiche, S. Benierbah, et M. Khamadja, « Distributed video coding based on vector quantization: Application to capsule endoscopy », *Journal of Visual*

Bibliographies

- Communication and Image Representation*, vol. 49, p. 14-26, nov. 2017, doi: 10.1016/j.jvcir.2017.07.007.
- [135] J. Ahn, H. Nguyen Loc, R. Krishna Balan, Y. Lee, et J. Ko, « Finding Small-Bowel Lesions: Challenges in Endoscopy-Image-Based Learning Systems », *Computer*, vol. 51, n° 5, p. 68-76, mai 2018, doi: 10.1109/MC.2018.2381116.
- [136] D. Salomon, *Data Compression: The Complete Reference*. Springer Science & Business Media, 2004.
- [137] P. C. Cosman, R. M. Gray, et R. A. Olshen, « Evaluating quality of compressed medical images: SNR, subjective rating, and diagnostic accuracy », *Proceedings of the IEEE*, vol. 82, n° 6, p. 919-932, juin 1994, doi: 10.1109/5.286196.
- [138] J. Urban, « How Chroma Subsampling Works », *Component*, 14 septembre 2017. <https://blog.biamp.com/how-chroma-subsampling-works/>.
- [139] A. K. Sekuboyina, S. T. Devarakonda, et C. S. Seelamantula, « A convolutional neural network approach for abnormality detection in Wireless Capsule Endoscopy », in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, avr. 2017, p. 1057-1060. doi: 10.1109/ISBI.2017.7950698.
- [140] X. Jia et M. Q.-H. Meng, « A deep convolutional neural network for bleeding detection in Wireless Capsule Endoscopy images », in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, août 2016, p. 639-642. doi: 10.1109/EMBC.2016.7590783.
- [141] « Artificial Intelligence in Medicine - 1st Edition ». <https://www.elsevier.com/books/artificial-intelligence-in-medicine/xing/978-0-12-821259-2>.
- [142] « Jetson Nano: Deep Learning Inference Benchmarks », *NVIDIA Developer*, 25 avril 2019. <https://developer.nvidia.com/embedded/jetson-nano-dl-inference-benchmarks>.
- [143] K. Sayood, *Introduction to Data Compression*. Morgan Kaufmann, 2017.
- [144] B. Sushma et B. Fatimah, « Wyner-Ziv Coding of Chroma in Wireless Capsule Endoscopy Image Compression using Deep Side Information Generation », in *2020 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*, août 2020, p. 58-62. doi: 10.1109/WiSPNET48689.2020.9198564.