

**« Implication des futurs techniciens en radiologie médicale dans
l'élaboration d'un modèle d'intelligence artificielle dédié à la
reconnaissance d'images radiologiques ?
Quel intérêt pour la formation ? »**

Travail de Bachelor

LOCARNINI DAMIEN

N° matricule : 17-594-458

PERRÉARD DAVID

N° matricule : 17-593-476

SIERRA PAULINO BRUNO

N° matricule : 17-592-999

Directeur : SCHMID Jérôme - Professeur HEdS filière TRM

Membres du jury : CHÊNES Christophe - Collaborateur scientifique

LOKAJ Belinda - Assistante HES

Genève, Août 2020

Avertissement

La rédaction et les conclusions de ce travail n'engagent que la responsabilité de ses auteur-e-s et en aucun cas celle de la Haute École de Santé Genève, du Jury ou du Directeur ou Directrice de Travail de Bachelor.

Nous attestons avoir réalisé seuls le présent travail, sans avoir utilisé d'autres sources que celles indiquées dans la liste des références bibliographiques.

Septembre 2020

Locarnini Damien / Perréard David / Sierra Paulino Bruno

Remerciements

Tout d'abord, nous aimerions remercier toutes les personnes ayant contribué à l'élaboration ainsi qu'au bon déroulement de ce travail.

En premier lieu, nous remercions notre directeur de travail, Monsieur Jérôme SCHMID, Professeur à la Haute École de Santé de Genève, qui nous a guidés et offert tous les outils pour mener à bien ce mémoire.

Nous souhaitons également remercier, Monsieur Christophe CHÊNES, pour son implication dans la réalisation du travail pratique ainsi que pour l'aide fournie à notre directeur de travail tout au long de ce projet.

Ensuite, un grand merci aux élèves de 3^{ème} année de la filière Technique en radiologie médicale pour leur participation active dans le cadre du travail pratique.

Pour finir, nous adressons nos meilleures pensées à nos proches, familles et amis qui nous ont soutenus tout au long de notre parcours et particulièrement pour ce travail.

Un grand merci à celles et ceux qui nous ont apporté leurs conseils et aidé dans la correction de ce travail.

Résumé

Objectif : L'imagerie médicale est un domaine en constante évolution et l'intelligence artificielle en fait partie intégrante. Grâce à l'apprentissage profond, cette étude a pour but de démontrer la capacité d'un système de classification à catégoriser des images radiologiques du membre supérieur en fonction des régions anatomiques et incidences d'épaules. En parallèle, nous avons participé à l'élaboration d'un travail pratique (TP), en vue d'introduire certains concepts de l'apprentissage supervisé aux étudiants de 3^{ème} année de notre filière par le biais d'une tâche de labellisation (tagging) indirecte d'images radiologiques. Cela permettra de récolter des données sur les performances des étudiants afin de les comparer à celle des réseaux entraînés à réaliser cette même tâche. Notamment, ce travail s'intéresse à la pertinence d'un tel format d'enseignement dans la formation et l'intérêt d'acquérir ces connaissances pour les futurs professionnel(le)s en radiologie.

Méthodologie : Dans le but de créer un dataset afin d'entraîner notre système d'IA, nous avons tagué 336 radiographies du membre supérieur en différentes régions anatomiques. Dans un second temps, 8379 radiographies d'épaule ont été taguées selon les cinq incidences de l'épaule principales grâce à un outil de labellisation ergonomique (VIA). Nous avons utilisé 2 types d'architectures réseau et appliqué différentes stratégies d'entraînement afin de déterminer le meilleur modèle pour réaliser cette tâche de reconnaissance. De plus, nous avons réalisé un travail pratique impliquant les étudiants de 3^{ème} années de la filière en vue d'introduire les concepts de base de l'IA et réaliser une tâche de tagging indirect afin de récolter des données pour les comparer avec nos différents modèles pré-entraînés.

Résultats : En fonction des différentes stratégies d'IA testées, il apparaît que les réseaux Nasnet_a_large et Inception V4 démontrent de bonnes performances (Accuracy) dans la reconnaissance des régions anatomiques du membre supérieur (0.965 et 0.966). Concernant la classification par incidences d'épaule, une baisse des performances est observée pour l'ensemble des modèles (12 à 15 %), mais les stratégies appliquées au réseau Nasnet_a_large semblent impacter positivement ces performances globales (0.842) contrairement au réseau Inception V4. Dans le cadre du TP réalisé, les performances des étudiants demeurent supérieures à l'IA (+17,18%). Malgré une moins bonne fiabilité, la rapidité de l'IA est non négligeable au vu de la quantité de données qu'elle arrive à traiter dans le même intervalle de temps que les étudiants (< 0,04 seconde / image).

Conclusion : Nous avons participé à l'élaboration d'un système d'IA capable de reconnaître des images radiologiques du membre supérieur en fonction des régions anatomiques ou des incidences d'épaule concernées. En fonction des différentes stratégies appliquées, nous avons analysé les résultats et défini le(s) modèle(s) se conformant le mieux à nos données. La réalisation du travail pratique a permis d'introduire les bases du fonctionnement et de l'entraînement d'un réseau de neurones au travers du tagging indirect. Cela a permis aux étudiants de prendre conscience des perspectives offertes par ces technologies et d'entrevoir leur impact sur l'évolution du métier de TRM.

Mots clés : Radiologie, Imagerie médicale, Technicien en radiologie médicale, Pédagogie, Intelligence artificielle, Réseaux de neurones, Machine learning, Deep learning, Réseaux de neurones convolutifs, Reconnaissance d'image, Labellisation, Big data.

Table des matières

| | |
|---|-----------|
| Liste des abréviations | 7 |
| Liste des annexes | 9 |
| 1. Introduction | 10 |
| 2. Cadre théorique | 11 |
| 2.1. Historique..... | 11 |
| 2.1.1. Histoire de l'informatique..... | 11 |
| 2.1.2. Histoire de l'IA | 12 |
| 2.1.3. Big Data | 14 |
| 2.2. Intelligence artificielle et apprentissage machine..... | 15 |
| 2.2.1. L'intelligence artificielle..... | 15 |
| 2.2.2. Machine Learning | 17 |
| 2.2.3. Méthodes d'apprentissage..... | 18 |
| 2.2.4. Réseau de neurones..... | 20 |
| 2.2.5. Apprentissage profond | 22 |
| 2.2.6. Réseau de neurones convolutif | 24 |
| 2.3. Entraînement et évaluation d'une approche d'IA | 25 |
| 2.3.1. Entraînement..... | 25 |
| 2.3.2. Validation et évaluation des performances | 30 |
| 2.4. Domaines d'applications de l'IA..... | 33 |
| 2.4.1. Médecine / Santé..... | 34 |
| 2.4.2. Radiologie..... | 34 |
| 2.5. Pédagogie..... | 37 |
| 2.5.1. Générale | 37 |
| 2.5.2. Radiologie..... | 37 |
| 3. Méthodologie..... | 39 |
| 3.1. Problématique..... | 39 |
| 3.2. Objectifs | 39 |
| 3.3. Hypothèses | 40 |
| 3.4. Étude 1..... | 41 |
| 3.4.1. Base de données (images)..... | 41 |
| 3.4.2. Tagging | 44 |

| | | |
|-------------|--|------------|
| 3.4.3. | Entraînement | 46 |
| 3.4.4. | Architectures d'IA sélectionnées | 48 |
| 3.4.5. | Validation / Test (évaluation du modèle déjà entraîné) | 51 |
| 3.5. | Étude 2..... | 52 |
| 3.5.1. | Élaboration travail pratique (Bsc 3 ^{ème} année) | 52 |
| 3.5.2. | Déroulement du travail pratique | 54 |
| 3.5.3. | Fin du travail pratique | 57 |
| 4. | Résultats | 58 |
| 4.1. | Étude 1 : Performances des différentes stratégies d'IA..... | 58 |
| 4.1.1. | Qualité d'images | 58 |
| 4.1.2. | Régions anatomiques | 61 |
| 4.1.3. | Incidences | 63 |
| 4.2. | Étude 2 : Travail pratique - Performances étudiants | 65 |
| 4.2.1. | Phase 0 - Félines | 65 |
| 4.2.2. | Phase 1- Régions anatomiques..... | 66 |
| 4.2.3. | Phase 2 - Incidences de l'épaule (sans support)..... | 67 |
| 4.2.4. | Phase 3 - Incidences de l'épaule (avec support) | 70 |
| 4.2.5. | Travail pratique - Performance IA | 72 |
| 5. | Analyse & Discussion | 73 |
| 5.1. | Étude 1 - Analyse des performances de l'IA..... | 73 |
| 5.1.1. | Influence de l'architecture réseau | 73 |
| 5.1.2. | Influence de la stratégie | 74 |
| 5.1.3. | Tâches de classification | 80 |
| 5.2. | Étude 2 - Travail pratique (TP) | 85 |
| 5.2.1. | Aide avec fiche de rappel (Phase 2 et 3)..... | 85 |
| 5.2.2. | Pertinence pour la formation..... | 89 |
| 5.3. | Synthèse..... | 91 |
| 5.4. | Potentiel, limites et perspectives | 92 |
| 6. | Conclusion | 94 |
| 7. | Bibliographie | 95 |
| 8. | Annexes..... | 101 |

Liste des abréviations

| Sigle | Acronyme anglais | Équivalent en français |
|---------------|---|--|
| ACC | Accuracy | Exactitude |
| AI | Artificiel Intelligence | Intelligence artificielle (IA) |
| AUC | Air Under the Curve | Aire sous la courbe |
| CAD | Computer Aided Diagnosis | Aide au diagnostic assisté |
| CAM | Class Activation Map | Carte d'activation des classes |
| CNN | Convolutional Neural Network | Réseau de neurones convolutionnels |
| DICOM | Digital imaging and communications in medicine (E.U.) | |
| DL | Deep Learning | Apprentissage profond |
| FDA | Food and Drug Administration | Agence américaine des produits alimentaires et médicamenteux |
| GPU | Graphics Processing Unit | Processeur graphique |
| JPEG | Joint Photographic Experts Group | Groupe conjoint d'experts photographiques |
| ML | Machine Learning | Apprentissage machine |
| MURA | MUsculoskeletal RAdiographs | Base de données musculo-squelettique |
| PET-CT | Position Emission Tomography | Tomographie par émission de positrons |
| PNG | Portable Network Graphics | Format de données pour les images numériques |
| ROC | Receiver Operating Characteristic (E.U.) | Courbe sensibilité / spécificité |
| VGG | Visual Geometry Group | |
| VIA | VGG image annotator | |

| Sigle | Acronyme français |
|--------------|--|
| DEFR | Département fédéral de l'Économie de la Formation et de la Recherche |
| ECG | Électrocardiogramme |
| FN | Faux Négatif |
| FP | Faux Positif |
| HE | Humérus en rotation Externe |
| HEDS | Haute École de Santé |
| HI | Humérus en rotation Interne |
| HN | Humérus en rotation Neutre |
| HUG | Hôpitaux Universitaire de Genève |
| IRM | Imagerie par Résonance Magnétique |
| RNA | Réseaux de Neurones Artificiels |
| RNC | Réseau de Neurone Convolutif |
| TFP | Taux de Faux Positifs |
| TP | Travail Pratique |
| TRM | Technicien en Radiologie Médicale |
| TVP | Taux de Vrais Positifs |
| VN | Vrai Négatif |
| VP | Vrai Positif |
| VPN | Virtual Private Network |
| VPP | Valeur Prédicative Positive |

Liste des annexes

Intelligence artificielle

I.Méthode d'évaluation des performances (Courbes ROC, macro, micro average)

Étude 1 – Entraînement IA

II.Étude 1 : Résumé des stratégies et performances par modèle

III.Résumé des erreurs par modèle et par classe

IV.Matrice de confusion des classes par modèle

Étude 2 – Travail pratique

V.Résumé des stratégies et performances par modèle

VI.Résumé d'erreurs par modèle et par classe

VII.TP IA – Protocole

VIII.TP IA – Consignes

IX.TP IA – Équipes

X.TP IA - Résultats des étudiants (Excel)

XI.Phase 0 – Images de félins

XII.Phase 3 - Fiche de rappel (rotations de l'humérus)

1. Introduction

Grâce aux importants progrès réalisés dans le domaine de l'intelligence artificielle (IA) cette dernière décennie, de nombreuses applications ont vu le jour. Le domaine de l'imagerie médicale bénéficie déjà de systèmes performants basés sur le Machine learning (ML), notamment dans la détection d'anomalies. Actuellement, l'IA semble être suffisamment fiable pour apporter une aide auprès des médecins (cliniciens et radiologues) dans le diagnostic et l'orientation de la prise en charge thérapeutique. Parmi les professionnels concernés, les Techniciens en Radiologie Médicale (TRM) seront témoins de l'implémentation progressive de ces technologies au sein des outils informatiques et électroniques utilisés dans l'exercice de leur métier. Cette évolution constitue un défi pour la formation des futurs professionnels qui se verront contraints d'acquérir de nouvelles connaissances pour ne pas se voir dépassés. Ainsi, l'IA devra trouver sa place au sein de la formation pour satisfaire les futures exigences en milieu clinique et dans la recherche.

En ce sens, ce travail étudie la potentielle implication des futurs TRM dans l'entraînement et la validation d'un modèle d'IA ainsi que son rôle bénéfique pour la formation. Une première étude visera à entraîner et évaluer la capacité d'un modèle de réseau de neurones profond à reconnaître des radiographies des membres supérieurs et les classer en différentes catégories (régions anatomiques et incidences de l'épaule). La seconde étude s'articule autour d'un atelier pratique réalisé auprès des étudiants de 3^{ème} année et basé sur le thème de l'IA.

Pour cela, une étape préliminaire de récolte et de traitement de données (images radiologiques) devra permettre d'entraîner un ou plusieurs modèles de classification et d'identifier les facteurs limitant l'apprentissage et les performances. En fonction de quoi, nous testerons différentes stratégies d'optimisation et identifierons les points forts et faibles de l'IA.

En profitant des connaissances radio-anatomiques acquises au cours de ces 3 années de formation, le recours aux étudiants nous permettra de mettre en perspective la fiabilité de l'être humain et de la machine dans l'accomplissement d'une tâche de labellisation d'images radiologiques. Ainsi, nous élaborerons une piste de réflexion pour l'enseignement de cette discipline auprès de la filière concernée et évaluerons l'utilité et la pertinence d'une telle application à des fins pédagogiques ou cliniques.

2. Cadre théorique

2.1. Historique

2.1.1. Histoire de l'informatique

L'informatique moderne est l'aboutissement d'un long parcours à travers l'histoire. En effet, avant de voir apparaître les machines que nous connaissons aujourd'hui, bon nombre d'étapes préliminaires ont dû être franchies.

Nous entendons par machine un « appareil ou ensemble d'appareils capable d'effectuer un travail ou de remplir une certaine fonction, sous la conduite d'un opérateur ou de manière autonome » (Larousse, 2019). Avant l'avènement de l'ère de l'électronique, leur fonctionnement était purement mécanique.

Nous retrouvons, par exemple, le « métier à tisser de Jacquard », inventé au début du XIX^e siècle par Joseph Marie Jacquard, et qui n'est autre que la première machine 100% mécanique programmable de l'histoire. Cet exploit a été rendu possible par l'utilisation de cartes (ou ruban) perforées permettant de sélectionner certaines aiguilles et pas d'autres en vue de créer des motifs complexes (Dowek, 2011). De la même façon qu'une mélodie jouée par un orgue de barbarie.

Bien qu'encombrante et limitée, cette technologie a permis la naissance de nombreuses machines pour des applications diverses et variées (principalement le calcul). Malgré tout, leur architecture (composants) et leur mode de fonctionnement ne permettaient de réaliser que des opérations simples.

Il faudra attendre le XX^{ème} siècle et la parution de l'article « On Computable Numbers, with an Application to the Entscheidungsproblem » (1936) rédigé par le célèbre mathématicien, Alan Turing, pour conceptualiser l'idée d'une machine polyvalente ou « machine universelle » dont l'architecture permettrait d'exécuter n'importe quel programme et d'accomplir des calculs complexes de manière automatique en se basant sur des règles mathématiques. Considéré comme le « père de l'informatique », Alan Turing a établi les fondements de la logique informatique et décrit le fonctionnement théorique des ordinateurs avant même leur apparition au milieu du XX^{ème} siècle.

Favorisée par la Deuxième Guerre Mondiale, l'informatique a progressé de manière fulgurante, notamment grâce à la transition électronique, et a permis l'émergence des premières générations d'ordinateurs. Par la suite, les architectures ont évolué jusqu'à aboutir (1945), grâce au mathématicien John Von Neuman, au modèle de fonctionnement que nous connaissons aujourd'hui (unité de calcul (CPU), mémoire, entrée-sortie, etc.).

2.1.2. Histoire de l'IA

Plusieurs courants de pensée ont encouragé le questionnement et la recherche autour de l'IA. À la croisée des sciences cognitives, mathématiques et informatiques, nombreux sont les chercheurs qui ont permis une avancée dans la conceptualisation de l'intelligence, son fonctionnement et sa modélisation. Parmi eux, Warren MacCulloch (neurophysiologiste) et Walter Pitts (logicien) ont conçu, en 1943, le premier modèle formel du neurone biologique et introduit le concept de « connexionnisme » qui perçoit l'intelligence comme étant le reflet d'un agencement complexe d'unités de calcul élémentaire (neurones) interconnectées (Memmi, 1990). En théorie, cette approche permettrait, moyennant une puissance de calcul hors-norme, de reproduire les processus neurobiologiques à l'échelle d'une machine.

Entrant en compétition avec cette dernière, le courant de pensée « cognitiviste » alloue à l'information une valeur sémantique plus profonde et abstraite, qualifiée de « symbolique » (Vittet, 2009). La maîtrise des symboles pourrait ainsi permettre aux machines de manipuler des représentations de haut niveau de complexité et d'abstraction, à l'instar de l'esprit humain.

Entre les années 1940 et 1950, l'importante évolution des technologies informatisées a considérablement changé notre vision des machines et de leur potentiel en termes d'applications. À un point tel qu'en 1950, Alan Turing publie un article « Computing Machinery and Intelligence » dans lequel il soumet la question suivante : « Les machines peuvent-elles penser ? » (Turing, 1950).

Cette idée va résonner au sein de la communauté scientifique et inspirer l'émergence de cette discipline dont le nom et la définition ne seront évoqués qu'en 1956 lors d'une conférence au Dartmouth College (Hanover, USA). À partir de quoi, plusieurs projets et problématiques vont émerger. Notamment, sur la question de l'apprentissage automatique.

Un an plus tard (1957), Frank Rosenblatt, inspiré des travaux de McCulloch & Pitts, crée le premier réseau de neurones artificiels (Perceptron) informatisé : Le Mark I. Dotée de cellules photo-électriques la rendant capable de simuler la vision humaine, cette machine était en mesure de reconnaître et classer des formes simples (ex. traits). Elle représente l'un des premiers pas vers l'Apprentissage automatique (Machine Learning ou Apprentissage machine).

Malgré l'effervescence, les promesses et les fantasmes qui gravitent autour, l'IA va subir, entre 1970 et 1980, une période hivernale. En effet, les limitations imposées par la faible puissance de calcul des ordinateurs de l'époque ont provoqué une grande déception. De longs débats scientifiques et philosophiques sur l'intelligence et ses limitations au sein d'une machine, vont conduire les chercheurs à redéfinir de nouveaux objectifs plus réalistes en vue de créer des systèmes industriellement exploitables.

Durant cette même période vont apparaître les premiers « Systèmes experts » qui se définissent comme étant des « outils capables de reproduire les mécanismes cognitifs d'un expert, dans un domaine donné » (Nabli, 2010). En effet, non basés sur une méthode d'apprentissage automatique, ces premiers systèmes « intelligents » sont néanmoins capables, par inférence (« si...alors... »), d'accomplir des tâches très spécifiques en suivant un raisonnement explicitement programmé à la main sur la base des connaissances d'un expert du domaine en question. Cependant, cette méthode est rapidement dépassée lorsqu'aucune expertise n'existe en la matière.

En parallèle, les réseaux de neurones regagnent l'intérêt des chercheurs. De nombreuses innovations techniques, mathématiques et algorithmiques voient le jour et permettent d'améliorer l'autonomie d'apprentissage des systèmes. Au début des années 1980, Terry Sejnowski et Geoff Hinton développent de nouvelles architectures de réseaux de neurones dotés de plusieurs couches et mettent en application l'algorithme de « Rétro-propagation de Gradient » qui permet aux réseaux d'auto-ajuster ses paramètres pour minimiser l'erreur finale, et donc, d'apprendre (Cardon & al., 2018).

Ce n'est qu'au milieu des années 90' que l'essor de l'informatique va donner un nouvel élan à l'IA. La mémoire ainsi que la puissance de calcul des ordinateurs semblent ouvrir le champ des possibilités d'application et d'industrialisation de ces systèmes.

Durant cette même période, Yann Le Cun, considéré aujourd'hui comme l'un des pères de l'Apprentissage profond (Deep Learning (DL) en anglais), développe le premier réseau de neurones convolutif dont l'agencement est directement inspiré du cortex visuel des mammifères. Particulièrement adaptée à la reconnaissance d'images, cette innovation a permis l'émergence au milieu des années 90' de l'une des premières applications mondiales de l'IA pour la lecture automatique des chèques bancaires. Malgré cela, les ordinateurs manquent encore de puissance pour garantir le succès de ces technologies (Le Cun, 2015).

La recherche connaît une nouvelle période hivernale jusqu'en 2012 où l'émergence et l'utilisation de nouveaux processeurs graphiques (GPU - Graphics Processing Unit) donne un regain d'intérêt pour les réseaux de neurones, notamment grâce à l'écrasante victoire d'un algorithme de reconnaissance d'images dans le cadre d'une compétition internationale (ImageNet).

Dans le même temps, la globalisation massive des technologies informatiques et l'émergence d'internet intensifient la création et les échanges de données à travers le monde. En conséquence, les centres de données (Data Center) se multiplient aux quatre coins du monde pour contenir cette vague numérique.

2.1.3. Big Data

Les données ont souvent constitué une limite à la conception et à l'apprentissage de systèmes d'IA performants. À l'heure actuelle, l'explosion des données numériques offre un environnement propice au développement d'algorithmes capables d'appréhender ces données à des fins prédictives (finances, marketing, épidémiologie, etc.).

Apparu à la fin du XXème siècle, le terme « Big data » (Mégadonnées ou Données massives) désigne un phénomène d'accroissement exponentiel des données rendant difficile leur traitement par des outils informatiques classiques (Bastin & Tubaro, 2018). Plus qu'une simple définition, il s'agit aujourd'hui d'une discipline à part entière qui vise à dompter et traiter ce flux conséquent de données, qu'elles soient structurées (tableau de données) ou non (images, texte, sons, etc.). Ce phénomène s'articule autour de 4 axes (4V), à savoir, le volume (quantité de données), la vélocité (quantité de données traitée par unité de temps (le flux de données)), la variété (hétérogénéité des formats de données telles que : texte, image, vidéo, sons, etc.) et la véracité (fiabilité et légitimité des données - qualité, exactitude).

Dans une optique de valorisation des données, l'utilisation de l'IA est amenée à devenir un outil analytique incontournable. Le secteur de la santé est certainement l'un des meilleurs candidats à l'implémentation de ces technologies, notamment sur la question du développement de l'e-santé et de la télémédecine (Béranger, 2016). Dans le domaine de la radiologie, nous estimons à 3,6 milliards le nombre d'exams d'imagerie réalisés annuellement (Frija, 2018). L'accroissement et le vieillissement de la population impacteront certainement ces chiffres à la hausse. Dans ce contexte croissant, les performances de l'IA pourraient fournir une aide précieuse.

2.2. Intelligence artificielle et apprentissage machine

2.2.1. L'intelligence artificielle

Étant la première espèce à tenter d'en donner la définition, l'intelligence a longtemps été considérée comme une propriété inhérente à l'être humain. D'après Gottfredson (1997), l'intelligence est une

« aptitude très générale qui, entre autres, implique la capacité de raisonner, de planifier, de résoudre un problème, de penser de manière abstraite, de comprendre des idées complexes, d'apprendre rapidement et de tirer des enseignements de l'expérience. »

Jusqu'alors, ces facultés dépendaient de facteurs biologiques, mais depuis peu, cette idée a été remise en question par l'évolution technologique.

Quand nous parlons d'IA, nombreuses sont les définitions existantes. En effet, la littérature peine à établir un consensus quant à la façon de définir l'IA, tant le sujet est vaste et complexe. D'après le Larousse (2019), il s'agit d'un « ensemble de théories et de techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine ». Yann Le Cun (2016) décrit ce domaine comme un « ensemble de techniques permettant à des machines d'accomplir des tâches et de résoudre des problèmes normalement réservés aux humains et à certains animaux ».

L'histoire nous apprend que l'IA est directement inspirée du fonctionnement cérébral et des milliards de neurones qui le composent. Néanmoins, l'intelligence n'est pas un phénomène inné, mais dépend en grande partie de l'interaction avec l'environnement et de la faculté

d'apprendre, comprendre et agir sur ce dernier. Depuis de nombreuses années, la question de l'apprentissage est au cœur même du développement de cette technologie.

Le but étant de parvenir à simuler, par le biais de machines, les facultés cognitives humaines telles que : l'apprentissage, le raisonnement, le calcul, la perception, la mémorisation, et même la découverte et la créativité (Bousquet, 2019).

La complexité des processus biochimiques à l'origine de ces phénomènes cognitifs constitue un obstacle majeur à la conception d'une forme d'intelligence universelle dotée de ces mêmes capacités. À l'heure actuelle, il est même impossible de statuer sur la possibilité de voir un jour apparaître une forme d'IA autonome et autodidacte égale ou supérieure à l'homme capable de créativité, de jugement, d'anticipation, d'adaptation ou de remise en question. Ce fantasme d'une IA « forte » continue toutefois de nourrir l'espoir de nombreux chercheurs.

Malgré les progrès technologiques fulgurants réalisés dans ce domaine, nous parlons encore d'IA « faible ». Ce terme, quelque peu réducteur, définit un système spécialisé dans la reproduction d'une tâche explicitement programmée ou dans l'apprentissage de cette même tâche sur la base de millions d'exemples. Elle n'a donc pas vocation à reproduire l'intelligence humaine dans toute son intégralité, mais simplement d'en imiter le raisonnement pragmatique lors de la réalisation d'une tâche élémentaire pour laquelle nous connaissons déjà la solution. Dans le cas d'un système informatique, cette réalisation est rendue possible par la traduction en langage de programmation sous la forme d'algorithmes.

Un algorithme se définit comme un « ensemble de règles opératoires dont l'application permet de résoudre un problème énoncé au moyen d'un nombre fini d'opérations » (Larousse, 2020). Nous pouvons concevoir l'algorithme comme une recette de cuisine. Dans le cas d'un ordinateur, les ingrédients représentent les données à traiter.

L'avantage des ordinateurs par rapport à l'homme réside dans leur capacité à trier, traiter et stocker d'importantes quantités d'informations. Permettant ainsi de réaliser des tâches de grande envergure en un temps record.

Cette puissance de calcul et de stockage informatique a rendu possible l'accès aux traitements d'une quantité grandissante de données en vue d'entraîner des modèles algorithmiques toujours plus performants dans la réalisation de tâches de plus en plus complexes, comme la reconnaissance d'images.

En fonction du domaine d'application et de la tâche à réaliser, il existe une multitude d'algorithmes et techniques envisageables. Comme le témoigne le tableau ci-dessous (Fig. 1), nous pouvons les considérer selon deux axes :

- Méthode d'apprentissage (supervisé ou non supervisé)
- Types de problèmes à traiter (régression ou classification)

| Algorithme | Mode d'apprentissage | Type de problème à traiter |
|------------------------------------|----------------------|------------------------------|
| Régression linéaire univariée | Supervisé | Régression |
| Régression linéaire multivariée | Supervisé | Régression |
| Régression polynomiale | Supervisé | Régression |
| Régression régularisée | Supervisé | Régression |
| Naive Bayes | Supervisé | Classification |
| Régression logistique | Supervisé | Classification |
| Clustering hiérarchique | Non supervisé | - |
| Clustering non hiérarchique | Non supervisé | - |
| Arbres de décision | Supervisé | Régression ou classification |
| Random forest | Supervisé | Régression ou classification |
| Gradient boosting | Supervisé | Régression ou classification |
| Support Vector Machine | Supervisé | Régression ou classification |
| Analyse en composantes principales | Non supervisé | - |

Figure 1. Tirée de : Eric Biernat & Michel Lutz, 2015. *Data Science : Fondamentaux et études de cas - Machine learning avec Python et R. [tableau 1-4]. Classification des algorithmes.*

Nous remarquons donc que les deux grands types de problèmes à résoudre sont la régression et la classification. La régression permet de prédire une valeur continue (ex. le prix d'une maison) à partir d'autres variables numériques (surface en m², nombre de pièces, etc.). Dans ce cas, la prédiction peut prendre une infinité de valeurs. La classification, quant à elle, permet de prédire une valeur discrète (ex. malin / bénin) à partir d'un ensemble de variables numériques (ex. taille d'une tumeur, temps de rehaussement artériel, etc.). Elle permet donc d'attribuer des données à une ou plusieurs classes (ou catégories).

2.2.2. Machine Learning

Depuis les débuts de la discipline, la question de l'apprentissage est au cœur de l'IA, de son fonctionnement et de son succès. Si bien que ces deux concepts sont désormais indissociables.

Le Machine learning (ML), Apprentissage automatique (ou Apprentissage machine) est une « discipline scientifique centrée sur le développement, l'analyse et l'implémentation de méthodes automatisables, qui offrent la possibilité à une machine d'évoluer grâce à un processus d'apprentissage, et ainsi accomplir des tâches qui auraient été difficiles, voire impossibles à réaliser avec des algorithmes plus classiques » (Bousquet, 2019).

Cette méthode offre donc la possibilité à une machine d'apprendre à partir de données, sans recourir à une méthode de programmation explicite et laborieuse en laissant opérer des modèles d'algorithmes évolutifs capables d'auto-ajustements.

En ce sens elle représente une solution efficace pour :

- Résoudre un problème pour lequel aucune solution n'existe.
- Résoudre un problème pour lequel beaucoup d'ajustements manuels sont nécessaires pour parvenir à la solution.
- Analyser et traiter d'importante quantité de données.
- Analyser des environnements fluctuants (Géron, 2017).

2.2.3. Méthodes d'apprentissage

Comme nous l'avons vu précédemment, il existe une multitude de méthodes d'apprentissage qui diffèrent en fonction du degré d'intervention de l'homme dans ledit processus d'entraînement.

nous distinguons :

- Supervisé
- Non-supervisé
- Par renforcement

Apprentissage supervisé

Cette méthode par induction consiste à fournir au système des données d'entraînement (ex. images) ainsi que les solutions correspondantes. Nous disons alors qu'elles sont « étiquetées » (couple entrée - sortie). Cette étape de labellisation (tagging ou labelling en anglais) est une condition essentielle.

Elle s'apparente à un enseignement scolaire classique dans lequel un professeur (opérateur) fournit un problème mathématique (entrées) ainsi que son résultat (sortie). L'élève (IA) doit alors trouver, lui-même, la bonne formule (fonction ou modèle) qui relie le problème au résultat (LeCun, 2016).

Apprentissage non-supervisé

Contrairement à l'apprentissage supervisé, cette méthode consiste à fournir au système des données d'entraînement dépourvues d'étiquettes (entrée uniquement). À partir de celles-ci, le modèle étudie les ressemblances, les différences et les liens observables entre elles.

Cette méthode d'apprentissage par observation (autodidacte) est très courante dans le monde animal, humain y compris (LeCun, 2016).

Lors du Salon de la Data (2017), Alexia Audevart nous explique que cette méthode est notamment utilisée pour des problèmes de segmentation (regroupement ou clustering) qui permet de séparer un groupe de données en paquets distincts et homogènes (cluster). Nous pouvons aussi l'utiliser pour découvrir des tendances (ex. moteur de recommandation de films et séries), extraire les caractéristiques (features) importantes (réduction de dimension) ou détecter des anomalies au sein d'un important volume de données.

Apprentissage par renforcement

Cette méthode est basée sur la valorisation d'une action par un système de récompense numérique permettant au système d'adapter son comportement face à une situation spécifique afin de maximiser cette récompense, essai après essai.

Nous pouvons comparer ce procédé au dressage animalier où l'animal est récompensé lorsqu'il parvient à réaliser l'action souhaitée.

Cette technique connaît un succès grandissant dans la robotique et permet d'atteindre des performances hors normes dans le domaine des jeux, en témoigne la victoire d'Alpha Go en 2015 (LeCun, 2016).

2.2.4. Réseau de neurones

Parmi les nombreuses techniques algorithmiques existantes pour faire de l'apprentissage machine, il en est une à qui nous devons la hausse des performances dans ce domaine depuis maintenant quelques années.

Les réseaux de neurones artificiels (RNA) sont directement inspirés par le fonctionnement du cerveau humain et, plus particulièrement, par la structure et l'agencement interconnectés de ses unités fonctionnelles élémentaires : les neurones.

Ce schéma biologique est à la base du Neurone formel (ou Neurone artificiel) (Fig.2). Concrètement, ce type de neurone opère de la même façon que son modèle biologique qui, à partir des signaux électrochimiques (entrées) reçus via ces connexions synaptiques avec les autres neurones, décidera ou non de s'activer (transmettre l'information). Informatiquement, cette action binaire se traduit par : 0 = pas d'activation / 1 = activation.

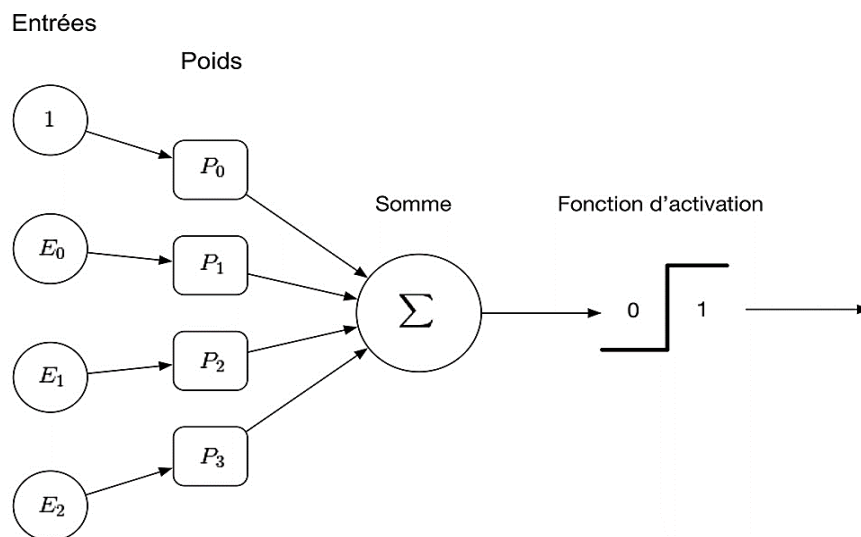


Figure 2. Tirée de : D. Cardon, J-P. Cointet & A.Mazières, 2018. *La revanche des neurones. L'invention des machines inductives et la controverse de l'intelligence artificielle.* (p. 173-220). Modèle formel d'un neurone artificiel à seuil binaire.

En se basant sur le modèle de neurone artificiel, Nous remarquons que l'activation dépend de plusieurs paramètres :

Poids

Le poids (coefficient ou poids synaptique) correspond à l'importance accordée au signal (entrée) reçu via une synapse, et donc, de sa contribution à l'activation finale du neurone. Il existe des synapses positives (activatrices) et négatives (inhibitrices). Ce paramètre évolue dans la phase d'apprentissage (Marc Parizeau, 2004).

Fonction d'activation

La fonction d'activation (ou de transfert) effectue une transformation (linéaire ou non) de la somme des entrées (multipliées par leur poids respectif) et définit ainsi le seuil d'activation (de 0 à 1). Elle est comparable au potentiel d'action (Simon, 2018) et permet de moduler les conditions d'activation du neurone. Une fonction non linéaire (sigmoïde) est par définition plus complexe qu'une fonction linéaire ce qui, en théorie, devrait permettre de générer des modèles plus sophistiqués. Il en existe une multitude, mais certaines semblent plus populaires (Fig. 3).

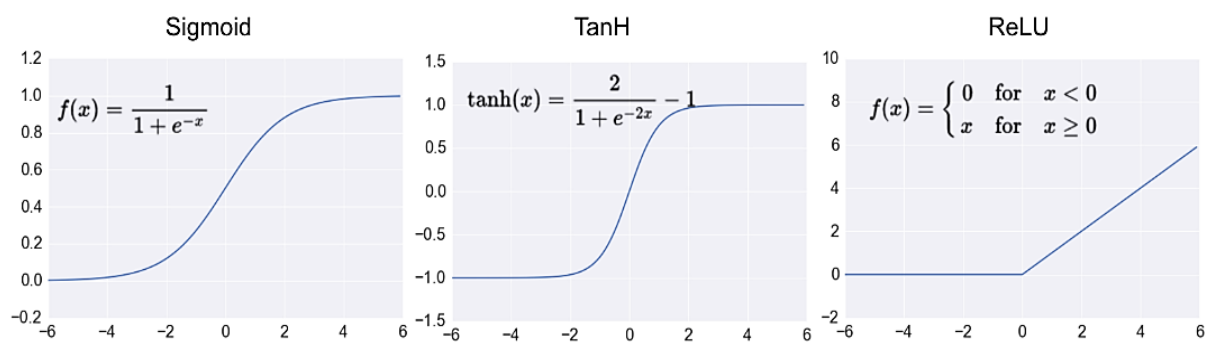


Figure 3. Tirée de : A. Moujahid, 2016. *A Practical Introduction to Deep Learning with Caffe and Python. Activation functions.*

Biais

Le biais est un paramètre externe au neurone qui favorise de façon arbitraire l'activation du neurone en modifiant son seuil d'activation. Il agit donc comme un poids d'entrée supplémentaire, mais qui ne dépend pas d'une synapse. Tout comme les poids, ce paramètre n'est pas fixe, mais évolue pendant la phase d'entraînement.

Schématiquement, un réseau de neurones se présente comme un empilement de couches de neurones (Fig. 4). Dans une situation basique, chaque neurone se ramifie pour établir des connexions avec l'ensemble des neurones de la couche suivante.

Chaque neurone de la première couche expédie les données d'entrée aux couches suivantes (cachées) qui effectuent des opérations mathématiques (sommations, multiplications, etc.), avant de transmettre (ou non) l'information aux neurones de la couche suivante et ainsi de suite jusqu'à la couche de sortie (résultat).

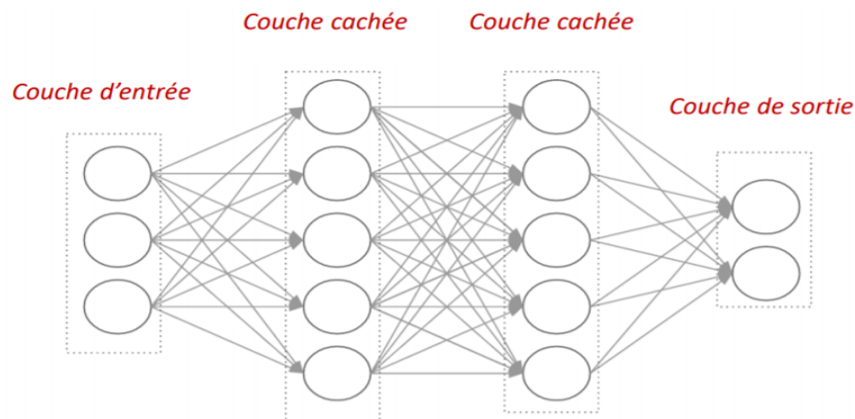


Figure 4. Tirée de : A. Moujahid, 2016. *A Practical Introduction to Deep Learning with Caffe and Python. Feedforward Neural Network with 2 hidden layers.*

Le nombre de neurones par couche et la façon dont ces derniers sont interconnectés représentent autant de facteurs pouvant varier d'un réseau à l'autre. Cette liberté de conception offre une infinité de possibilités d'architectures réseau et constitue un domaine de recherche à part entière. Le choix d'une architecture va dépendre de la tâche.

2.2.5. Apprentissage profond

Au fil des années, le nombre de couches n'a cessé d'augmenter, de même que les performances de l'IA. De cette augmentation est né le terme d'apprentissage profond (ou Deep Learning (DL)). Ce domaine fait partie intégrante du Machine learning (ML).

En théorie, plus le nombre de couches augmente, plus les performances d'apprentissage croissent de façon exponentielle, rendant possible la réalisation de tâches qui jusqu'alors semblaient impossibles ou limitées par les techniques classiques d'apprentissage automatique (ML).

Lors d'une Conférence (USI Event 2018), Yann LeCun, nous explique que l'avènement du DL a permis une hausse des performances dans de nombreux domaines tels que : reconnaissance de la parole, reconnaissance et catégorisation d'images, reconnaissance faciale, classification de texte, etc.

Le DL utilise des méthodes d'apprentissages identiques à celle du ML (supervisé, non supervisé et renforcement). L'avantage de cette technique réside dans l'autonomie d'apprentissage des caractéristiques liées aux données (Fig. 5). En effet, la méthode classique (ML) consiste à extraire manuellement des caractéristiques d'intérêt (features) au sein d'un jeu de données (dataset) avant d'entraîner le système. Cette tâche peut prendre un temps considérable et nécessite une bonne connaissance des données. Inversement, le DL est capable d'extraire lui-même les caractéristiques importantes pour parvenir au résultat final.

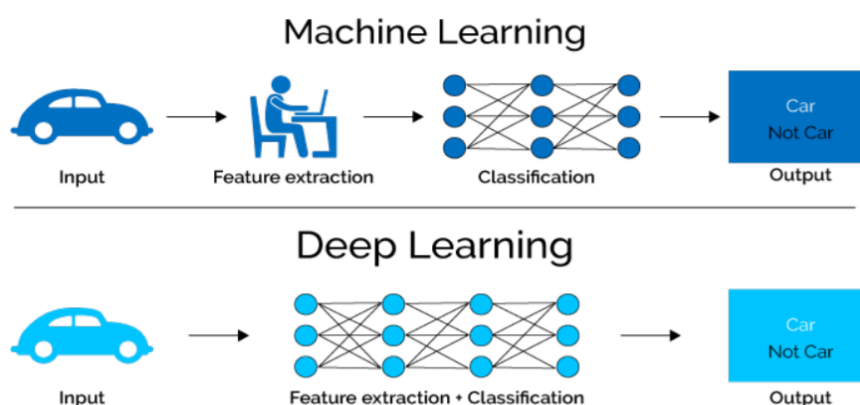


Figure 5. Tirée de : J. Kaur Gill, 2018. *Automatic Log Analysis using Deep Learning and AI. Différence entre approche ML et DL.*

Chaque couche (ou module) apprend à relever des tendances (pattern) au sein des données et parvient, par un cheminement déductif, à trouver la solution finale.

Pour parvenir à affiner un modèle de prédiction, ces réseaux complexes requièrent une importante quantité de données (images, texte, son, etc.). De plus, les ressources de calcul ainsi que le temps d'entraînement sont eux aussi plus conséquents. Une autre problématique concerne l'interprétation du fonctionnement de ces modèles qui, en raison de leur complexité, rendent difficile l'analyse de leur raisonnement.

Dans un souci éthique et juridique, ce phénomène de « boîte noire » constitue une limite à l'implémentation de ces systèmes dans certains domaines (ex. médical) où leur transparence est primordiale. Actuellement, les chercheurs travaillent sur le développement de méthodes d'inspection pour garantir leur transparence.

D'ores et déjà, il existe de nombreuses méthodes. En prenant l'exemple des réseaux de neurones convolutifs pour la reconnaissance d'images et/ou d'objet, il est possible de voir à travers les yeux du réseau grâce à de différentes méthodes telles que : Saliency map (ou Carte de saillance), Class Activation Map (CAM), Gradient Activation Map (Grad-CAM). Celles-ci permettent de générer, à partir d'une image (ex. une radiographie), une carte qui révélera, grâce à une échelle colorimétrique, les zones (pixels) ayant le plus contribué à l'activation du réseau et donc, à la prédiction finale (B. Zhou & al., 2019).

2.2.6. Réseau de neurones convolutif

Parmi les nombreuses architectures existantes, le Réseau de Neurones Convolutif ou Convolutional Neural Network (CNN) en anglais est un type de réseau multicouche directement inspiré d'une zone cérébrale : le cortex visuel. Ce type de réseau est donc particulièrement adapté au traitement et à la reconnaissance d'images en tout genre.

La convolution (Fig. 6) est une opération qui consiste à appliquer successivement des filtres de taille (matrice) définie. Cette matrice de convolution va, pour chaque zone de l'image, extraire les caractéristiques numériques propres à chaque pixel (ex. luminosité). En sommant les valeurs de ces caractéristiques et en les pondérant avec les valeurs de la matrice de convolution, ce procédé permet d'obtenir une unique valeur représentative de la zone de pixels. Pour chaque filtre appliqué, une image de dimension plus petite est générée puis sous-échantillonnée (pooling) de façon à réduire la quantité d'information à traiter et donc, optimiser le temps de calcul. L'opération est répétée autant de fois qu'il y'a de filtres et de couches de convolution.

Chaque couche est spécialisée dans l'extraction d'un nombre fini de caractéristiques (features) abstraites de l'image (traits, contours, formes, etc.) qui seront ensuite recombinaées en motifs (patterns) de plus en plus précis. Ces caractéristiques résumant l'entièreté de l'image initiale serviront ensuite de données d'entrée à un réseau de neurones plus classique qui, dans le cas d'une classification, déterminera l'appartenance de l'image à une classe en particulier (ex. animal, voiture, etc.).

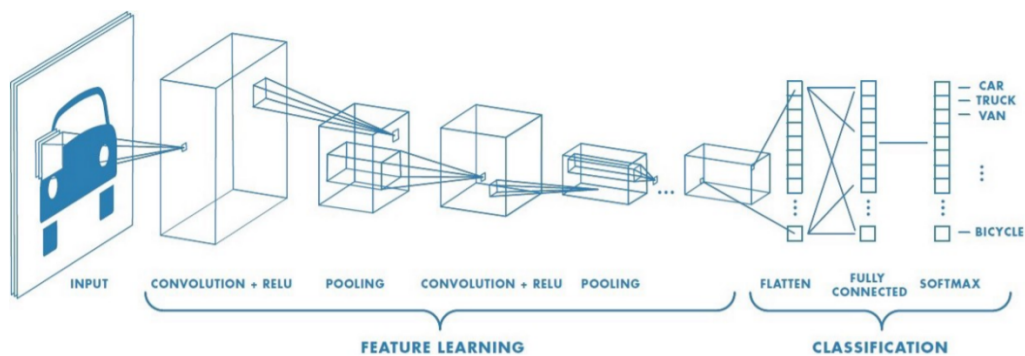


Figure 6. Tirée de : S. Saha, 2018. *A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way. Principe de convolution d'un réseau de neurones artificiels.*

2.3. Entraînement et évaluation d'une approche d'IA

Comme énoncé, tout système intelligent doit passer par une phase d'entraînement afin d'acquérir l'habileté de raisonnement nécessaire à la tâche pour laquelle il est destiné. Après quoi ses performances peuvent être évaluées sur un nouveau dataset.

2.3.1. Entraînement

Plusieurs étapes préliminaires sont nécessaires à l'entraînement d'un modèle d'apprentissage machine.

En premier lieu, le choix des données constitue une étape importante à l'élaboration du modèle. La quantité de données ainsi que leur qualité constituent un facteur important. En effet, les performances d'un modèle sont directement liées au volume du dataset d'entraînement. Cependant, il n'est pas toujours simple d'avoir accès à un volume conséquent de données. Leur qualité joue également un rôle déterminant. Pour cela, une phase de « nettoyage » est parfois nécessaire pour garantir que l'entraînement ne sera pas biaisé par des données superflues (Audevert, 2017).

Une étape de prétraitement des données (features engineering) est parfois nécessaire pour extraire les caractéristiques importantes sur lesquelles le modèle s'appuiera pour faire sa prédiction. Contrairement au réseau de neurones profond, cette étape s'effectue manuellement pour des méthodes ML classiques.

Dans un second temps, en fonction des données à traiter (ex. images) et de la tâche à réaliser (ex. classification), Nous choisirons un modèle algorithmique en particulier (ML ou DL). Parmi les réseaux de neurones, certains sont disponibles en libre accès. Le choix de réinitialiser ou

conserver les paramètres du réseau pour l'entraîner avec de nouvelles données représente deux modes d'entraînement : complet ou par transfert (Amazon Web Service, 2020).

Certains modèles sont plus adaptés à certaines méthodes d'apprentissage (supervisé, non supervisé, par renforcement). Aussi cette méthode va dépendre du dataset (étiquetées ou non) et de ce que nous voulons en tirer. Dans le cadre d'un apprentissage supervisé, les données sont labellisées. Le réseau sera alors chargé de découvrir la fonction reliant l'entrée (ex. image de chat) à sa sortie (label « chat »).

Ce processus d'apprentissage nécessite un certain nombre d'itérations durant lesquelles le réseau va optimiser ses paramètres (par ex. poids et biais) de façon à minimiser ses erreurs de prédiction. Étant donné l'importante quantité de paramètres ajustables, ce processus est rendu possible par un algorithme dit d'optimisation.

Responsable du phénomène d'apprentissage des réseaux de neurones profond, c'est l'algorithme de « Rétro-propagation de gradient » (ou Back propagation) qui effectue cette tâche (Fig. 7). Ce processus automatisé consiste à calculer la sensibilité de la sortie par rapport à l'ensemble des paramètres ajustables. Il permet donc de calculer l'implication de chaque paramètre dans l'erreur finale et d'indiquer la direction et l'amplitude avec laquelle ils doivent être modifiés pour que la sortie corresponde à celle qui est attendue (LeCun, 2018).

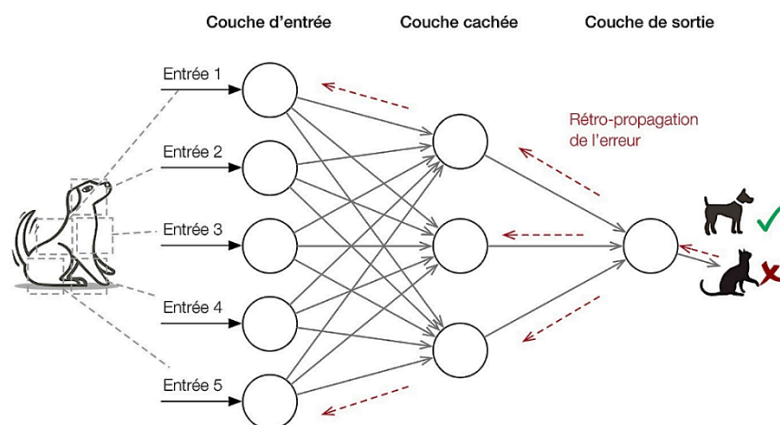
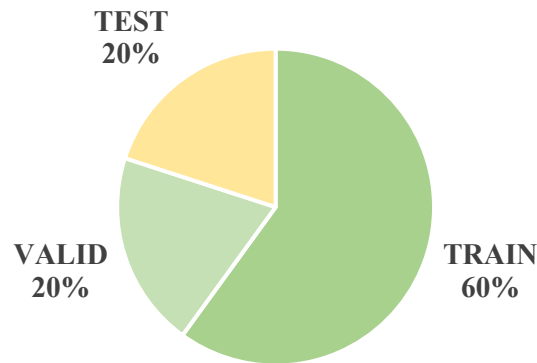


Figure 7. Tirée de : D. Cardon, J-P. Cointet & A.Mazières, 2018. La revanche des neurones. L'invention des machines inductives et la controverse de l'intelligence artificielle. (p. 173-220). Principe de rétro-propagation de gradient.

Il existe différentes méthodes d'entraînement, mais l'une des plus communes consiste à diviser le dataset en 3 parties distinctes : entraînement (TRAIN), validation (VALID) et évaluation (TEST). Cette méthode alloue une partie des données à l'entraînement (TRAIN et VALID) du modèle et l'autre à son évaluation (TEST ou Hold-out set) (graphique 1).



Graphique 1. Exemple courant de division du dataset.

Les données « TRAIN » servent à l'entraînement pur du modèle et l'ajustement de ses paramètres (poids, biais, etc.).

Les données « VALID » permettent, en cours d'entraînement, d'approximer, par un indice de performance, la capacité du modèle à effectuer des prédictions sur de nouvelles données et déterminent donc sa capacité de généralisation. Ces données n'influencent pas l'entraînement du modèle, mais permettent de rectifier le choix de certains hyper-paramètres régissant l'efficacité de l'entraînement lui-même (Leonel, 2019).

Nous entendons par « hyper-paramètres », des paramètres dont le choix influence la structure du réseau (ex. nombre de couches, nombre d'unités par couche) et la façon dont il est entraîné. Ils sont déterminés de façon fixe par l'opérateur lui-même en amont de l'entraînement et peuvent, si nécessaire, faire l'objet d'ajustement pour améliorer les performances du modèle (Microsoft, 2020). Il existe une multitude d'hyper-paramètres (tableau 1) et de choix de configurations possibles (Radhakrishnan, 2017).

| | |
|---|---|
| Taux d'apprentissage (Learning rate) | Définit la vitesse à laquelle le réseau ajuste ces paramètres et donc la vitesse d'apprentissage. |
| Taille du lot (Batch size) | Correspond à la quantité (lot) de données traitée avant ajustement des paramètres du réseau au cours de l'entraînement. |
| Nombre d'époques (Number of epoch) | Correspond au nombre de fois que l'ensemble des données d'entraînement transite par le réseau au cours de l'entraînement, soit le nombre d'itérations effectuées. |

Tableau 1. Exemples d'hyper-paramètres.

La validation du modèle en cours d'entraînement, par le biais des données « VALID », permet aussi de limiter le « sur-apprentissage » (« overfitting » en anglais). Ce phénomène caractérise le moment où les paramètres du modèle s'adaptent parfaitement aux données d'entraînement. En cours d'entraînement, ceci se caractérise par un écart significatif entre les performances réalisées sur les données « TRAIN » et « VALID ». Nous remarquons alors que le modèle est excellent sur les données d'entraînement, mais médiocre sur de nouvelles données (« VALID » et TEST). Nous disons alors qu'il n'est pas généralisable (Carpentier, 2019).

Plusieurs méthodes existent pour pallier ce problème. La plus simple consiste à récolter plus de données. Ceci n'est malheureusement pas toujours possible. Toutefois, à partir des données à disposition, il est possible d'en générer de nouvelles en appliquant de légères modifications (rotation, agrandissement, rétrécissement, translation, compression, filtres, colorisation, etc.). Ce procédé d'augmentation artificielle des données (Data augmentation) permet d'accroître considérablement la quantité de données d'entraînement (Ruizendaal, 2017). Cependant, le choix de la stratégie d'augmentation doit être adapté aux données qui seront traitées par le système (ex. changement de couleur pour des images radiologiques en noir et blanc).

Si malgré tout la quantité de données demeure faible, le fait d'amputer une partie des données d'entraînement (TRAIN) du modèle au profit de sa validation (VALID) risque de limiter les performances de ce dernier (P. Gupta, 2017), Nous parlons alors d'« underfitting ». De plus, les résultats d'entraînement de cette méthode sont biaisés par le choix de la division des données qui peuvent être hétérogènes (Allibhai. 2018).

Pour cela, il existe une méthode alternative de validation permettant de maximiser l'entraînement du modèle en l'entraînant succinctement sur l'ensemble des données précédemment dédié à l'entraînement (TRAIN et VALID), tout en conservant un dataset indépendant pour évaluer les performances finales du modèle (TEST).

Cette méthode dite de « Validation croisée » (ou Cross-validation en anglais) englobe plusieurs méthodes. La plus populaire est la méthode « K-folds Cross-validation » qui consiste à diviser le dataset en k-groupes (folds), généralement 5, et réaliser plusieurs entraînements au sein desquels chacun de ces paquets de données servira à l'entraînement et à la validation (tableau 2).

TRAIN **TEST**

| Entraînements | Dataset | | | | | Mesure |
|---------------|---------|--------|--------|--------|--------|--------------|
| 1 | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Performances |
| 2 | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Performances |
| 3 | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Performances |
| 4 | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Performances |
| 5 | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Performances |

Tableau 2. Fonctionnement de la validation croisée.

Cette stratégie n'influence pas les performances du modèle mais en donne une meilleure estimation ce qui permet, en fonction des résultats obtenus pour chaque entraînement (k-folds), de prédire sa capacité à traiter de nouvelles données (Degila, 2019).

Une autre méthode permettant de maximiser l'utilisation des données d'entraînement consiste à ne pas effectuer de validation (no validation). L'ensemble des données (full) sont utilisées pour entraîner le modèle et une partie est réservée à l'évaluation de ces performances. Bien qu'avantageuse, cette technique d'entraînement à l'aveugle rend difficile de prévoir ses performances sur de nouvelles données (TEST). En effet, les résultats obtenus par le biais de la validation influencent directement le choix de certains paramètres censés améliorer ses performances finales. Sans cette étape, il devient alors difficile de savoir comment le modèle se comportera sur de nouvelles données.

Outre la quantité, l'équilibre des classes (ou catégories) au sein d'un dataset peut aussi être un facteur limitant les performances d'entraînement. En effet, un dataset non équilibré (unbalanced en anglais) engendrera un modèle dont les performances seront médiocres pour la reconnaissance des classes les moins représentées. Pour pallier ce problème, nous pouvons avoir recours à une stratégie dite de « balancing ». Pour cela, plusieurs méthodes existent : Sur-échantillonnage (upsampling) aléatoire des classes minoritaires, sous-échantillonnage (undersampling) aléatoire des classes majoritaires, etc. (Kushwaha, 2019).

2.3.2. Validation et évaluation des performances

Une fois le modèle configuré sur la base des données d'entraînement (TRAIN et VALID), les données « TEST » permettent d'évaluer et de valider les performances du modèle sur un dataset n'ayant pas participé à son élaboration. Cette étape permet la validation du modèle dont les performances peuvent être mesurées par divers indicateurs tels que la précision ou par une analyse graphique.

La précision (formule) mesure pour une classe donnée, la proportion de prédictions correctes (vrais positifs ou VP) par rapport à l'ensemble des prédictions jugées correctes (vrais positifs + faux positifs ou FP), soit la « valeur prédictive positive (VPP) ». Elle détermine donc la capacité du modèle à prédire la bonne classe.

$$Précision = \frac{VP}{VP + FP}$$

Un autre indicateur de performance souvent confondu avec la précision est l'« Accuracy » (formule). Elle permet de mesurer la proportion de prédictions correctes (VP et VN) par rapport à l'ensemble des prédictions effectuées.

$$Accuracy = \frac{VP + VN}{VP + FP + VN + FN}$$

Une autre métrique couramment utilisée pour représenter les performances d'un classificateur est l'Aire sous la courbe (Area Under Curve (AUC) en anglais) (Baya-Laffite & al., 2018). Cette méthode d'analyse graphique consiste à mesurer, et ce pour chaque classe, la surface en dessous d'une courbe sensibilité / spécificité, aussi appelée : Courbe ROC (Receiver Operating Characteristic) (Fig. 8).

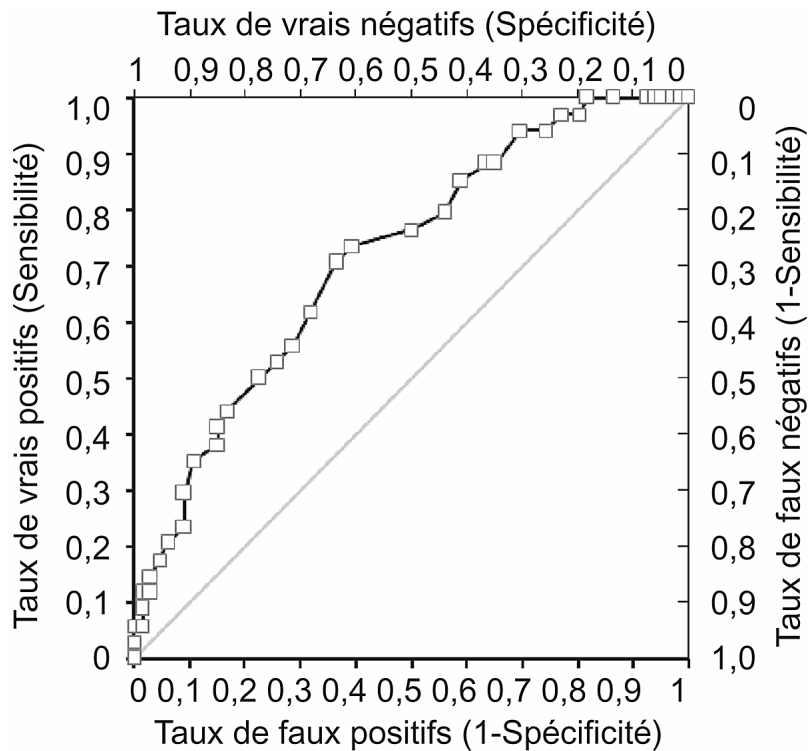


Figure 8. Tiré de : D. Bertrand, J. Fluss, C. Billard & J. C. Ziegler, 2010. *Efficacité, sensibilité, spécificité : comparaison de différents tests de lecture*. pp. 299-320. Courbe ROC (Receiver Operating System).

En guise de rappel, la sensibilité (ou rappel) (formule) définit la proportion (%) de vrais positifs (taux de vrais positifs) par rapport à l'ensemble des réponses jugées positives (VP + FN). Elle détermine donc la capacité du modèle à détecter les réponses positives (correspondant à la classe). Plus la sensibilité est élevée (proche de 1), plus la proportion de faux négatifs diminue (Bertrand & al., 2010).

$$\text{Sensibilité} = \frac{VP}{VP + FN}$$

La spécificité (formule), quant à elle, définit la proportion (%) de vrais négatifs par rapport à l'ensemble des réponses négatives attendues. Elle détermine donc la capacité du modèle à détecter les réponses négatives (ne correspondant pas à la classe). Plus la spécificité est élevée (proche de 1) plus la proportion de faux positifs diminue (Bertrand & al., 2010).

$$\text{Spécificité} = \frac{VN}{FP + VN}$$

Ainsi, la courbe ROC est une représentation graphique montrant la relation entre le taux de vrais positifs (TVP, sensibilité ou rappel) en fonction du taux de faux positifs (TFP = 1 – spécificité), c'est-à-dire le « taux de détection correcte d'éléments appartenant à la classe, en fonction du taux d'éléments détectés par erreur comme appartenant à la classe », et ce pour chaque valeur de seuil possible (Larlus & al., 2006).

Nous entendons par « seuil », la valeur (de 0 à 1) à partir de laquelle le modèle estime que sa prédiction est juste. En effet, chaque réponse en faveur d'une classe est accompagnée d'un « score » (de 0 à 1) qui établit la confiance que le modèle a envers sa prédiction (juste = 1 / faux = 0). Plus le seuil est bas, moins le modèle est discriminant et plus la quantité de mauvaises réponses augmente. Le but étant d'obtenir un seuil discriminant au mieux les bonnes réponses (VP + VN) des mauvaises (FP + FN) (Bertrand & al., 2010).

Idéalement, un modèle théoriquement parfait devrait pouvoir classifier les images sans aucune erreur et fournir une aire sous la courbe égale à 1.

L'analyse ROC est une méthode normalement utilisée pour traiter des problèmes à deux classes (ex. oui ou non). La généralisation à des problèmes multi-classes peut s'avérer problématique dans la mesure où chaque valeur de prédiction est influencée par l'ensemble des classes présentes, ce qui augmente le risque de confusion et d'erreurs possibles (Oufella, 2008). Or, cette méthode trace des courbes pour chaque classe indépendamment des autres. Sauf que le modèle de prédiction, quant à lui, tient compte de toutes les classes. Ainsi, lorsque nous souhaitons évaluer les performances globales d'un modèle, il est difficile de reconnaître la contribution de chaque classe au score final, d'autant plus si ces dernières sont mal équilibrées.

2.4. Domaines d'applications de l'IA

En 1956, des chercheurs (Herbert Simon, Newell Allen et Shaw Cliff) ont développé « Logic Theorist » qui est considéré aujourd'hui comme le premier programme « intelligent ». Conçu pour imiter les capacités humaines dans la résolution de problèmes mathématiques, ce programme était capable de démontrer 38 des 52 théorèmes existants à l'époque (Gugerty, 2006). Cette révolution dans le domaine de l'IA a prouvé qu'il était possible de créer des applications pertinentes et fonctionnelles. Les nombreuses simulations cognitives développées par ses 3 chercheurs ont eu un impact considérable dans le domaine du développement de la psychologie du traitement de l'information. Ce modèle est encore beaucoup utilisé de nos jours, pour simuler des tâches complexes étudiées en psychologie humaine (ou cognitive).

Par la suite, l'IA s'est imposée dans de nombreux secteurs et a vu naître une multitude d'applications. En effet, depuis les débuts de son implémentation à l'échelle industrielle, l'IA offre un potentiel de développement dans de nombreux domaines de la société actuelle, tels que : Administration publique, Agriculture, Commerce, Éducation, Recherche, Environnement, Finance, Industries, Juridique, Loisirs et médias, Sécurité, Télécommunications, Transport, Santé, etc. (Dubois & al., 2019).

Au cours de ces dernières années, ces observations ont permis à l'IA de s'imposer pour devenir un thème majeur dans la recherche technologique. La recherche s'affaire à reproduire les trois grandes catégories de tâches cognitives que sont : la perception de l'environnement, la compréhension d'une situation et la prise de décision.

De nos jours, même si les capacités de l'IA ne sont pas toujours de grande envergure, il existe des domaines où l'investissement et la recherche semblent se focaliser davantage, en témoignent la proportion d'études au sein des différents secteurs. Ce constat permet d'estimer le classement des secteurs qui seront le plus impacté par l'IA, notamment : Télécommunication et technologies, Industrie manufacturière et Santé.

2.4.1. Médecine / Santé

La littérature nous montre que l'IA aura un impact majeur dans le futur, tant au niveau social que professionnel. En effet, elle s'apprête à bouleverser de nombreux corps de métier, notamment dans la santé. Cette situation étant principalement due à l'importante accumulation de données rendant de plus en plus difficile leur exploitation par les professionnels qui se voient contraints d'utiliser des systèmes de gestions informatiques de plus en plus complexes.

La particularité de ce domaine est qu'il regorge de variables cliniques, biologiques et pathologiques, rendant le travail des médecins laborieux et constituant une source d'erreur potentielle. En effet, autour de chaque patient gravitent une multitude de données d'anamnèse (âge, sexe, historique médical, etc.) et cliniques (imagerie diagnostic, données génétiques, test électro-physiologique, symptomatologie, traitement, etc.) qui tendent à noyer les informations pertinentes (Jiang & al., 2017). Or, les réseaux de neurones sont capables d'exploiter toutes ces variables et d'établir une corrélation entre elles, contrairement à l'être humain qui n'en a pas toujours la capacité, ni même le temps. D'ores et déjà, elle apporte une aide dans la réalisation de nombreuses tâches, telles que : Recherche, Diagnostic, Thérapie, Dépistage, Planification de traitement, etc.

Cependant, cet engouement autour de l'IA a fait naître des enjeux éthiques et juridiques qui représentent désormais un obstacle à sa libre utilisation. Pour Odile Thiéry (2018), il existe une multitude de problèmes qui doivent être réglés avant d'envisager son utilisation par le plus grand nombre. Notamment : la confidentialité des données, le suivi continu de nos actions (objets connectés, etc.) et le rôle du médecin qui pourrait être dévalorisé. Sur ce dernier point, nombreux sont les professionnels qui craignent de voir leur profession disparaître.

2.4.2. Radiologie

Le domaine de la radiologie est fortement tributaire des avancées technologiques. Depuis maintenant plusieurs années, l'IA fait partie intégrante des projets de développement dans ce secteur. Actuellement, les applications en routine clinique se focalisent davantage sur la reconnaissance d'images et l'aide au diagnostic. Le but étant de réduire les erreurs médicales (faux positifs, faux négatifs) en fournissant un diagnostic plus fiable et rapide, permettant d'augmenter le flux de travail et le nombre de prestations fournies (Holzberger & Kim, 2018).

Depuis maintenant plusieurs années, certains instituts utilisent des systèmes d'aide pour détecter des tumeurs cancéreuses sur des mammographies, identifier des arythmies cardiaques sur des électrocardiogrammes (ECG) ou de détecter des cancers du poumon sur des scanners (Dubois & al., 2019).

Parmi eux, nous pouvons citer le CAD (Computer Aided Diagnosis). Ce système de diagnostic assisté par ordinateur a été développé dans le but d'améliorer la précision des examens, de permettre une meilleure cohérence dans l'interprétation des images, de mieux définir le pronostic et de soutenir le processus décisionnel (Koenigkam Santos & al., 2019). Il s'agit d'une forme précoce d'IA qui s'appuie sur les connaissances et le raisonnement d'experts du domaine. Bien que son fonctionnement n'ait pas toujours été basé sur des algorithmes d'apprentissage automatique (ML ou DL), les chercheurs y ont depuis intégré des algorithmes intelligents afin d'obtenir de meilleurs résultats (Deng & al., s.d.).

D'ores et déjà, ces applications permettent la détection, la segmentation automatique ainsi que la caractérisation (maligne ou bénigne) des lésions radiologiquement observables. De plus, elle permet d'effectuer le suivi temporel des lésions dans le cadre des examens de surveillance. À l'heure actuelle, il est principalement utilisé dans la détection des micro-calcifications, des masses ou des opacités dans le cadre du dépistage du cancer du sein en mammographie, dans la recherche de nodule pulmonaire au scanner ou pour diagnostiquer un polype en colonoscopie virtuelle (Grenier & Neuenschwander, 2006).

La recherche en IA dans le domaine de la radiologie clinique, et plus particulièrement l'apprentissage profond, semble s'intensifier depuis plusieurs années, en témoigne le nombre d'articles sur le sujet, et ce pour diverses modalités d'imagerie.

En radiologie conventionnelle, un système a été entraîné à détecter et localiser avec précision la présence d'un pneumothorax (Gooßen & al., 2019).

En médecine nucléaire, de nombreuses études démontrent l'efficacité des réseaux de neurones convolutifs à identifier et segmenter les lésions suspectes en tomographie par émission de positrons (PET-CT) (Santini & al., 2020).

En échographie, le recours à ces mêmes réseaux permet d'améliorer significativement la détection et la reconnaissance des nodules thyroïdiens (Song & al., 2019).

En imagerie par résonance magnétique (IRM), un réseau a été entraîné à reconstruire des séquences faiblement injectées au gadolinium sans perdre l'information sur le rehaussement, permettant de réduire jusqu'à 10 fois la dose injectée au patient (Gong & al., 2018).

En radio-oncologie, la segmentation automatique des structures et organes par une approche basée sur les réseaux profonds démontre une précision supérieure par rapport aux méthodes conventionnelles sollicitant un opérateur (Zhou, 2020).

De plus en plus, ces technologies s'emparent de nos appareils radiologiques (machines, stations de travail, etc.). Grâce à l'apprentissage profond, l'arrivée de nouveaux scanners dotés d'IA permet d'améliorer la qualité des images grâce à l'implémentation d'algorithmes de reconstruction (ex. Advanced intelligent Clear-IQ Engine (AiCE)) plus performants et rapides, permettant d'améliorer le rapport S/B, la netteté des images et la détectabilité à bas contraste, tout en permettant une réduction de la dose au patient (Canon, 2020). De plus, ces algorithmes trouvent également leur place dans l'ajustement du positionnement des patients permettant une meilleure reproductibilité dans la prise en charge, une amélioration de la qualité des images (rapport S/B) et l'optimisation de la dose au patient (General Electric, 2020). Ces nouveaux outils pourraient permettre un gain de temps considérable pour les TRM et leur donnant la possibilité de se focaliser sur d'autres volets de la prise en charge tels que : communication, confort, sécurité, etc.

Le Dr Johannes T. Heverhagen (2018) affirme que nous ne sommes qu'à l'aube de cette nouvelle ère et qu'il va falloir beaucoup d'études et de validation avant d'espérer leur utilisation sur le terrain. La question n'étant plus de savoir si ces technologies vont être utilisées, mais plutôt d'anticiper quand elles feront partie intégrante de la pratique professionnelle. En ce sens, de nouvelles normes et réglementations doivent voir le jour pour que ces outils, catégorisés de « produits de bien-être général », soient approuvés par les agences de régulation (ex. Food and Drug Administration (FDA)) responsables de leur commercialisation (Jiang & al., 2017).

2.5. Pédagogie

2.5.1. Générale

À l'heure actuelle, de nombreux chercheurs suggèrent l'utilisation de l'apprentissage et de la communication numérique pour les étudiants afin d'améliorer l'enseignement. Selon le Département Fédéral de l'Économie, de la Formation et de la Recherche (DEFR, 2019), l'IA pourrait apporter des opportunités afin d'améliorer l'enseignement et l'apprentissage. Ils mettent en avant que son utilisation dans le domaine de la pédagogie est très peu utilisée aujourd'hui, mais évoluera rapidement.

En effet, l'IA pourrait aider au développement de plateformes d'enseignement qui permettront une automatisation et une individualisation du processus d'enseignement, avec un suivi du comportement et des performances d'apprentissage. Ainsi, ce modèle d'interaction virtuel pourrait offrir un enseignement plus adapté aux besoins de l'apprenant.

2.5.2. Radiologie

Comme nous l'avons évoqué, le domaine de la radiologie est amené à utiliser davantage l'intelligence artificielle en routine clinique, dans les petites comme dans les grandes institutions. Dans le cadre de la formation de TRM, la littérature nous révèle une absence d'intérêt pour le développement d'outils pédagogiques en lien avec l'IA.

Cependant, la généralisation des outils professionnels utilisant l'IA forcera certainement les écoles à reconsidérer l'enseignement et les connaissances jugées nécessaires sur le terrain. Par ailleurs, il faudra faire la distinction entre les métiers produisant l'IA et ceux qui l'exploitent. Dans la seconde situation, il sera nécessaire d'acquérir une compréhension intuitive des algorithmes afin d'en comprendre les possibilités et les limites (DEFR, 2019).

Dans ce cas, l'expertise des professionnels de l'imagerie médicale (TRM et radiologue) pourrait, en collaboration avec des ingénieurs d'application, représenter un atout considérable pour la conception d'outils intelligents. Par leur compréhension des protocoles et critères de qualité, ces derniers seront susceptibles de fournir ou vérifier l'exactitude des informations transmises au système et celles émises par ce dernier en vue d'améliorer la qualité de ces prédictions. Dans une dimension plus pédagogique, cette tâche imposera de réfléchir davantage sur les aspects fondamentaux de notre pratique et permettre une avancée dans la recherche, la

production et l'utilisation de ces outils qui ne sauraient exister sans le transfert des connaissances expertes des utilisateurs qui en bénéficient.

La sollicitation des TRM et radiologues, professionnels ou étudiant dans la recherche et le développement de ces outils sont des points importants dans le succès de leur utilisation. De plus, nous pouvons imaginer que l'expertise naissante des étudiants du domaine pourrait apporter une importante plus-value dans le développement de ces outils.

En effet, les TRM bénéficient d'une bonne connaissance des processus d'acquisition, une capacité de jugement sur leur qualité et, à moindre mesure, d'une capacité de lecture et d'interprétation des anomalies susceptibles d'être détectées. En collaboration avec les autres corps de métier, cette plurivalence justifie l'intégration de notre profession au sein de la recherche clinique. Désormais, la thématique de l'IA prend une place importante dans la conscience conjointe des instituts radiologiques et des écoles spécialisées. Sous la forme de travaux académiques, des projets codirigés sont menés en ce sens.

Parmi eux, nous pouvons citer le travail d'Azal Al-Musibli et Onour Cagdas, « Intelligence artificielle en imagerie ostéo-articulaire : le rôle du TRM ? » (2019), mené à la Haute École de Santé (HEdS) à Genève en collaboration avec les Hôpitaux Universitaires de Genève (HUG). Outre la démonstration technologique de l'IA dans la détection et la segmentation des os du carpe, ce travail encourage la participation active des TRM et témoigne du rôle de notre profession dans la mise sur pied d'un tel projet.

En ce sens, notre sujet de recherche à toute son importance dans l'évaluation de la pertinence d'une éventuelle implication des étudiants dans cet apport d'expertise.

3. Méthodologie

3.1. Problématique

Notre question de recherche est :

« Quelle est l'implication des futurs techniciens en radiologie médicale dans l'entraînement d'un modèle d'IA dédié à la reconnaissance d'images radiologiques ? Quel intérêt pour la formation ? »

3.2. Objectifs

- Participer au développement d'un système de classification d'images radiologiques du membre supérieur en différentes instances (régions anatomiques / incidences de l'épaule).
- Comparer les performances des modèles de réseaux en testant différentes stratégies d'entraînement.
- Participer à l'élaboration d'un travail pratique (TP), en vue d'introduire les concepts de l'apprentissage supervisé par le biais du tagging indirect des images radiologiques du membre supérieur en différentes instances (régions anatomiques / incidences d'épaule).
- Récolter des données sur les performances des étudiants afin de les comparer à celle d'un ou plusieurs réseaux pré-entraînés.
- Évaluer l'utilité et la pertinence de ce format d'enseignement dans la formation et l'intérêt d'acquérir ces connaissances pour les futurs professionnel(le)s en radiologie.

3.3. Hypothèses

- La quantité d'images est un facteur déterminant dans l'entraînement et les performances du modèle. Cependant, nous pensons qu'une grande quantité d'images de mauvaise qualité ne constitue pas un réel avantage et que ces deux paramètres sont importants dans l'élaboration d'un système performant.
- Le choix d'une architecture réseau et d'une stratégie d'entraînement donné a un impact significatif sur les performances du modèle.
- Malgré leurs connaissances des images radiologiques, les étudiants TRM ont plus de difficulté à distinguer les incidences d'épaule que les régions anatomiques, car elles sont plus complexes et subtiles dans l'analyse. Nous pensons que ces difficultés se vérifieront également sur les performances du réseau. Néanmoins, nous pensons que l'IA demeure plus performante que les étudiants dans la classification des images par incidence.
- Nous pensons qu'après un bref rappel par le biais d'une « fiche de rappel » (phase 3), nous observerons une amélioration de la fiabilité du tagging des incidences d'épaule par les étudiants démontrant une meilleure fiabilité par rapport à l'IA.
- Nous pensons que le format du TP constitue une bonne amorce à l'acquisition des concepts de bases à l'entraînement d'un réseau de neurones et qu'il s'inscrit dans la transition à venir. Nous pensons que les étudiants prendront conscience de leur rôle dans le développement et l'amélioration de ces systèmes.

Notre travail est divisé en deux parties :

Étude 1 → Entraînement d'un système d'IA basé sur les réseaux de neurones profonds en vue de classer des radiographies du membre supérieur selon différentes catégories (régions anatomiques / incidences d'épaules).

Étude 2 → Élaboration d'un travail pratique (TP) pour les étudiants TRM de 3^{ème} année en vue d'introduire les concepts fondamentaux de l'IA et de l'entraînement d'un réseau de neurones profond par le biais du « tagging » indirect.

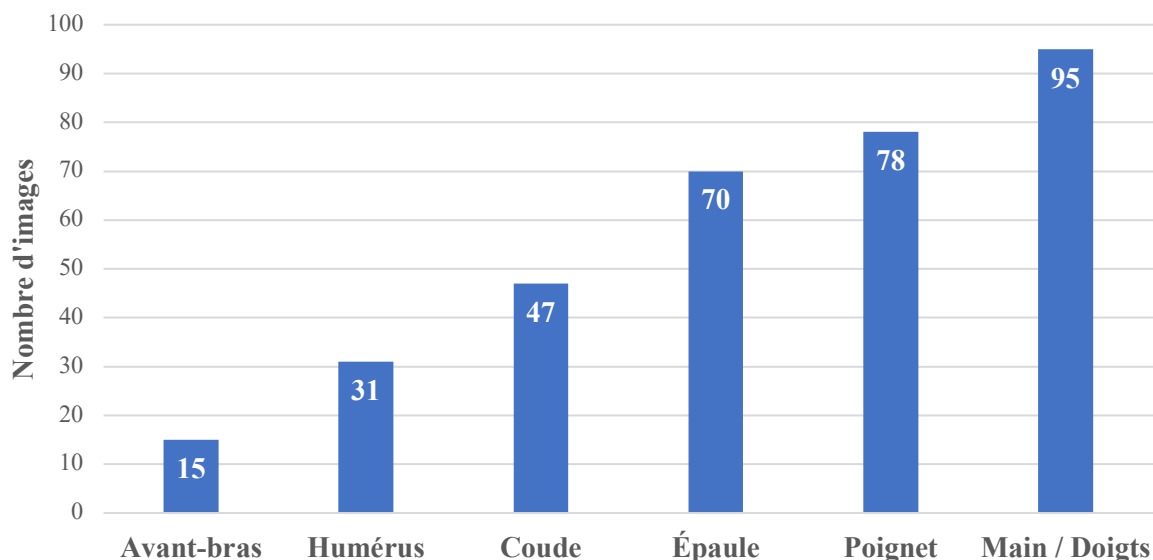
3.4. Étude 1

Le système d'IA utilisé dans le cadre de cette étude a été développé et déployé sur un serveur de calcul distant dédié à l'apprentissage machine. Pour accéder à ce serveur, nous avons obtenu des accès VPN (Virtual Private Network) et utilisé un logiciel open source (WinSCP) permettant la connexion entre un ordinateur local et un ordinateur distant.

3.4.1. Base de données (images)

Dataset HEDS

Ce dataset comprend 336 images radiographiques musculo-squelettiques récoltées auprès de différents partenaires genevois tels que les Hôpitaux Universitaires de Genève (HUG) (Graphique 2). Ces images ont été acquises en format DICOM (.dcm) et ont été converties en PNG (Portable Network Graphics) afin de répondre aux contraintes des réseaux de neurones utilisés (support exclusif d'une dynamique d'image de 8 bits) tout en conservant une qualité d'image interprétable (compression sans perte). Dans le cadre de cette étude, seules les radiographies du membre supérieur ont été retenues (Fig. 9).



Graphique 2. Base de données HEDS – Répartition des classes (ordre croissant).



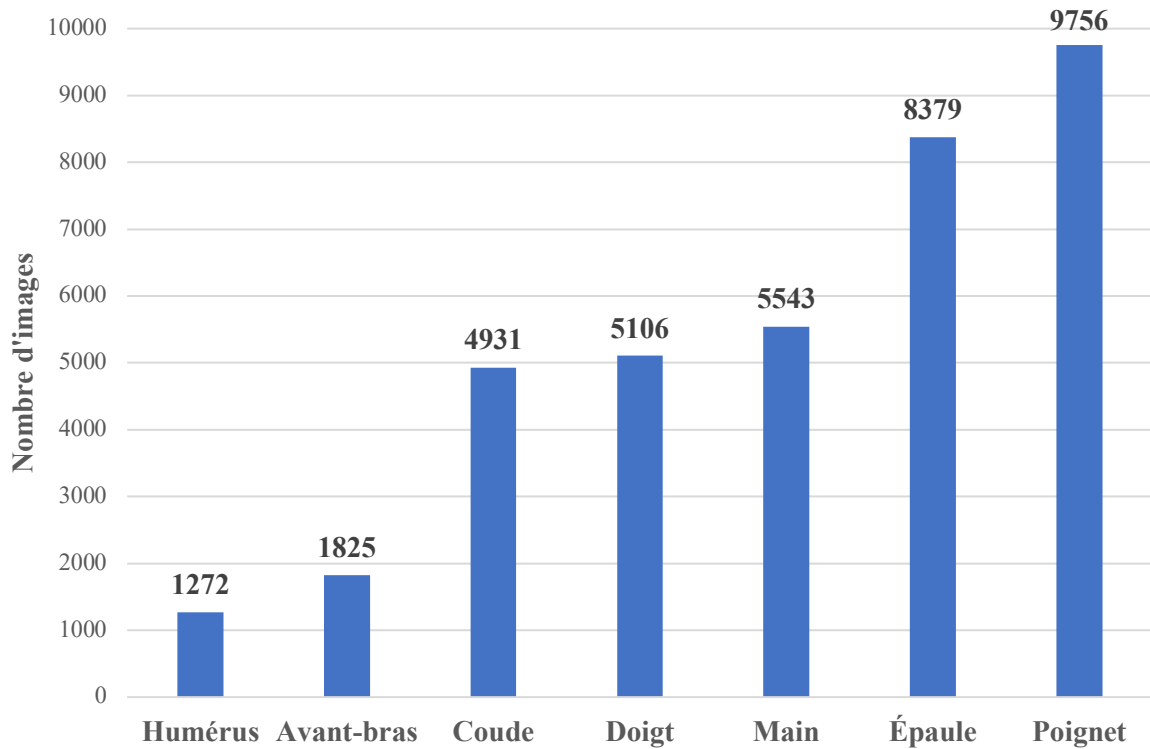
Figure 9. Exemples d'images du Dataset HEdS. a) Épaule, b) Main, c) Coude, d) Poignet, e) Humérus, f) Avant-bras.

Dataset MURA

Ce dataset comprend 40'561 images radiologiques musculo-squelettiques (graphique 3) classées par régions anatomiques et catégorisées selon la présence ou non (positive ou négative) d'une pathologie (fracture(s) et/ou prothèse) (Fig. 10). Ce travail a été effectué par 6 radiologues de l'université de Stanford dans le but d'entraîner un réseau à détecter et catégoriser les images pathologiques ou non et comparer ses performances par rapport à celles des radiologues (Rajpurkar & al., 2018).



Figure 10. Exemple d'images avec fracture et/ou prothèse (positive). Incidence Épaule de face (rotation indifférente).



Graphique 3. Base de données MURA – Répartition des classes (ordre croissant).

Étant donné la quantité d'images qui la compose, la base de données MURA ne sera pas traitée dans son intégralité durant ce travail. Nous nous sommes volontairement restreints aux radiographies d'épaules (Fig. 11), car elles sont plus pertinentes d'un point de vue pédagogique étant données les connaissances nécessaires à l'identification des différentes incidences que comprend cette région anatomique, notamment pour distinguer les degrés de rotation de l'humérus (neutre, interne, externe) sur les clichés de face.

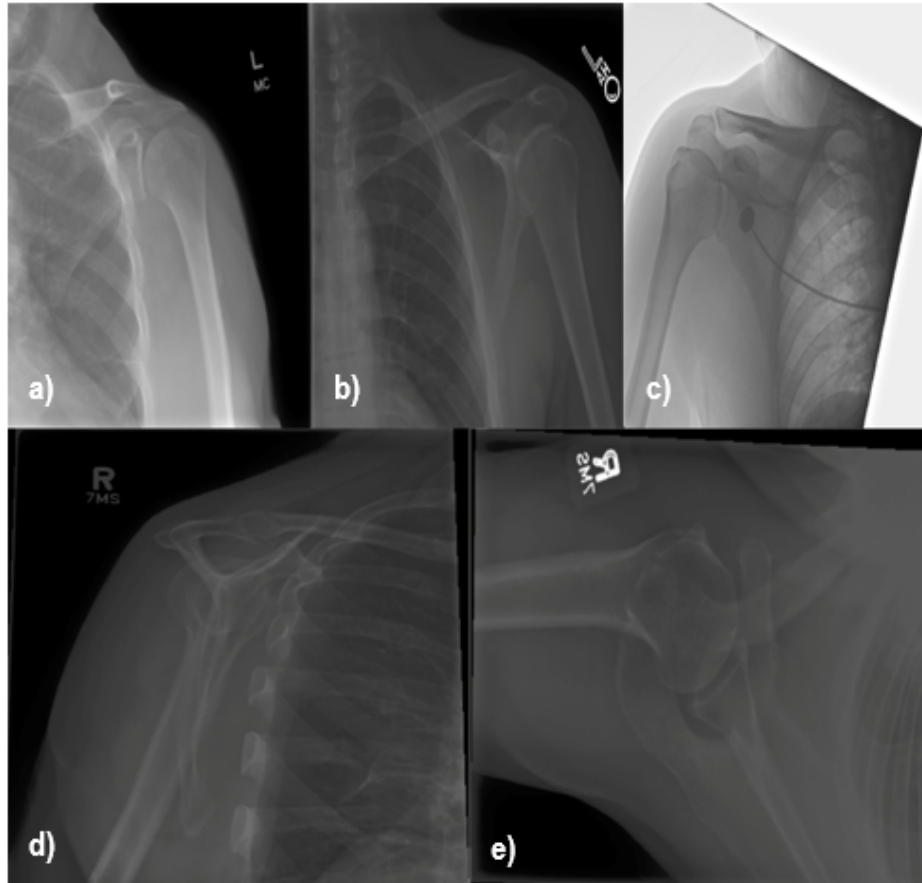


Figure 11. Exemples d'image du dataset de MURA (incidences de l'épaule). a) Épaule de face rotation neutre, b) Épaule de face rotation interne, c) Épaule de face rotation externe, d) Profil de la Coiffe des rotateurs (Neer), e) Épaule axiale.

3.4.2. Tagging

Pour rappel, le « tagging » consiste à associer une étiquette (« tag » en anglais) à une image et représente une condition essentielle à l'apprentissage supervisé. Les images peuvent être étiquetées de différentes manières selon les objectifs. Dans notre cas précis, nous avons choisi d'étiqueter les images selon plusieurs degrés en suivant une logique d'entonnoir. Cela consiste, à partir d'une étiquette globale, à aboutir à un tag plus précis. Ici, cela consiste, à partir des images du membre supérieur (tag global ou 1^{er} degré), à identifier les régions anatomiques (tag de 2^{ème} degré), puis les incidences d'épaule (tag de 3^{ème} degré).

- Tag de 1^{er} degré** → Membre supérieur
- Tag de 2^{ème} degré** → Épaule, bras, coude, poignet, avant-bras, mains / doigts
- Tag de 3^{ème} degré** → Épaule de face et rotations (neutre / interne / externe), Axiale, Profil de la coiffe des rotateurs (Neer), Acromio-claviculaire.

Démarches Dataset HEDS

Étant donné que les images étaient déjà étiquetées selon le « tag de 1^{er} degré », nous nous sommes chargés de classer ces images en différentes classes (régions anatomiques) en vue d'alimenter la base de données d'entraînement et parfaire la capacité du réseau à reproduire cette tâche de classification (graphique 2). Par la suite, la classe « main et doigts » a volontairement été retirée du dataset d'entraînement, raison pour laquelle cette classe n'apparaît pas dans les graphiques de performance (Section – 4.1.2).

Par sa faible quantité d'images (336), les données ont été organisées et triées à l'aide du logiciel Excel (Microsoft Office). Néanmoins, bien qu'utile pour le tagging d'un nombre limité d'images, ce logiciel n'est pas adapté pour l'étiquetage de larges bases de données telle que MURA.

Démarche dataset MURA

Déjà classées par régions anatomiques, nous avons passé en revue les 8379 images radiologiques de l'épaule et avons, pour chacune d'elles, attribué l'incidence qui lui correspond.

Étant donné le nombre conséquent d'images appartenant à cette classe (8379), il aurait été extrêmement fastidieux d'effectuer ce « tagging » avec le programme Excel. C'est pourquoi nous avons utilisé un logiciel d'annotation d'image nommé « VGG Image Annotator (VIA) ».

Ce programme a été développé à l'Université d'Oxford pour annoter les données (images, audio, vidéo) à travers une plateforme ergonomique (Fig. 12) accessible par un navigateur Web. Proposant de nombreuses fonctionnalités, ce logiciel a été utilisé dans plusieurs projets d'apprentissage machine (Dutta & Zisserman, 2019).

Cependant, dans le cas particulier de notre étude, ce logiciel ne permet pas de modifier le fenêtrage (contraste et luminosité) des images radiologiques pour en améliorer la qualité. Par chance, cette limitation n'a pas porté préjudice à l'étiquetage des images.

Par son ergonomie, cette application nous a offert un gain de temps considérable dans l'étiquetage des images comparé à une méthode classique consistant à relier chaque image à son étiquette dans un document annexe.

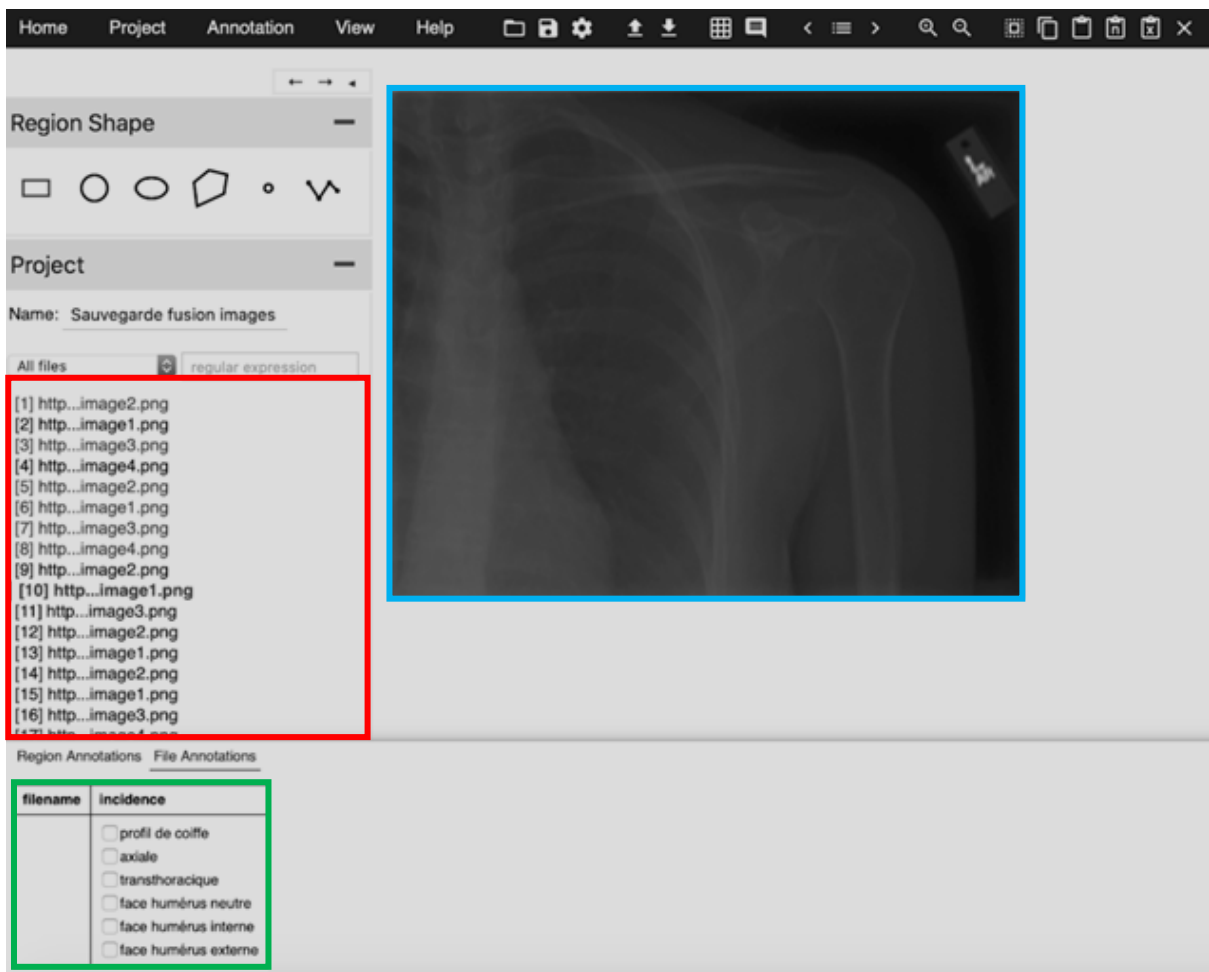


Figure 12. Interface de l'outil VIA. Encadré rouge : Base de données (images) ; encadré bleu : image sélectionnée ; encadré vert : Liste d'attributs (tags - incidences).

3.4.3. Entraînement

Les bases de données MURA et HEDS a été divisé en deux catégories distinctes (proportions en %) : Données d'entraînement (TRAIN – 80%), réservées à l'entraînement du modèle ; Données d'évaluation (TEST – 20%), réservées à l'évaluation des performances du modèle (graphique 1 - Section 2.3.1).

De plus, les données d'entraînement (TRAIN) ont été divisées en deux sous-catégories : Données d'entraînement pur (TRAIN – 80%) ; Données de validation de l'entraînement (VALID – 20%).

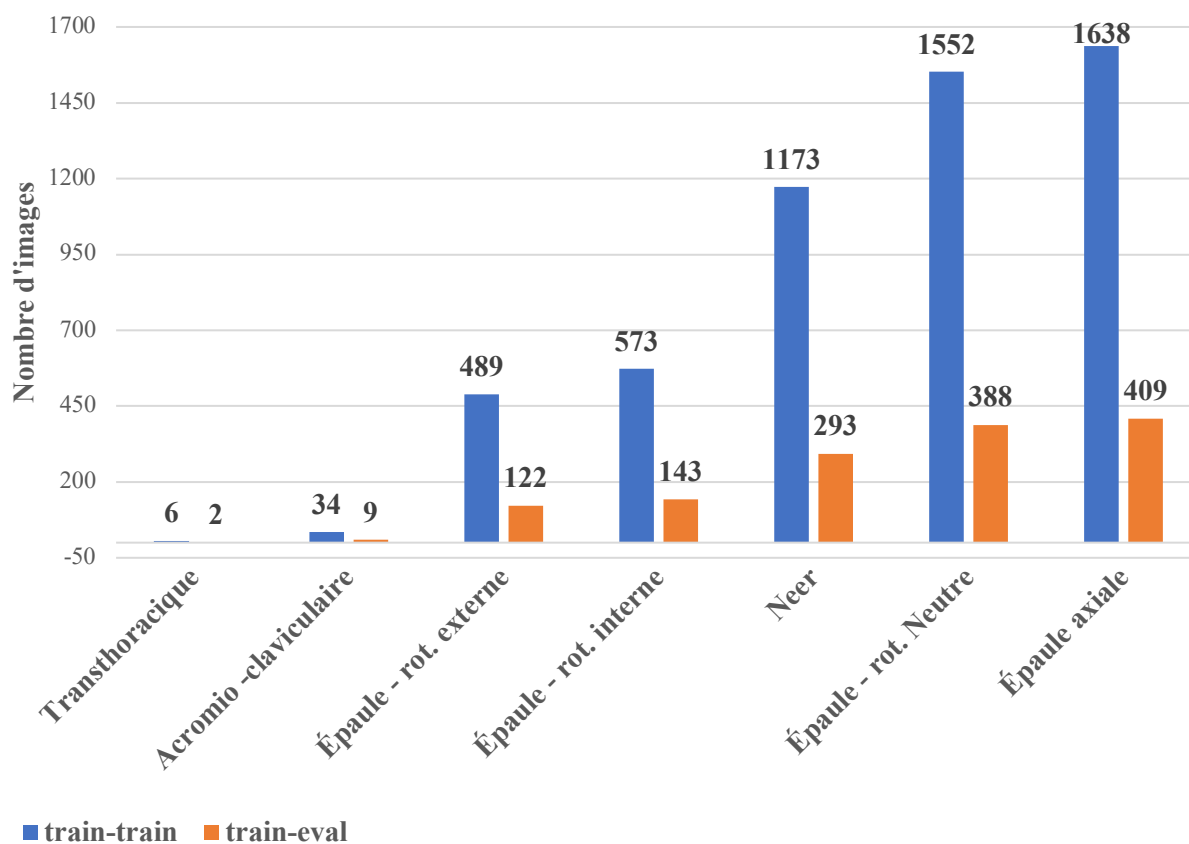
MURA régions anatomiques

Pour la détection des régions anatomiques (tag de 2^{ème} degré), le système a été entraîné sur les données d'entraînement (TRAIN MURA). Ces performances ont été évaluées sur la base de données HEdS ainsi que les données TEST MURA.

MURA incidences de l'épaule

Pour la détection des incidences d'épaule (tag de 3^{ème} degré), seule la base de données MURA a été utilisée. Le nombre d'images d'épaule HEdS étant trop faible (70). Tout comme pour la détection des régions anatomique, les données ont été divisées en deux catégories (proportions en %) : Données d'entraînement (train-train – 80%), dédiées à l'entraînement pur ; Données d'évaluation (train-éval – 20%) dédiées à l'évaluation des performances du modèle (graphique 4).

Les données d'entraînement (train-train) comprennent 2 sous catégories : Données d'entraînement pur (train – 80%) ; Données de validation de l'entraînement (valid – 20%).



Graphique 4. Base de données MURA – Répartition des données TRAIN / TEST (ordre croissant).

3.4.4. Architectures d'IA sélectionnées

En vue d'accomplir cette tâche de classification, deux réseaux de neurones profonds ont été sélectionnés : Inception V4 et NasNet-a-large.

Ces réseaux profonds de types « convolutionnels » comprennent des couches optimisant le traitement et l'extraction des caractéristiques de l'image (Section 2.2.6).

Inception V4

Proposé par Google, Inception est un réseau dont le développement a permis une avancée majeure dans l'amélioration des performances des réseaux convolutionnels. Il s'agit d'un réseau sophistiqué basé sur des stratégies d'optimisation complexes permettant d'améliorer la précision et la vitesse du réseau. Les différentes stratégies mises en œuvre ont donné naissance à plusieurs versions : Inception V1, Inception V2, Inception V3, Inception V4, Inception ResNet (Raj, 2018).

L'architecture de base (V1) contient 22 couches de convolution et 5 millions de paramètres. Cela étant, sa dernière version (V4) contient 43 millions de paramètres permettant une nette amélioration des performances d'entraînement (Karim, 2019).

Les informations qui définissent une image sont très variables et nécessitent une adaptation du choix des paramètres du réseau (nombre de couche, filtres, taille de matrice de convolution, etc.). De ce fait, ce réseau compile une multitude d'opérations avec différents paramètres permettant une extraction des informations plus performante d'une convolution à l'autre, limitant ainsi le choix des paramètres par l'opérateur. Le réseau étant déjà performant dans cette tâche, il devient alors possible de l'appliquer sur une base de données de nature différente aux données d'entraînement (principe de transfert de l'apprentissage). Ainsi, seule la portion du réseau de classification (couches entièrement connectées) est entraînée sur la nouvelle tâche de classification (Sarkan, 2018).

NASNet-a-large

Le modèle NasNet (Neural Architecture Search Network), également proposé par Google, est un type de réseau différent dans la conception de son architecture qui est générée automatiquement en fonction des données auxquelles nous l'appliquons. En effet, bien que son architecture de base soit prédéfinie, ce réseau, composé de « blocs » réalisant les opérations de convolutions classiques (Kaenampornpan & al., 2018), est capable de générer sa propre architecture grâce à une méthode dite « AutoML » basée sur l'utilisation d'un réseau de neurones récurrent (Recurrent Neural Network en anglais) capable de réarranger ces blocs et de recréer une architecture de bout en bout et sans intervention humaine (Seif, 2018).

Plusieurs de ces blocs se retrouvent en partie dans l'architecture Inception, mais leur organisation s'effectue manuellement. Un tel réseau offre une plus grande flexibilité et de meilleures performances sur une nouvelle base de données tout en offrant une bonne transférabilité (Zoph & al., 2018).

Ce réseau pré-entraîné sur la base de données « ImageNet » possédant 1000 classes, a été utilisé sur nos images d'épaules. La dernière couche du réseau étant dépendante du nombre de classes utilisées, nous avons apporté des modifications afin que cela corresponde à nos 12 classes (2^{ème} et 3^{ème} degré).

Stratégies d'entraînement

Pour améliorer l'entraînement et les performances de nos modèles, différentes stratégies et paramétrages ont été appliqués (tableau 3).

| Stratégies utilisées | Caractéristiques |
|---|--|
| Data augmentation (par défaut) | <i>Pour des images quelconques (couleur ou non)</i> |
| | Translation (x ; y), rotation, changement d'échelle (agrandissement et rétrécissement), changement de perspective, luminosité, contraste, niveau de compression JPEG (artéfact), teinte / saturation (couleur), courbe gamma |
| Data augmentation Xray (Radiologique) | <i>Pour des images radiologiques (noir et blanc)</i> |
| | Translation (x ; y), rotation, changement d'échelle (agrandissement et rétrécissement), changement de perspective, rehaussement de bords, luminosité, contraste, miroir horizontal (flip) |
| Balancing | Ré-équilibrer les différentes classes entre elles (Upsampling) |
| No validation (Full) | Utilisation de l'ensemble des données d'entraînement (y compris VALID) et pas de validation en cours d'entraînement. |
| k-folds | Méthode de validation croisée (tableau 2 - Section 2.3.1) |

Tableau 3 – Caractéristiques des stratégies d'entraînement.

3.4.5. Validation / Test (évaluation du modèle déjà entraîné)

Pour chaque stratégie d'entraînement, nous avons procédé à l'évaluation du modèle avec les données « TEST ». Cela dans le but d'analyser l'efficacité des différentes stratégies d'entraînements et d'identifier les paramètres les mieux adaptés à la tâche de classification d'images radiologiques. En vue de comparer ces différentes stratégies, 11 évaluations (tests) ont été effectuées sur la base de données MURA (tableau 4).

Pour toutes les stratégies évoquées, nous avons utilisé 20 « Epoch (e) » et un « Batch size (bs) » de 18 (tableau 1 - Section 2.3.1).

| Architecture | | Stratégie | | | | Paramètres |
|-------------------------------|-----------------------|-------------------|---------------------|---------------|---------|-------------|
| | | Data augmentation | Balancing | No validation | k-folds | |
| Régions anatomiques | | | | | | |
| 1) | NASNet_a_large | Par défaut | X | X | X | bs 18 / e20 |
| 2) | Inception V4 | Par défaut | X | X | X | bs 18 / e20 |
| 3) | NASNet_a_large | X | X | X | X | bs 18 / e20 |
| Incidences de l'épaule | | | | | | |
| 4) | NASNet_a_large | Par défaut | Sur-échantillonnage | X | X | bs 18 / e20 |
| 5) | NASNet_a_large | Par défaut | X | X | X | bs 18 / e20 |
| 6) | Inception V4 | Par défaut | Sur-échantillonnage | X | X | bs 18 / e20 |
| 7) | NASNet_a_large | X | X | X | X | bs 18 / e20 |
| 8) | NASNet_a_large | Xray | Sur-échantillonnage | Full | 5 | bs 18 / e20 |
| 9) | NASNet_a_large | Xray | Sur-échantillonnage | X | X | bs 18 / e20 |
| 10) | NASNet_a_large | Par défaut | Sur-échantillonnage | Full | X | bs 18 / e20 |
| 11) | Inception V4 | Par défaut | X | X | X | bs 18 / e20 |

Tableau 4. Approches stratégiques réalisées (modèles).

Pour analyser les performances de notre réseau, nous utiliserons les indicateurs quantitatifs suivants :

- Exactitude (Accuracy) globale
- Exactitude (Accuracy) par classe

Pour pousser notre analyse plus loin, nous allons également nous appuyer sur les rapports d'erreurs (annexe III) fournis à l'issue de chaque entraînement et procéder, lorsque cela est possible, à une analyse visuelle des échecs en exploitant la méthode d'interprétation dite de CAM pour révéler les zones auxquelles le modèle a accordé le plus d'intérêt pour effectuer sa prédiction (Section 2.2.5).

3.5. Étude 2

3.5.1. Élaboration travail pratique (Bsc 3^{ème} année)

Pour donner suite à l'étude 1, nous avons pensé qu'il serait pertinent de mettre en place un travail pratique (TP) pour les étudiants de 3^{ème} année. Ce travail a été préparé plusieurs semaines à l'avance en vue de l'implémenter à la suite d'un cours d'introduction à l'intelligence artificielle enseigné par M. Christophe Chênes afin d'introduire les concepts de bases de cette discipline.

Ce TP avait pour but de faire participer, à leur insu, des élèves à une tâche de labellisation d'images de la base de données MURA à l'aide du même logiciel que nous avons utilisé (VGG Images Annotator (VIA)) (Section 3.4.2) et, de les instruire au concept d'apprentissage supervisé au travers du « tagging indirect ». Les données récoltées au travers de ce TP nous ont permis d'évaluer les performances de l'échantillon d'étudiants dans cette tâche et de les comparer aux modèles de réseau de neurones entraînés (Section 3.4.5).

Réalisé dans le plus grand secret, ce TP s'est déroulé sur une demi-matinée (deux périodes de 45 minutes) et a demandé la participation de 21 étudiants (volée 2017), dont les auteurs de ce travail. Nous avons formé des équipes en binômes (10) et une personne seule (observatrice). Chaque étudiant / groupe était assigné à un poste informatique.

En vue de rendre ce TP plus attractif et ludique, nous avons déposé devant chaque poste une feuille attribuant à chaque membre et binôme un pseudonyme de superhéros (ex. Wolverine), le groupe de héros auquel il appartient (ex. Xmen) ainsi que son logo (ex. logo Xmen) (annexe IX).

Nous avons conçu ce TP en 4 phases et rédigé un protocole (annexe VII) décrivant leurs caractéristiques :

- Durée
- Équipe
- Description tâche
- Base de données (images)
- Quantité
- Consignes (générales et spécifiques)
- But

Pour faciliter les manœuvres informatiques liées à l'utilisation du logiciel de tagging et éviter toute erreur de la part des étudiants, nous leur avons conçu et distribué une feuille de consignes (annexe VIII) détaillant la marche à suivre pour chaque étape :

1. Préparation de l'outil de tagging ;
2. Chargement des images et liste d'attribut (tag) ;
3. Consigne par phase et par groupe (ex. Taguer les images en vous référant aux consignes dans l'enveloppe) ;
4. Sauvegarde du fichier « tagging » dans le bon répertoire ;

Chaque fichier sauvegardé contient l'ensemble des « tags » réalisé par l'étudiant ou le groupe au terme de chaque phase. Dans l'intervalle, ces résultats étaient traités et organisés dans un tableur Excel afin de générer, pour chaque attribut, des graphiques (annexe X) reflétant les performances des étudiants ou groupes :

- Nombre d'images taguées
- Précision
- Accuracy
- Courbe ROC

La veille du TP, nous avons réalisé une simulation sur les postes informatiques afin de vérifier que le programme fonctionne correctement et élaborer une stratégie de récolte d'informations pendant le TP.

En effet, n'étant pas informés du lien existant entre ce TP et notre Travail de bachelor, les étudiants ne se doutaient pas que l'observateur avait pour but de prendre note de leurs commentaires et réactions au cours du débriefing réalisé par notre directeur de travail entre chaque phase du TP.

3.5.2. Déroulement du travail pratique

Phase 0

Cette phase, d'une durée de 15 minutes, est effectuée de façon individuelle par les étudiants. Elle consiste à les familiariser avec le logiciel de « tagging » en vue des étapes suivantes.

Durant cette étape, les étudiants ont pour tâche d'associer différentes images de félins aux 5 classes correspondantes (ex. chat, guépard, lion, lynx, tigre) (Fig. 13). Certaines images piègeuses (ex. Fig. 13e) ou représentant deux classes (ex. Fig. 13f), sont volontairement présentes pour complexifier la tâche et révéler des disparités dans les choix des étudiants. Ces derniers ont pour consigne de taguer correctement les 32 images (annexe XI).



Figure 13. Exemple d'images de félins. a), Chat, b), Guépard, c), Lion, d), Lynx, e), Image piègeuse, f), Image à deux classes.

Phase 1

Cette phase d'une durée de 15 minutes est effectuée en binôme (10 groupes + une personne seule). Elle consiste à taguer des images radiologiques du membre supérieur par régions anatomiques (tag de 2^{ème} degré) (Section 3.4.2).

Nous avons utilisé la base de données MURA contenant des radiographies des membres supérieurs déjà classées par régions anatomiques (graphique 3 – Section 3.4.1). Seules 3 classes ont été retenues (épaule, coude et main). Les données sont équilibrées de sorte que les différentes classes apparaissent de façon régulière et plus au moins égale.

Nous avons volontairement choisi d'utiliser un nombre conséquent d'images (18'853) afin que les étudiants n'arrivent pas à traiter l'ensemble dans le temps imparti et que les groupes puissent se démarquer davantage.

Chaque groupe reçoit des consignes spécifiques contenues dans une enveloppe :

- Taguer tout correctement ; (19 enveloppes)
- Taguer tout faux ; (1 enveloppe)
- Taguer partiellement faux (1 erreur sur 3) ; (1 enveloppe)

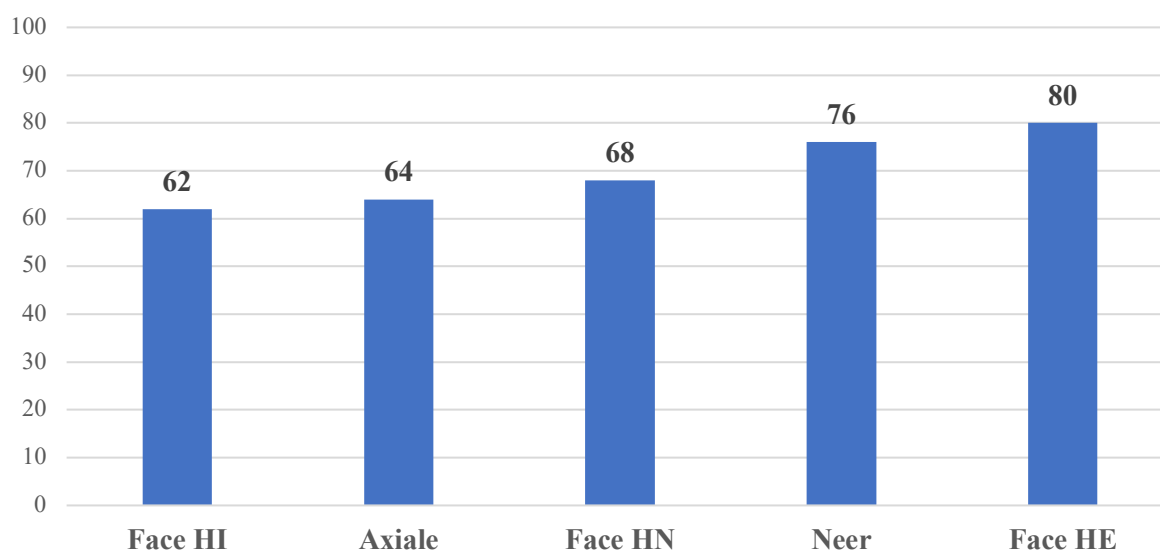
La consigne générale est de taguer correctement le maximum d'images dans le temps imparti.

Cette phase comporte 3 objectifs principaux. Premièrement, l'expérience est cette fois-ci axée sur des images radiologiques, faisant plus de sens pour des étudiants TRM. Deuxièmement, les consignes spécifiques créent des variations au sein de l'échantillon et démontrent des performances excellentes comme négligentes (expert vs profane), de même que leur capacité à respecter les consignes. Troisièmement, cette étape introduit le concept de « tagging indirect » en leur présentant des images dont nous connaissons initialement le « tag ». Démontrant l'importance de faire appel à des personnes munies de connaissances adaptées à la labellisation d'une telle base de données d'images.

Phase 2

Cette phase d'une durée de 20 minutes est effectuée en binôme (10 groupes + une personne seule). Elle consiste à taguer des images radiologiques d'épaule par incidences (tag de 3^{ème} degré) (Section 3.4.2).

Nous avons aléatoirement récupéré 350 images issues de la base de données MURA contenant les radiographies d'épaule classées par incidences. Seules 5 classes ont été retenues (Axiale, Face-HE, Face-HI, Face-HI, Neer), les deux autres (Acromio-claviculaire et Trans-thoracique) ayant un nombre d'images trop faible. Les données sont équilibrées de sorte que les différentes classes apparaissent de façon régulière et plus au moins égale (graphique 5).



Graphique 5. Phase 2 - Incidences d'épaule – Répartition des données.

La consigne générale est de taguer correctement le maximum d'images dans le temps imparti.

Le but principal de cette phase est de montrer un exemple de tagging plus complexe et révélateur des difficultés que peuvent rencontrer les TRM dans un cadre pédagogique ou professionnel.

Phase 3

Presque identique à la phase précédente, cette phase d'une durée de 20 minutes est effectuée individuellement. Elle consiste à taguer des images radiologiques d'épaule par incidences (tag de 3^{ème} degré).

Nous avons réutilisé les 350 images de la « phase 2 » issues de la base de données MURA contenant des radiographies d'épaule classées par incidences. Les données sont toujours équilibrées de sorte que les différentes classes apparaissent de façon régulière et plus au moins égale.

Certaines incidences d'épaule sont difficiles à caractériser, notamment les différents degrés de rotation de l'humérus (neutre, interne et externe) sur les radiographies de l'épaule de face. C'est pourquoi entre la phase 2 et 3, une « fiche de rappel » (annexe XII), récapitulant les critères anatomiques de l'humérus en fonction de sa rotation, est distribuée aux étudiants afin qu'ils puissent s'améliorer dans la tâche.

La consigne générale est de taguer correctement le maximum d'images dans le temps imparti.

Le but principal de cette phase est d'observer une hausse des performances de la part des étudiants après avoir pris connaissance de la fiche de rappel et après avoir acquis de l'aisance dans l'utilisation du logiciel VIA durant les 3 précédentes phases.

3.5.3. Fin du travail pratique

À la fin de la matinée, un bilan général nous a permis de révéler à l'ensemble de la classe notre implication réelle dans ce TP et son importance pour notre Travail de bachelor. Le but de cette procédure de tagging et son rôle dans l'entraînement d'un réseau de neurones profond leur ont été exposés au travers des explications fournies afin de leur faire prendre conscience des compétences que requiert cette tâche et du rôle que peuvent jouer les professionnels et futurs professionnels du domaine concerné.

Les étudiants ont été sollicités pour donner leur avis sur différentes thématiques liées au TP et son potentiel dans la formation et la recherche. Après quoi, nous les avons chaleureusement remerciés pour leur intérêt et leur participation tout au long de la matinée.

4. Résultats

4.1. Étude 1 : Performances des différentes stratégies d'IA

4.1.1. Qualité d'images

Afin de comparer les différentes stratégies d'entraînement, il semble important de comparer la qualité des données ayant participé à leur entraînement et évaluation. Comme nous l'avons vu, les jeux de données HEDS et MURA ne sont pas équivalents en termes de quantité (336 vs 40'561). De plus, leur qualité respective diffère.

Dataset HEDS

Le positionnement est optimal et respecte les critères de réussites (Fig. 14). Le niveau de contraste ainsi que la luminosité sont bons et nous permettent de différencier les différentes densités tissulaires (graisse, os, muscle et air). Malgré les conditions précaires de réalisation engendrées par la présence de certaines pathologies (fractures, matériel d'ostéosynthèse, etc.), la qualité des images demeure largement acceptable. Cependant, il existe certaines radiographies de moindre qualité (Fig. 15).



Figure 14. Image Dataset HEDS bonne qualité. a), Épaule, b), Poignet, c), Main et doigts de face, d), Avant-bras, e), Humérus, f), Coude.

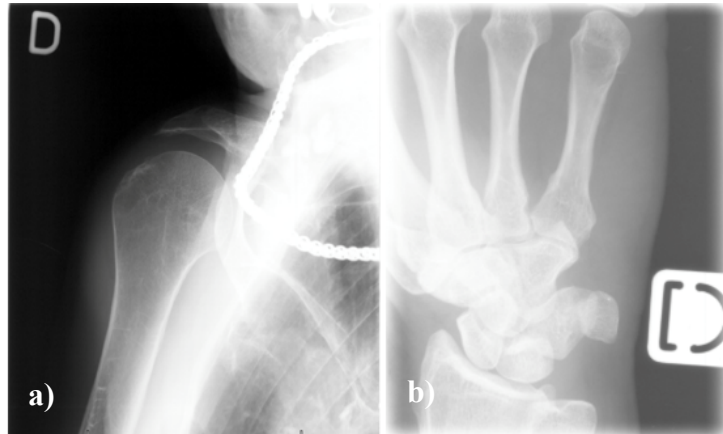


Figure 15. Image Dataset HEdS mauvaise qualité. a), Épaule de face (artéfact métallique), b), Poignet mal diaphragmé ou autre incidence.

Dataset MURA

Provenant des États-Unis (USA), les images qui composent cette base de données sont de qualités variables et répondent parfois à des normes et/ou critères radiologiques différents. La luminosité, le contraste et la netteté ne sont pas toujours optimaux (Fig. 17), de même que le centrage ou le cadrage (diaphragmes). Pour cette raison, certaines incidences sont difficiles à identifier et présentent un aspect peu harmonieux.

De plus, elle englobe une multitude de clichés de moindre qualité justifiés par la présence de pathologies (fractures) ou de matériel d'ostéosynthèse (Fig. 18) rendant difficile le respect des critères positionnement radiologiques, utiles dans le cadre de notre étude. En effet, la présence d'anomalies anatomo-pathologiques et d'artéfacts métalliques pourrait empêcher le réseau de neurones d'acquiescer une bonne représentation des images radiologiques et des incidences.

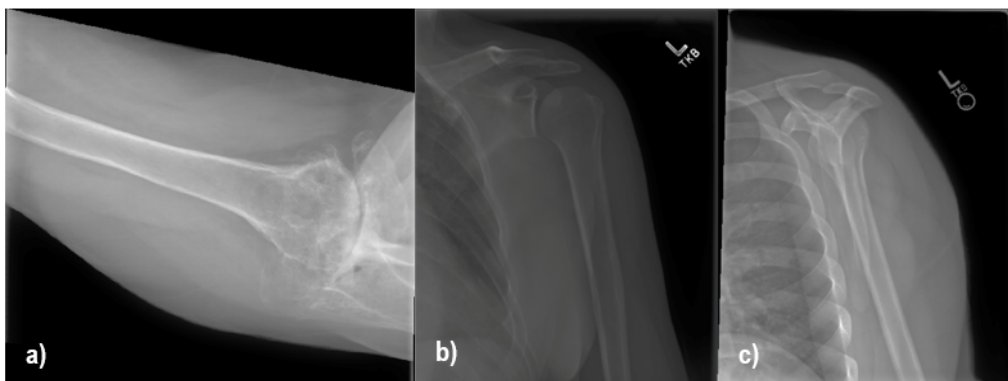


Figure 16. Images Dataset MURA bonne qualité (incidences de l'épaule) : a), Épaule en axiale, b), Épaule de face rotation externe, c), Profils de coiffe (Neer).

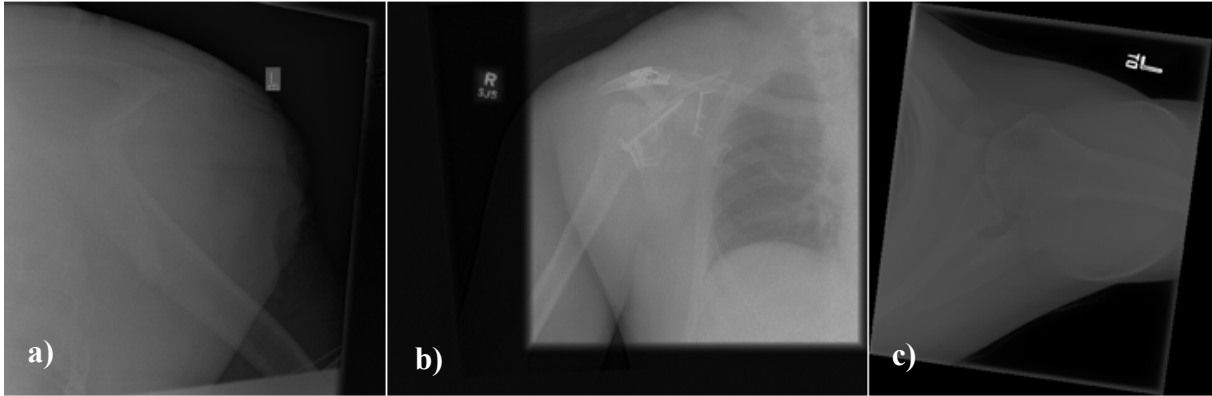


Figure 17. Images Dataset MURA mauvaise qualité. a), Profils de coiffe (Neer), b), Épaule de face rotation interne, c), Épaule en axiale.

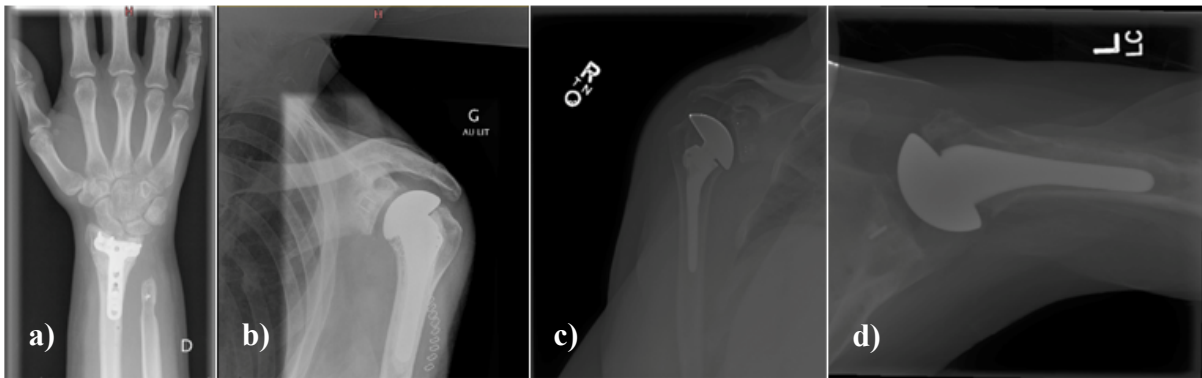
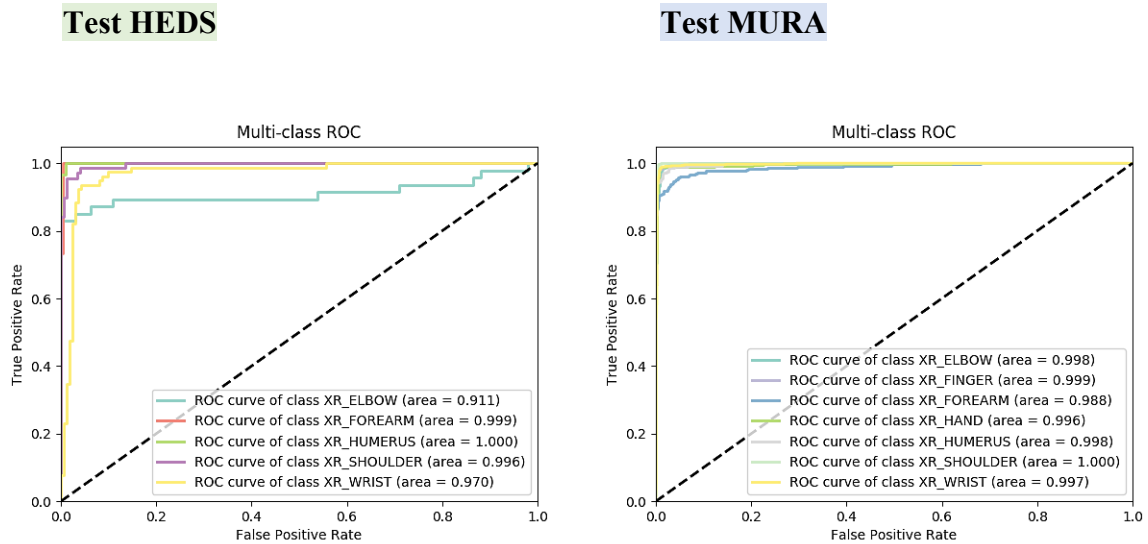


Figure 18. Image avec matériel d'ostéosynthèse. a), Poignet de face. b), Épaule de face rotation neutre, c), Épaule de face rotation neutre, d), Épaule en axiale.

4.1.2. Régions anatomiques

Voici les résultats (tableaux 5, 6 et 7) obtenus en fonction des différentes approches (stratégies) appliquées pour la tâche de classification par régions anatomique (2^{ème} degré).



Graphique 6. Exemple : Courbe ROC multi-classes (test n°1 Dataset MURA).

| N° modèle | Accuracy par classe | | | | | | | Accuracy globale |
|-----------|---------------------|------------|---------|--------|---------|-------|--------|------------------|
| | Coude | Avant-bras | Humérus | Épaule | Poignet | Main | Doigts | |
| 1) | 0.808 | 1.00 | 1.00 | 0.956 | 0.884 | X | X | 0.912 |
| | 0.963 | 0.870 | 0.944 | 0.995 | 0.986 | 0.970 | 0.972 | 0.965 |
| 2) | 0.872 | 1.00 | 0.774 | 0.971 | 0.833 | X | X | 0.883 |
| | 0.978 | 0.867 | 0.938 | 0.995 | 0.983 | 0.974 | 0.970 | 0.966 |
| 3) | 0.970 | 0.854 | 0.924 | 0.998 | 0.983 | 0.983 | 0.974 | 0.965 |

Tableau 5. Résumé performances (Accuracy par classe et globale).

| | |
|------------------|-----------------|
| Test HEDS (vert) | Valeur minimale |
| Test MURA (bleu) | Valeur maximale |

| N° modèle | Coude | Avant-bras | Humérus | Épaule | Poignet | Main | Doigts | Total |
|-----------|-------|------------|---------|--------|---------|------|--------|-------|
| 1) | 9 | 0 | 0 | 3 | 9 | X | X | 21 |
| | 17 | 39 | 16 | 3 | 9 | 14 | 13 | 111 |
| 2) | 6 | 0 | 7 | 2 | 13 | X | X | 28 |
| | 10 | 40 | 18 | 3 | 11 | 12 | 14 | 108 |
| 3) | 14 | 44 | 22 | 1 | 11 | 8 | 12 | 112 |

Tableau 6. Résumé des erreurs.

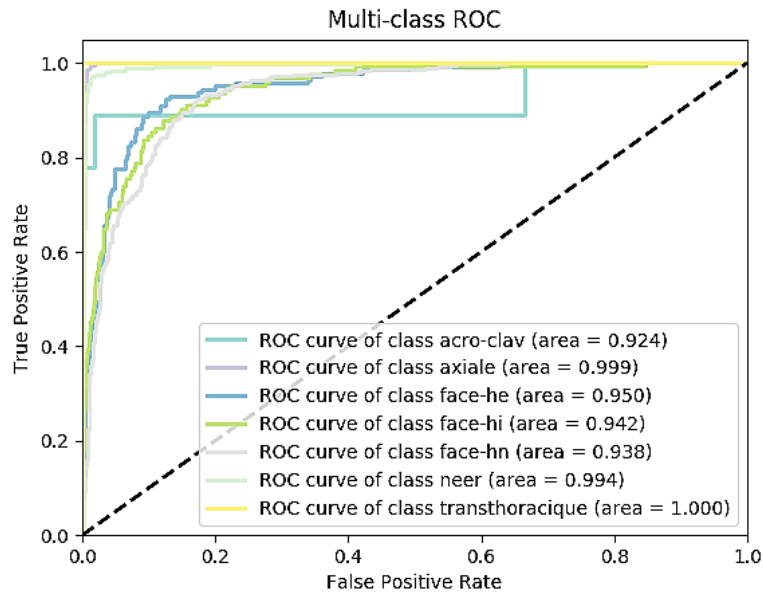
| | | Erreurs | | | | | | |
|------------|-------|------------|---------|--------|---------|------|--------|--|
| Régions | Coude | Avant-bras | Humérus | Épaule | Poignet | Main | Doigts | |
| Coude | | 4 | 5 | 5 | | | | |
| Avant-bras | 13 | | 4 | 8 | 16 | 1 | | |
| Humérus | 6 | 8 | | 5 | 1 | | 1 | |
| Épaule | | | 1 | | | | | |
| Poignet | | 2 | | 1 | | 7 | 1 | |
| Main | | | | | 7 | | 1 | |
| Doigts | | | | 1 | 1 | 10 | | |

Tableau 7. Exemple de confusion des classes (modèle n°3).

Nous observons (tableau 5) que les performances sont variables entre les différentes approches. Étant entraîné sur la base de données MURA, nous remarquons que les performances des modèles testés sur les radiographies issues de ce même dataset sont meilleures, avec une Accuracy quasi-parfaite pour la région de l'épaule. Cependant, nous remarquons que la présence de deux classes supplémentaires (Main et Doigts) entraîne une augmentation du nombre d'erreurs (tableau 6) et une baisse de l'Accuracy pour la classe « Avant-bras » qui est confondue avec les classes « poignet » et « coude » (annexe IV), ces deux régions apparaissant dans les clichés de l'avant-bras. Nous pouvons également noter que, pour les tests « HEDS », la région anatomique du coude révèle des performances inférieures. Le modèle le confond majoritairement avec le poignet, et ce indépendamment de la stratégie appliquée.

4.1.3. Incidences

Voici les résultats (tableau 8, 9 et 10) obtenus en fonction des différentes approches (stratégies) appliquées pour la tâche de classification par incidence d'épaule (3^{ème} degré).



Graphique 7. Exemple : Courbe ROC multi-classes (Test n°9).

| N° modèle | Accuracy par classe | | | | | | | Accuracy (ACC) |
|-----------|---------------------|--------|---------|---------|---------|-------|------------------|----------------|
| | Acro-clav | Axiale | Face HE | Face HI | Face HN | Neer | Trans-thoracique | |
| 4) | 0.778 | 1.00 | 0.734 | 0.795 | 0.639 | 0.952 | 0 | 0.838 |
| 5) | 0.889 | 0.992 | 0.056 | 0.582 | 0.861 | 0.969 | 0 | 0.813 |
| 6) | 0.444 | 0.988 | 0.839 | 0.672 | 0.533 | 0.969 | 0 | 0.806 |
| 7) | 0.444 | 0.995 | 0.224 | 0.434 | 0.873 | 0.959 | 0 | 0.816 |
| 8) | 0.666 | 0.992 | 0.489 | 0.680 | 0.752 | 0.959 | 0 | 0.833 |
| 9) | 0.777 | 0.995 | 0.594 | 0.713 | 0.729 | 0.955 | 0.500 | 0.842 |
| 10) | 0.666 | 0.995 | 0.657 | 0.680 | 0.703 | 0.959 | 0 | 0.837 |
| 11) | 0.666 | 0.992 | 0.706 | 0.606 | 0.734 | 0.966 | 0 | 0.845 |

Tableau 8. Résumé performances (Accuracy par classe et globale).

| N° modèle | Acro-clav | Axiale | Face HE | Face HI | Face HN | Neer | Trans-thoracique | Total |
|-----------|-----------|--------|---------|---------|---------|------|------------------|-------|
| 4) | 2 | 0 | 38 | 25 | 140 | 14 | 2 | 221 |
| 5) | 1 | 3 | 135 | 51 | 54 | 9 | 2 | 255 |
| 6) | 5 | 5 | 23 | 40 | 181 | 9 | 2 | 265 |
| 7) | 5 | 2 | 111 | 69 | 49 | 12 | 2 | 250 |
| 8) | 3 | 3 | 73 | 39 | 96 | 11 | 2 | 228 |
| 9) | 2 | 2 | 58 | 35 | 105 | 13 | 1 | 216 |
| 10) | 3 | 2 | 49 | 39 | 115 | 12 | 2 | 222 |
| 11) | 3 | 3 | 42 | 48 | 103 | 10 | 2 | 211 |

Tableau 9. Résumé des erreurs.

| Incidences | Erreurs | | | | | | |
|------------|---------|--------|----|----|----|------|--------|
| | Acro. | Axiale | HE | HI | HN | Neer | Trans. |
| Acro. | | 1 | | 1 | | | |
| Axiale | | | | 1 | 1 | | |
| HE | | | | 8 | 46 | 4 | |
| HI | | 4 | 1 | | 28 | 2 | |
| HN | | 11 | 29 | 62 | | 3 | |
| Neer | | | 6 | 3 | 4 | | |
| Trans. | | | | | | 1 | |

Tableau 10. Exemple de confusion des classes (test n°9).

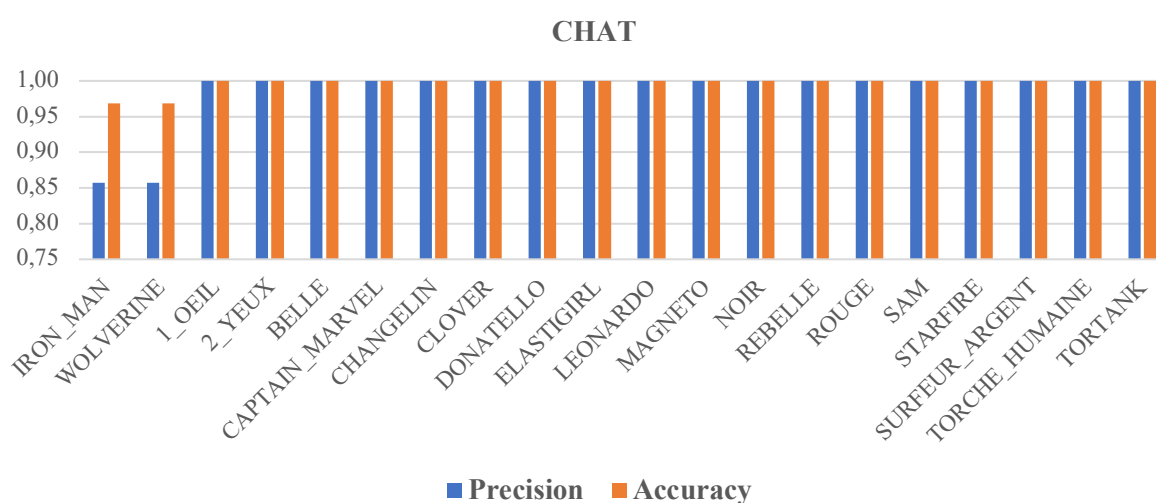
De façon générale, nous pouvons remarquer une baisse significative des performances entre la classification par régions anatomiques et par incidences, la complexité étant objectivement plus grande entre ces deux tâches. La plus grande difficulté semble portée sur les incidences de l'épaule de face et les détections des degrés de rotations de l'humérus (neutre, interne et externe), en témoigne le nombre d'erreurs commises sur ces 3 classes (tableau 9), notamment sur les rotations neutres qui sont majoritairement confondues avec les rotations internes (annexe IV). Néanmoins, les différents modèles montrent une plus grande facilité pour reconnaître l'incidence « axiale » et démontrent une Accuracy quasi-maximale.

4.2. Étude 2 : Travail pratique - Performances étudiants

4.2.1. Phase 0 - Félines

20 personnes ont tagué 32 images de félines en 15 minutes avec la consigne de taguer correctement.

Pour chacune des classes, nous avons recueilli les valeurs d'Accuracy et de Précision en les classant de la plus faible (min.) à la plus élevée (max.) (graphique 8).



Graphique 8. Exemple valeur d'Accuracy et précision par équipe (ordre croissant).

Afin de montrer les écarts de performance entre les différents étudiants, nous avons répertorié les valeurs d'Accuracy et de Précision minimale (min.) et maximales (max.) pour chaque classe (tableau 11).

| | Chat | Guépard | Lion | Lynx | Tigre |
|-----------------------|------|---------|------|------|-------|
| Accuracy max. | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Accuracy min. | 0.97 | 0.94 | 0.91 | 0.94 | 0.91 |
| Précision max. | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Précision min. | 0.86 | 0.83 | 0.86 | 0.80 | 0.75 |

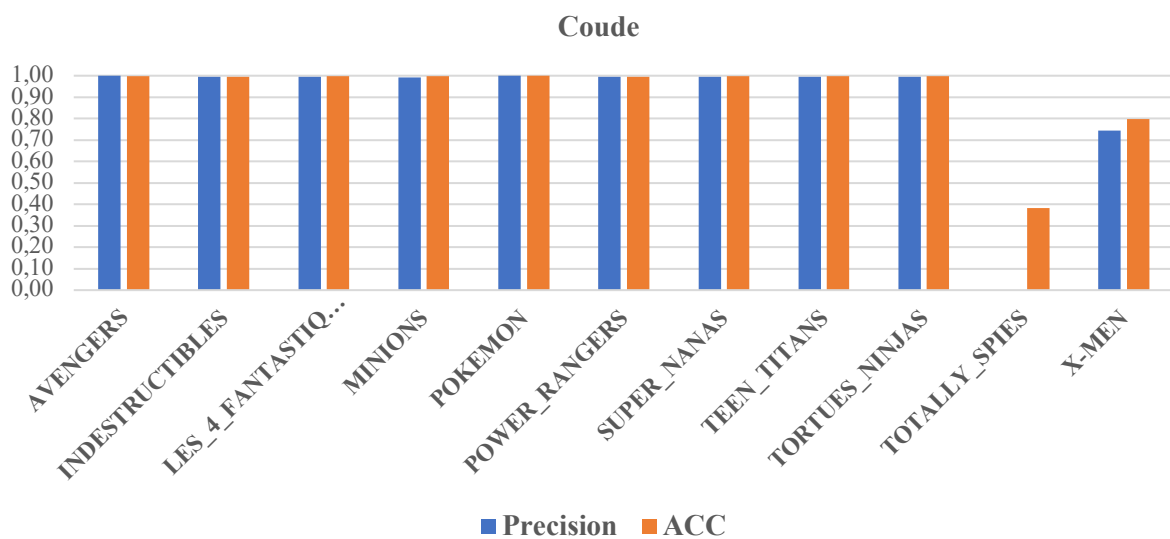
Tableau 11. Résumé Accuracy et Précision minimale (min.) et maximale (max.).

4.2.2. Phase 1- Régions anatomiques

11 équipes de deux personnes ont tagué entre 285 et 486 images par régions anatomiques en 15 minutes avec des consignes variables.



Graphique 9. Nombre d'images taguées par équipe (ordre croissant).



Graphique 10. Exemple valeur d'Accuracy et précision par équipe (ordre croissant).

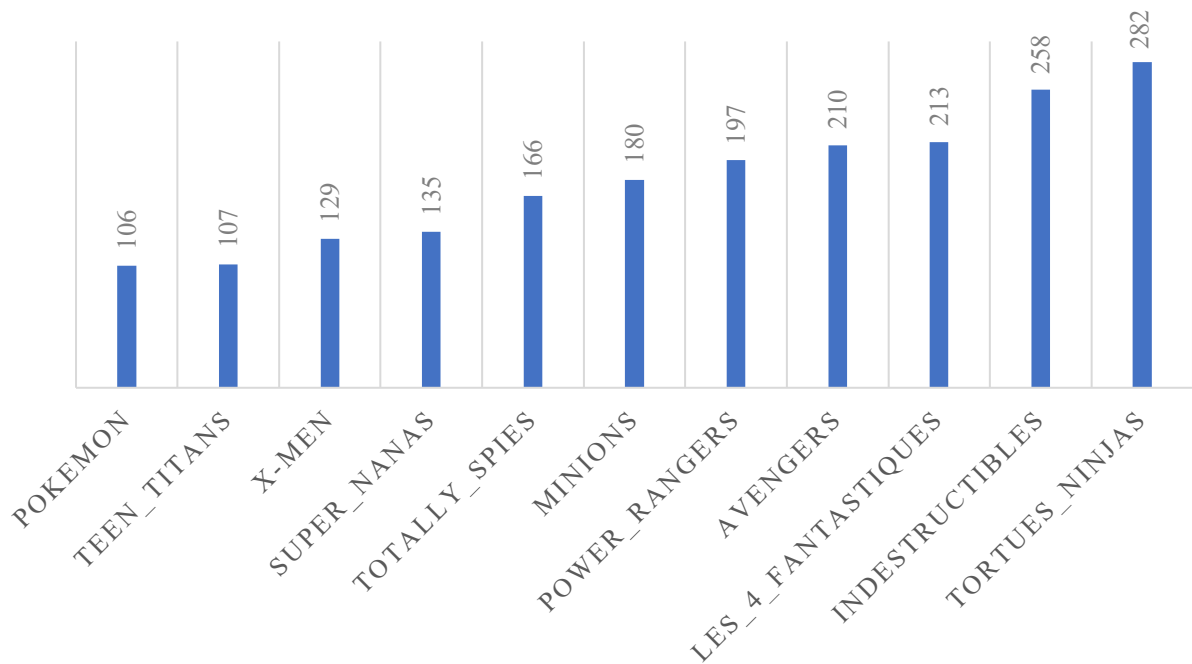
Afin de montrer les écarts de performance entre les différents étudiants, nous répertorié les valeurs d'Accuracy et de Précision minimale (min.) et maximales (max.) pour chaque classe en fonction des consignes reçues (tableau 12).

| | Épaule | Coude | Main |
|--|--------|-------|------|
| Consigne : Taguer tout faux | | | |
| Accuracy | 0.32 | 0.38 | 0.29 |
| Précision | 0 | 0 | 0 |
| Consigne : Taguer partiellement juste | | | |
| Accuracy | 0.76 | 0.80 | 0.85 |
| Précision | 0.64 | 0.75 | 0.73 |
| Consigne : Taguer tout juste | | | |
| Accuracy max. | 1.00 | 1.00 | 1.00 |
| Accuracy min. | 1.00 | 1.00 | 1.00 |
| Précision max. | 1.00 | 1.00 | 1.00 |
| Précision min. | 0.99 | 0.99 | 0.99 |

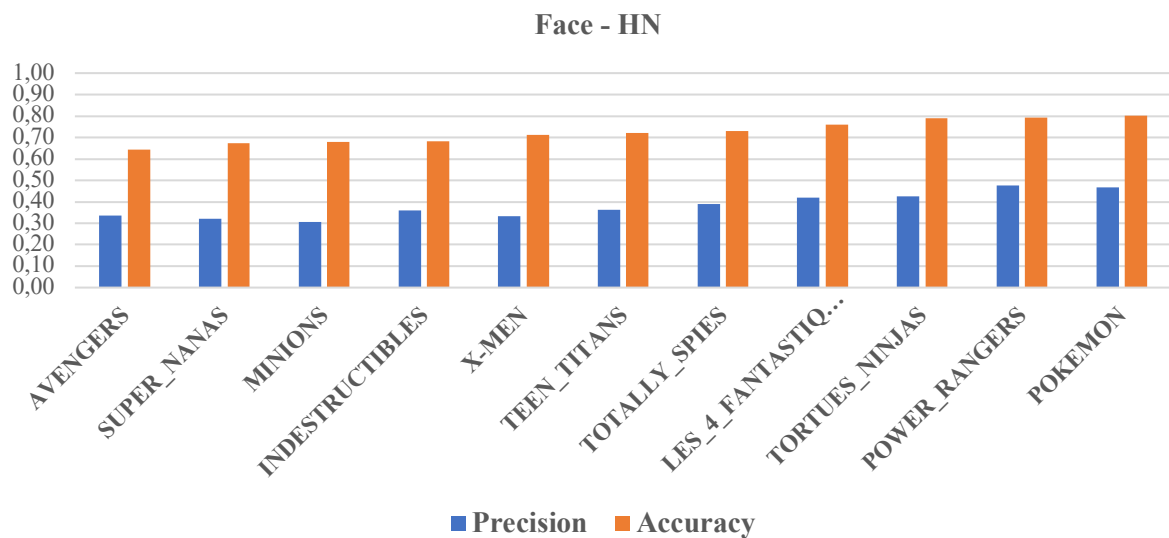
Tableau 12. Résumé Accuracy et Précision minimale (min.) et maximale (max.).

4.2.3. Phase 2 - Incidences de l'épaule (sans support)

11 équipes de deux personnes ont tagué entre 106 et 282 images par incidences d'épaule en 20 minutes avec la consigne de taguer correctement le maximum d'image dans le temps imparti.



Graphique 11. Nombre d'images taguées par équipe (ordre croissant).



Graphique 12. Exemple valeur d'Accuracy et précision par équipe (ordre croissant).

Afin de montrer les écarts de performance entre les différents étudiants, nous avons répertorié les valeurs d'Accuracy et de Précision minimale (min.) et maximales (max.) pour chaque classe (tableau 13) et réalisé un tableau montrant les performances moyennes de l'ensemble des équipes pour chaque classe (tableau 14).

| | Neer | Axiale | Face-HN | Face-HI | Face-HE |
|-----------------------|------|--------|---------|---------|---------|
| Accuracy max. | 1.00 | 1.00 | 0.80 | 0.88 | 0.85 |
| Accuracy min. | 0.97 | 0.98 | 0.67 | 0.63 | 0.69 |
| Précision max. | 1.00 | 1.00 | 0.48 | 0.71 | 0.83 |
| Précision min. | 0.86 | 0.98 | 0.31 | 0.14 | 0.13 |

Tableau 13. Résumé Accuracy et Précision minimale (min.) et maximale (max.).

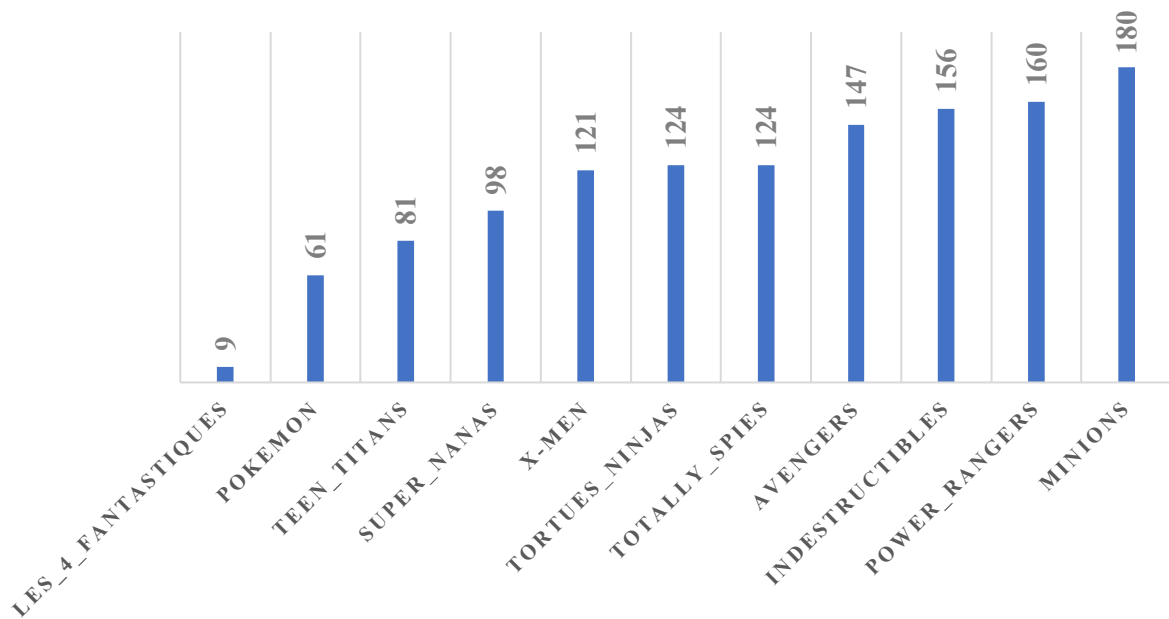
Remarque : Nous avons été obligés de retirer l'équipe des « 4 fantastiques », car leurs résultats ont été impactés par un problème informatique de telle sorte que seules 9 images ont été taguées lors de la phase suivante. Afin d'obtenir des résultats plus fiables, nous avons décidé de retirer cette équipe lors des calculs des Accuracy pour les deux dernières phases (tableau 15 et 16).

| Incidences | Accuracy globale |
|---------------------------------|------------------|
| Profils de coiffe (Neer) | 0.986 |
| Axiale | 0.993 |
| Face-HN | 0.721 |
| Face-HI | 0.786 |
| Face-HE | 0.764 |
| Total | 0.850 |

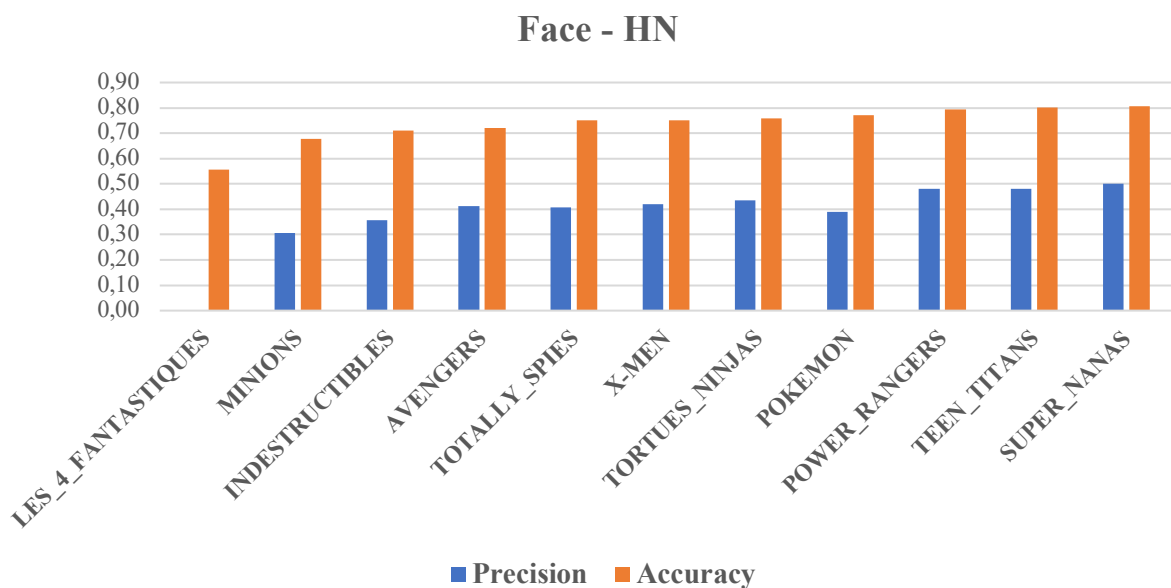
Tableau 14. Phase 2 - Valeurs d'Accuracy globale par classe (sans fiche de rappel).

4.2.4. Phase 3 - Incidences de l'épaule (avec support)

11 équipes de deux personnes ont tagué entre 61 et 180 images par incidences d'épaule en 20 minutes avec la consigne de taguer correctement le maximum d'image dans le temps imparti. La tâche est identique à la phase 2, mais cette fois-ci, les étudiants se voient distribués une fiche de rappel décrivant les critères radiologiques des différents degrés de rotation de l'épaule de face.



Graphique 13. Nombre d'images taguées par équipe (ordre croissant).



Graphique 14. Exemple valeur d'Accuracy et précision par équipe (ordre croissant).

Tout comme pour la phase précédente, nous avons répertorié les valeurs d'Accuracy et de Précision minimale (min.) et maximales (max.) pour chaque classe (tableau 15) et réalisé un tableau montrant les performances moyennes de l'ensemble des équipes pour chaque classe (tableau 16).

Le but étant de voir si la fiche de positionnement distribuée entre ces deux dernières phases entraîne une amélioration des performances des étudiants.

| | Neer | Axiale | Face-HN | Face-HI | Face-HE |
|-----------------------|------|--------|---------|---------|---------|
| Accuracy max. | 1.00 | 1.00 | 0.81 | 0.90 | 0.88 |
| Accuracy min. | 0.67 | 0.44 | 0.56 | 0.78 | 0.69 |
| Précision max. | 1.00 | 1.00 | 0.50 | 0.86 | 0.86 |
| Précision min. | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 |

Tableau 15. Résumé Accuracy et Précision minimale (min.) et maximale (max.).

| Incidences | Accuracy globale |
|---------------------------------|------------------|
| Profils de coiffe (neer) | 0.989 |
| Axiale | 0.992 |
| Face-HN | 0.754 |
| Face-HI | 0.842 |
| Face-HE | 0.814 |
| Total | 0.873 |

Tableau 16. Phase 3 - Valeurs d'Accuracy globale par classe (avec fiche de positionnement radiologique).

4.2.5. Travail pratique - Performance IA

En vue de comparer les résultats obtenus par les étudiants avec ceux obtenus par l'IA, nous avons sélectionné 2 modèles (n°4 et n°9) et testé leur performance (tableau 17) sur les données traitées par les étudiants lors de la phase 2 et 3 (incidences d'épaule). Ainsi, ces modèles ont été ré-entraînés sur toutes les radiographies d'épaule du Dataset MURA, puis testés sur les 350 images taguées lors de la phase 2 et 3 du TP.

| N° modèle | Accuracy par classe | | | | | Accuracy Globale |
|-----------|---------------------|---------|---------|---------|-------|------------------|
| | Axiale | Face HE | Face HI | Face HN | Neer | |
| 4) | 0.984 | 0.525 | 0.258 | 0.853 | 0.973 | 0.723 |
| 9) | 1.00 | 0.625 | 0.322 | 0.764 | 0.987 | 0.745 |

Tableau 17. Résumé performances (Accuracy par classe et globale).

| N° modèle | Axiale | Face HE | Face HI | Face HN | Neer | Total |
|-----------|--------|---------|---------|---------|------|-------|
| 4) | 1 | 38 | 46 | 10 | 2 | 97 |
| 9) | 0 | 30 | 42 | 16 | 1 | 89 |

Tableau 18. Résumé des erreurs.

| Incidences | Erreurs | | | | |
|------------|---------|----|----|----|------|
| | Axiale | HE | HI | HN | Neer |
| Axiale | | | 1 | | |
| HE | | | 2 | 35 | 1 |
| HI | | 3 | | 43 | |
| HN | | 5 | 2 | | 3 |
| Neer | | | 2 | | |

Tableau 19. Exemple :Matrice de confusion des classes (modèle n°4).

5. Analyse & Discussion

5.1. Étude 1 - Analyse des performances de l'IA

5.1.1. Influence de l'architecture réseau

Dans le cadre de cette étude, plusieurs tests ont été réalisés avec deux architectures différentes (Inception V4 et NasNet_a_large) afin de définir laquelle est la plus performante pour classifier nos images radiologiques selon différents degrés de tag (régions anatomiques vs incidences d'épaule). Afin de rendre cette analyse cohérente, nous avons sélectionné et comparé différents modèles ayant bénéficié de la même stratégie d'entraînement (tableau 3 – Section 3.4.4). De cette manière, il devient possible de comparer les performances de ces architectures, indépendamment de la stratégie appliquée.

En fonction des Accuracy obtenus dans les tableaux 5 (Section 4.1.2) et 8 (Section 4.1.3), voici les comparaisons effectuées (tableau 20).

Test HEDS

Test MURA

| Inception V4 | | Nasnet_a_large | |
|---------------------|-------|----------------|-------|
| Régions anatomiques | | | |
| Modèle n°2 | | Modèle n°1 | |
| 0.883 | 0.966 | 0.912 | 0.965 |
| Incidences d'épaule | | | |
| Modèle n°6 | | Modèle n°4 | |
| 0.806 | | 0.838 | |
| Modèle n°11 | | Modèle n°5 | |
| 0.845 | | 0.813 | |

Tableau 20. Régions anatomiques : Comparaison des performances (Accuracy globale).

La comparaison des différentes valeurs d'Accuracy obtenues (tableau 20) peine à fournir une démarcation nette des performances d'une architecture par rapport à l'autre. Pour le tag de 2^{ème} degré (régions anatomiques), les valeurs maximales obtenues sont très proches. Cependant, nous remarquons que le réseau Nasnet_a_large obtient de meilleures performances (+3,28 %) sur le dataset HEDS comparé au réseau Inception V4, ce qui témoigne une meilleure capacité d'adaptation de ce réseau à des données d'une autre provenance (HEDS vs MURA), bien que leurs performances soient quasi-équivalentes sur le dataset MURA.

Cependant, cet argument est remis en question lorsqu'il s'agit des tags de 3^{ème} degré (incidences d'épaule). En effet, le réseau Inception V4 semble montrer de meilleures performances sur cette tâche (0.845 vs 0.813) lorsque la stratégie d'entraînement utilise peu de paramètres. Inversement, le réseau Nasnet_a_large est meilleur (0.838 vs 0.806) lorsque la stratégie est plus riche en paramètres.

5.1.2. Influence de la stratégie

Pour entraîner un réseau profond de manière efficace, il importe de tenir compte de l'ensemble des facteurs qui influencent ses futures performances. Parmi ces facteurs, les données représentent un élément déterminant, que ce soit en termes de quantité ou de qualité. Deux réseaux entraînés à reproduire une même tâche avec des données différentes auront des performances diamétralement opposées si l'un ou l'autre de ces facteurs venait à faire défaut.

De la même façon qu'un radiologue acquiert de l'expérience en analysant des milliers d'images radiologiques, un réseau de neurones doit lui aussi analyser beaucoup d'images, si ce n'est davantage, pour simplement reconnaître l'image et la classer dans la bonne catégorie. À cela s'ajoute la qualité qui, pour des images radiologiques, est déterminante tant pour l'interprétation humaine à des fins diagnostics que pour la reconnaissance du cliché par un réseau de neurones profonds. En effet, dans la réalité, certains contextes (ex. traumatologique) peuvent rendre le respect de ces critères difficiles. À cela s'ajoutent les variations intrinsèques propres aux patients (ex. anatomiques), aux centres radiologiques, aux opérateurs et aux constructeurs qui peuvent affecter l'aspect général du cliché (dimensions, centrage, contraste, luminosité, harmonie, etc).

En partant du postulat que nos données sont de qualité hétérogène et que la quantité d'images par catégorie peut limiter l'apprentissage et la reconnaissance de certaines classes par rapport à

d'autres, la question sous-jacente serait d'estimer quelle stratégie est la plus adaptée pour permettre de réaliser ces différentes tâches de classification.

Pour cela, nous allons comparer différentes stratégies d'entraînement (tableau 4 -Section 3.4.5), et ce pour une même architecture réseau (Nasnet_a_large) afin de démontrer son impact sur les performances du modèle.

Data augmentation

Pour rappel, il s'agit d'une stratégie visant à augmenter artificiellement la quantité de données en y appliquant de légères modifications (Section 2.4.1).

Afin de mesurer son impact sur les performances, nous avons comparé différents modèles entraînés avec et sans ce procédé (tableau 21).

| Nasnet_a_large | |
|--|--|
| Avec Data augmentation (par défaut) | Sans Data augmentation (par défaut) |
| Régions anatomiques | |
| Modèle n°1 | Modèle n°3 |
| 0.9653 | 0.9649 |
| Incidences d'épaule | |
| Modèle n°5 | Modèle n°7 |
| 0.813 | 0.816 |

Tableau 21. Data augmentation : Comparaison des performances (Accurarcy globale).

Lorsqu'il s'agit de différencier des régions anatomiques, nous remarquons que l'augmentation de données entraîne une très légère différence sur les performances du réseau (+0,041 %). Pour cette tâche, l'augmentation de données ne permet donc pas une amélioration significative des performances. Toutefois, bien que les performances globales semblent identiques, nous remarquons que l'Accuracy par classe (tableau 5 – Section 4.1.2) est légèrement supérieure pour les classes minoritaires (humérus et avant-bras), cependant, pour les classes majoritaires (coude, doigts, main, épaule et poignet), cela diminue légèrement les performances. Pour un dataset de qualité discutable, comme c'est le cas pour la base de données MURA, nous pouvons donc penser que cette stratégie est belle et bien bénéfique pour les classes faiblement représentées, mais qu'elle peut être contre-productive si les données sont déjà nombreuses.

Concernant les incidences d'épaule, nous observons que cette stratégie influence de façon négative les performances du modèle 5 qui sont inférieures (-0.37 %) par rapport au modèle 7. Bien que cette différence soit négligeable, nous pouvons penser que, tout comme pour la tâche de classification par régions anatomiques, la qualité des données peut limiter, voire inverser, l'efficacité de cette stratégie pour des tâches plus complexes comme c'est le cas pour la classification par incidence d'épaule.

Data augmentation Xray

Nous avons vu qu'il existe différents processus de Data augmentation (tableau 3 – Section 3.4.4). En fonction des images traitées, l'application d'une stratégie par défaut semble peu pertinente. Dans ce contexte, une autre stratégie plus adaptée aux images radiologiques (tableau 3 – Section 3.4.4) devrait améliorer les performances.

Afin de témoigner de son efficacité, deux modèles ont été comparés (tableau 22).

| Nasnet_a_large | |
|---|-------------------------------------|
| Data augmentation (par défaut) | Data augmentation (Xray) |
| Incidences | |
| Modèle n°4 | Modèle n°9 |
| 0.838 | 0.842 |

Tableau 22. Data augmentation Xray : Comparaison des performances (Accuracy globale).

Nous remarquons que le fait de tenir compte de la nature des images radiologiques, en appliquant des transformations spécifiques à ces images, semble améliorer légèrement les performances du réseau (+0,48 %).

La base de données MURA contient une quantité d'images relativement élevée permettant déjà d'atteindre des performances proches de 100%. Le fait d'augmenter la quantité de données permet au mieux d'améliorer faiblement les performances et, au pire, d'y nuire si elle n'est pas adaptée aux données, ou lorsque la tâche de classification devient plus complexe (régions anatomiques vs incidences). En ce sens, le choix d'une stratégie ciblée (Xray) semble plus indiqué lorsque la tâche est plus complexe, même si l'amélioration des performances demeure assez faible. Cependant, il est intéressant de remarquer que le modèle n°9 est le seul modèle ayant réussi à classer correctement l'une des deux images appartenant à la classe « Trans-thoracique ». Ce modèle est d'ailleurs 2^{ème} dans le classement des performances récoltées dans le cadre de cette étude. Raison pour laquelle il retient tout notre intérêt.

Équilibrage des données

La base de données MURA englobe une multitude d'images qui, une fois catégorisées, révèle un déséquilibre entre les différentes classes. Ce déséquilibre peut engendrer des performances variables en fonction des classes et de leur occurrence au sein du dataset d'entraînement.

Le procédé de « balancing » consiste à augmenter le nombre de données au sein des classes les moins représentées afin d'augmenter les performances globales du réseau. Pour démontrer l'impact isolé de cette stratégie, nous avons comparé plusieurs modèles et architectures ayant ou n'ayant pas recours à ce procédé (tableau 23).

| Incidences | |
|-----------------------|--------------------|
| Avec Balancing | Sans Balancing |
| Nasnet_a_large | |
| Modèle n°4 | Modèle n°5 |
| 0.838 | 0.813 |
| Inception V4 | |
| Modèle n°6 | Modèle n°11 |
| 0.806 | 0.845 |

Tableau 23. Balancing : Comparaison des performances (Accuracy globale).

L'utilisation de cette stratégie semble avoir un impact variable en fonction de l'architecture réseau sélectionnée. En effet, elle semble avoir un impact positif sur l'architecture Nasnet_a_large (+3,08 %), mais négatif sur l'architecture Inception V4 (-4,96 %).

De plus, si nous regardons l'évolution des performances sur les classes minoritaires (tableau 8 – Section 4.1.3), nous remarquons que cette stratégie n'engendre pas systématiquement une amélioration des performances. En prenant l'exemple de la classe « Acromio-claviculaire » (34 images), nous remarquons une baisse des performances pour le réseau Nasnet_a_large (-12,49 %) et Inception V4 (-33,33 %). Cependant, pour les classes intermédiaires, par exemple pour l'épaule de face en rotation interne (573 images), nous remarquons une hausse des performances pour Nasnet_a_large (+36,59 %) et Inception V4 (+10,89%). Il semblerait donc que cette stratégie soit plus efficace sur l'architecture Nasnet_a_large et de surcroît sur les classes de moyenne occurrence.

Full Training Set

Cette méthode (tableau 3 – Section 3.4.4), qui s’affranchit de la validation, consiste à utiliser l’ensemble des données ne servant pas à l’évaluation finale. Ainsi, les données normalement destinées à la validation (VALID) sont ajoutées au dataset d’entraînement (TRAIN), ce qui permet de potentialiser l’entraînement du modèle. Le reste servant à l’évaluation des performances (TEST).

Dans le cadre de cette étude, seuls deux modèles (n°8 et n°10) ont bénéficié de cette stratégie (tableau 4 – Section 3.4.5). Afin de juger de son efficacité, nous allons comparer deux modèles ayant pour seule différence l’utilisation de cette méthode (tableau 24).

| Nasnet_a_large | |
|-------------------------------|-------------------------------|
| Sans Full Training Set | Avec Full Training Set |
| Incidences | |
| Modèle n°4 | Modèle n°10 |
| 0.838 | 0.837 |

Tableau 24. Full training set : Comparaison des performances (Accuracy globale).

Nous remarquons que le recours à cette stratégie n’améliore pas les performances du modèle n°10 (-0,12%). Excepté pour les incidences d’épaules de face en rotation neutre (+10,1%) et Neer (+0,73%), les performances ont diminué partout ailleurs. En termes de quantité, ces deux classes sont les plus représentées au sein du dataset MURA, ce qui peut expliquer que le fait d’utiliser davantage de données de mauvaises qualités n’améliorent pas les performances.

5.1.3. Tâches de classification

Pour répondre à l'hypothèse selon laquelle le tagging des incidences est plus complexe que celui par régions anatomiques, nous avons comparé l'évolution des performances de plusieurs modèles ayant les mêmes caractéristiques (architecture et stratégie d'entraînement) en fonction de la tâche réalisée (tableau 25).

| Régions anatomiques | Incidences d'épaule |
|-----------------------|---------------------|
| Nasnet_a_large | |
| Modèle n°1 | Modèle n°5 |
| 0.965 | 0.813 |
| Modèle n°3 | Modèle n°7 |
| 0.965 | 0.816 |
| Inception V4 | |
| Modèle n°2 | Modèle n°11 |
| 0.966 | 0.845 |

Tableau 25. Régions anatomiques vs Incidences d'épaule : Comparaison des performances (Accuracy globale).

Comme attendu, l'évolution à la baisse des performances entre la classification par régions anatomiques et par incidences d'épaule témoigne de la différence de complexité entre ces deux tâches, indépendamment de l'architecture et de la stratégie appliquées. Cependant, nous remarquons que des deux architectures, Inception V4 est celle qui est le moins impactée par cette transition avec une baisse de performance d'environ 12,5 % contre ~ 15,5 % pour Nasnet_a_large. Nous pouvons donc affirmer que l'IA a davantage de facilité à différencier des radiographies selon la région anatomique, mais qu'elle peine encore à identifier parfaitement certaines incidences, notamment les 3 degrés de rotation de l'humérus sur les clichés d'épaule de face où elle commet proportionnellement beaucoup d'erreurs (annexe III).

Afin de mettre en évidence cette tendance à la baisse, indépendamment du fait que la tâche soit réalisée par un humain ou une machine, nous avons comparé l'évolution de la moyenne des performances globales obtenues par les étudiants (tableau 26), toutes classes confondues, entre les phases 1 (tableau 12 – Section 4.2.2) et 2 (tableau 14 – Section 4.2.3) du travail pratique.

| | Régions anatomiques (phase 1) | Incidences d'épaule (phase 2) |
|------------------|----------------------------------|----------------------------------|
| Accuracy moyenne | 1.00 | 0.850 |

Tableau 26. Phase 1 et 2 : Comparaison des performances des étudiants (Accuracy globale).

Au travers des résultats obtenus entre ces deux phases, nous remarquons également une baisse des performances des étudiants de l'ordre de 15 %. Naturellement, la reconnaissance des régions anatomiques est une tâche aisée pour des étudiants de 3^{ème} année, en témoigne l'Accuracy parfaite (1,00). Néanmoins, nous observons une plus grande disparité dans la labellisation des incidences d'épaule, et plus spécifiquement dans la différenciation des 3 degrés de rotation de l'humérus sur les clichés d'épaule de face, tout comme l'IA. Toutefois, les étudiants obtiennent proportionnellement de meilleurs résultats que l'IA.

Qu'il s'agisse d'être humain ou d'IA, la tâche de classification d'images radiologiques en catégories de complexité variable reflète à peu près les mêmes difficultés. Comme attendu, la classification par incidence nécessite une analyse plus subtile, ce qui accroît le risque de commettre des erreurs. Ce constat met en avant la nécessité de mobiliser des connaissances en imagerie médicale pour traiter ce type de données.

En faveur de notre hypothèse, le facteur temps constitue un argument de choix pour illustrer cette hausse de complexité. En effet, la durée imposée à chaque phase nous permet de mesurer le temps moyen nécessaire pour catégoriser une image.

Ainsi, pour la phase 1, les étudiants ont tagué entre 285 et 486 images selon leur appartenance au tag de 2^{ème} degré, ce qui correspond à un temps de réflexion compris entre 1,8 et 3,1 secondes par image.

Pour la phase 2, les étudiants ont tagué entre 106 et 282 images selon leur appartenance au tag de 3^{ème} degré, ce qui correspond à un temps de réflexion compris entre 4,25 et 11,3 secondes, soit 2 à 4 fois plus que pour la phase 1. Cette variation est principalement due au temps que nécessite la différenciation des degrés de rotation de l'humérus sur les clichés d'épaule de face. Le tableau 12 (Section 4.2.2) et 14 (Section 4.2.3) nous montre que, globalement, l'incidence de l'épaule de face avec l'humérus en rotation neutre (HN) est la plus compliquée à reconnaître. Tandis que sur ce dataset, l'IA semble avoir plus de peine avec les épaules de face en rotation interne (tableau 17 – Section 4.2.5).

Grâce à cette analyse, nous pouvons dire que la rotation de la tête humérale est la chose la plus difficile à identifier sur une radiographie de l'épaule de face. Cette difficulté est partagée entre les étudiants et l'IA, ce qui montre que même avec une bonne connaissance des images radiologiques, la tâche de labellisation peut s'avérer complexe et requiert une analyse minutieuse.

En effet, afin de distinguer correctement ces 3 incidences, il est important de connaître parfaitement l'anatomie de cette région. L'unique différence entre ces 3 clichés réside dans l'aspect de la tête humérale (Fig. 19) et la projection de 3 structures osseuses :

1. Tubercule majeur (Trochiter)
2. Coulisserie bicipitale (entre les deux tubercules)
3. Tubercule mineur (Trochin)

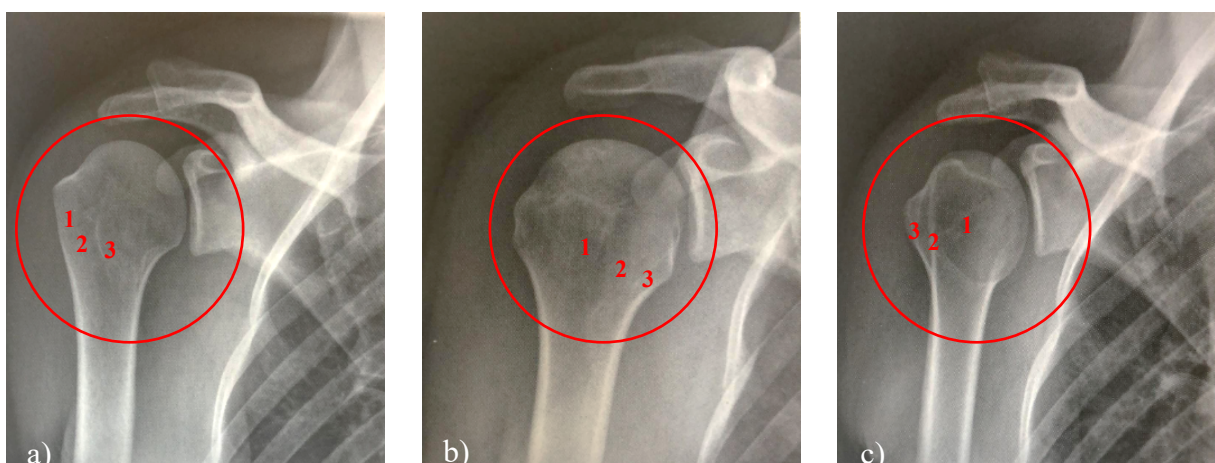


Figure 19. Incidences de l'épaule de face. a) Rotation neutre, b) Rotation interne, c) Rotation externe.

Pour identifier correctement ces incidences, les personnes entraînées savent exactement où regarder. Cependant, il paraît moins évident de comprendre à quoi l'IA prête attention pour reconnaître ces 3 incidences. Heureusement, grâce aux CAM (Class Activation Map) il est possible de voir les zones auxquelles le réseau de neurones a prêté le plus d'attention pour prédire la classe. Nous avons donc repris quelques exemples de « CAM » pour illustrer les capacités du réseau à percevoir les zones d'intérêts propres à ces incidences (Fig. 20).

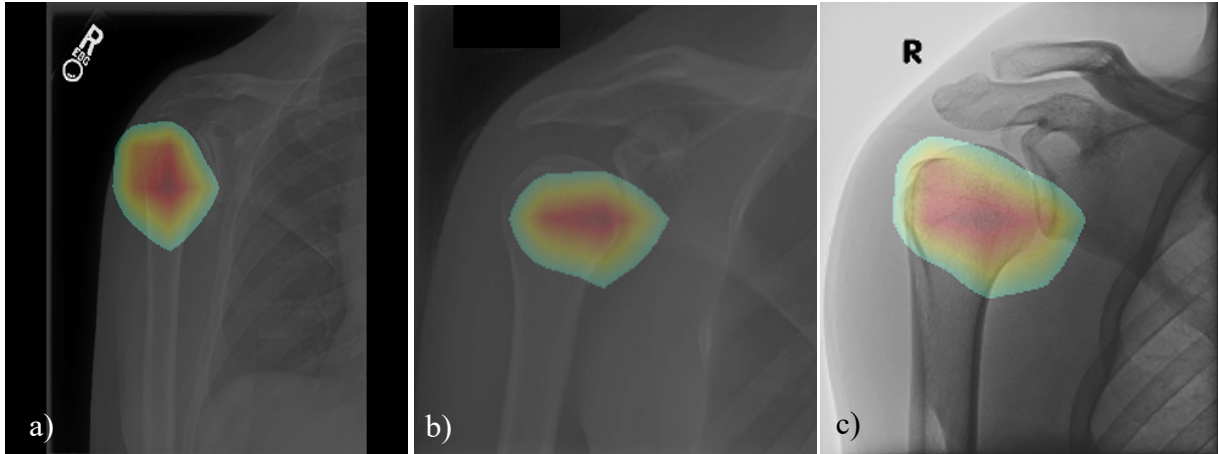


Figure 20. CAM « SUCCESS » - Incidences de l'épaule de face. a) Rotation neutre, b) Rotation interne, c) Rotation externe.

En observant ces images, nous remarquons que l'IA a appris d'elle-même à reconnaître l'importance de cette région pour distinguer ces 3 incidences. Plus encore, nous remarquons que la gerbe de scintillation est plus intense sur les régions où se projettent les structures d'intérêt (tubercules et coulisse bicipitale).

Cependant, il arrive que le système prête attention à d'autres régions, faussant ainsi ces prédictions. Pour illustrer cela, nous avons sélectionné quelques images de CAM montrant les erreurs commises par le système (Fig. 21).

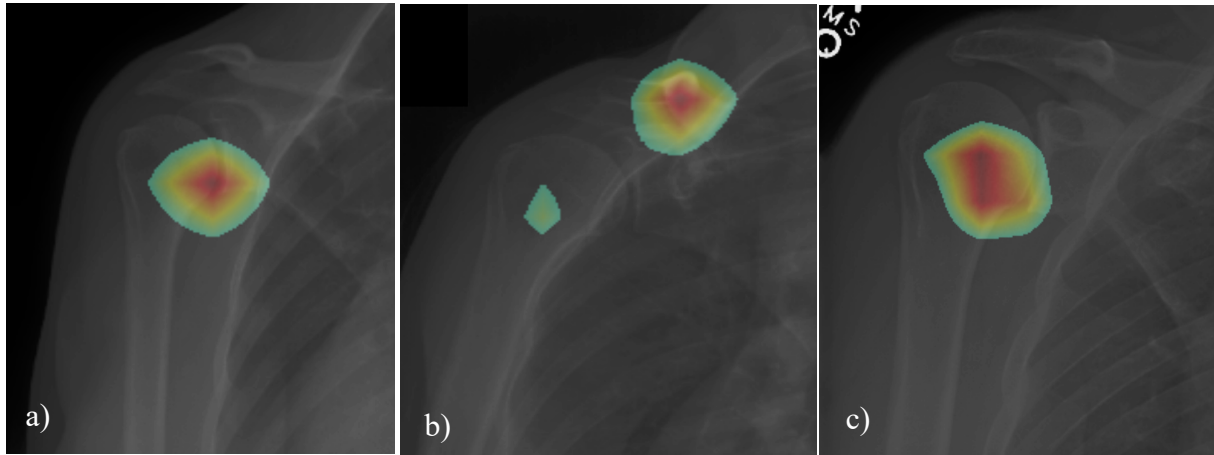


Figure 21. CAM « FAILURE » - Incidences épaule de face.

Ces images illustrent bien le fait que, le plus souvent, lorsque le système ne regarde pas au bon endroit, il se trompe.

Cependant, il arrive que l'IA nous corrige lorsque des images mal taguées apparaissent. En effet, sur le grand échantillon d'images taguées préalablement, il n'est pas rare que des erreurs surgissent. Certains exemples nous ont été révélés en analysant les CAM (Fig. 22).

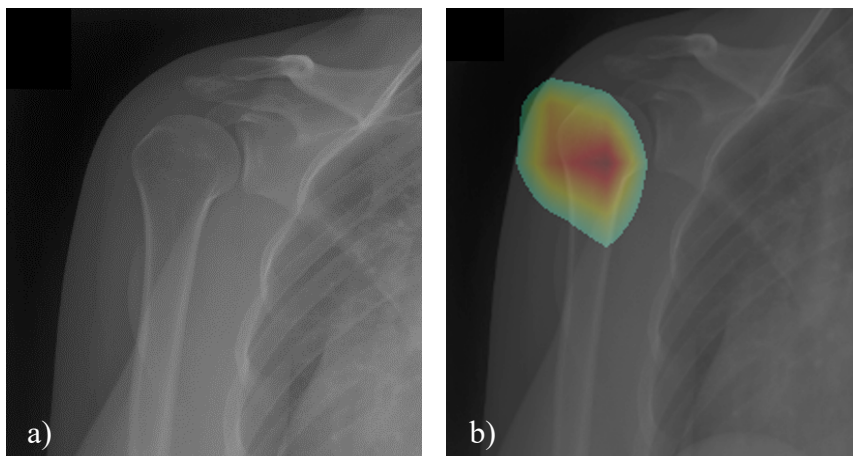


Figure 22. Exemple erreur de tagging. a) Tag : Rotation interne (faux), b) Tag : Rotation neutre (correct).

5.2. Étude 2 - Travail pratique (TP)

5.2.1. Aide avec fiche de rappel (Phase 2 et 3)

Nous l'avons vu, l'IA est capable de réaliser un travail colossal en un temps record. Toutefois, l'exactitude de ses prédictions n'est pas toujours à la hauteur. Malgré une quantité de données importante et une stratégie d'entraînement optimale, les gains de performance sont minimes et peuvent vite plafonner.

Inversement, l'être humain est capable d'apprendre vite avec un minimum d'exemples et sur un court intervalle de temps. C'est ce que nous avons cherché à démontrer entre les phases 2 (Section 4.2.3) et 3 (Section 4.2.4). En effet, dans la phase 2, les étudiants étaient chargés de catégoriser des images radiologiques d'épaule selon leur incidence. Nous avons prédit que les rotations de l'humérus sur les clichés d'épaule de face seraient les plus difficiles à reconnaître. C'est pourquoi, dans la phase 3, les étudiants ont reçu une feuille de rappel sur les critères anatomiques de rotation de la tête humérale sur les clichés de l'épaule de face (annexe XII) avant de passer en revue les mêmes images que lors de la phase 2.

Ainsi, les résultats obtenus à l'issue de ces deux phases vont nous permettre de confirmer l'utilité des fiches de rappel sur la base du constat de l'évolution des performances par incidences (tableau 27).

| Incidences | Accuracy globale | |
|--------------------------|-----------------------------------|-----------------------------------|
| | Phase 2 (sans fiche de rappel) | Phase 3 (avec fiche de rappel) |
| Profils de coiffe (Neer) | 0.985 | 0.989 |
| Axiale | 0.993 | 0.992 |
| Face-HN | 0.722 | 0.754 |
| Face-HI | 0.786 | 0.842 |
| Face-HE | 0.763 | 0.814 |
| Total | 0.850 | 0.878 |

Tableau 27. Phase 2 et 3 : Comparaison des performances (Accuracy par classe et globale).

Au terme de la phase 3, nous pouvons voir que les performances globales des étudiants ont augmenté (+3,19 %). Les fiches portant principalement sur les rotations de la tête humérale, nous remarquons aussi une augmentation des performances sur les incidences de l'épaule de face en rotation neutre (+4,43%), rotation interne (+7,12 %) et rotation externe (+6,68 %). Ce constat prouve que l'utilisation d'une fiche de rappel peut améliorer significativement les performances des étudiants.

D'autre part, cela nous montre que même si nous faisons appel à des personnes possédant les connaissances requises pour traiter ce type d'images, le risque de commettre des erreurs n'est pas nul. De plus, cette expérience témoigne de l'importance de suivre une phase d'instruction pour améliorer l'étiquetage des données en vue d'entraîner un réseau de neurones dans le cadre d'un projet d'apprentissage supervisé. En effet, cette étape aura un impact crucial sur la fiabilité des prédictions du modèle qui en découle.

Pour pouvoir comparer les performances de l'IA avec celles des étudiants, nous avons comparé (tableau 28) les résultats obtenus par les étudiants lors de la phase 3 avec ceux obtenus par le meilleur modèle d'IA (modèle n°9 - Nasnet_a_large) sur les mêmes images (tableau 17 - Section 4.2.5).

| Incidences | Accuracy globale | |
|---------------------------------|------------------|--------------|
| | Étudiants | IA |
| Profils de coiffe (neer) | 0.989 | 0.987 |
| Axiale | 0.992 | 1.00 |
| Face-HN | 0.754 | 0.764 |
| Face-HI | 0.842 | 0.322 |
| Face-HE | 0.814 | 0.625 |
| Total | 0.873 | 0.745 |

Tableau 28. Phase 3 : Comparaison Étudiants vs IA (Accuracy par classe et globale).

Globalement, nous remarquons que les étudiants ont un taux de bonnes réponses supérieur à l'IA (+17,18%), ce qui réfute l'hypothèse selon laquelle l'IA est plus performante que les étudiants dans cette tâche. Toutefois, notons que pour certaines incidences, l'IA semble légèrement meilleure que les étudiants, comme c'est le cas pour les axiales (+0,8%) et les épaules de face en rotation neutre (+1,32%). De même pour les profils de coiffe (Neer) qui montrent des performances quasi-identiques ($\Delta 0,2\%$). Ainsi, cette comparaison nous montre que, pour les incidences facilement différenciables, l'IA est en mesure d'égaliser voire dépasser les performances humaines.

Cependant, pour les incidences plus subtiles, comme pour les degrés de rotation des clichés d'épaule de face, nous remarquons un plus grand écart de performance. C'est notamment le cas pour les rotations internes où l'IA démontre des performances médiocres par rapport aux étudiants ($\Delta 161,49\%$). Tandis que pour les rotations externes, cette différence est beaucoup moins grande ($\Delta 30,24\%$).

Il est intéressant de porter attention à l'évolution des performances du modèle n°9 entre l'étude 1 (tableau 8 – Section 4.1.3) et l'étude 2 (tableau 17 – Section 4.2.5). En effet, nous pouvons remarquer que le fait de modifier les caractéristiques (données, quantités, etc.) du dataset d'évaluation (TEST) révèle des performances significativement différentes (0.842 vs 0.745).

Cela peut s'expliquer par les défauts intrinsèques du dataset MURA dont les classes ne sont pas équilibrées en termes de quantité (graphique 4 – Section 3.4.3). Ainsi, les erreurs commises sur les classes possédant moins de 1000 images (ex. épaule de face en rotation interne et externe) seront plus pénalisantes au niveau de performances mesurées. Or, pour permettre sa comparaison avec les étudiants, nous avons redéfini un nouveau dataset (graphique 5 – Section 3.5.2) au sein duquel les classes sont pratiquement équilibrées. De plus, 2 classes ont volontairement été retirées (acromio-claviculaire et trans-thoracique), ce qui rend la comparaison des performances globales difficile.

En conséquence, la sélection arbitraire des données, dont la qualité n'a pas été vérifiée pour réaliser ce test, a pu biaiser l'évaluation des performances du modèle et majorer ses faiblesses pour les incidences dont les performances n'étaient déjà pas excellentes, voire inverser la tendance pour d'autres (tableau 29).

| Accuracy par classe | | | | | | | Accuracy globale |
|---------------------|--------|---------|---------|---------|-------|------------------|------------------|
| Acro-clav | Axiale | Face HE | Face HI | Face HN | Neer | Trans-thoracique | |
| Étude 1 | | | | | | | |
| 0.777 | 0.995 | 0.594 | 0.713 | 0.729 | 0.955 | 0.500 | 0.842 |
| Étude 2 | | | | | | | |
| X | 1.00 | 0.625 | 0.322 | 0.764 | 0.987 | X | 0.745 |

Tableau 29. Étude 1 et 2 : Comparaison des performances du modèle n°9 sur les 2 dataset d'évaluation (Accuracy).

Toutefois, si nous faisons abstraction des limitations imposées par cette expérience, nous remarquons que, même en ayant recours à l'ensemble des données d'évaluation, les performances demeurent inférieures à celles obtenues par les étudiants, et ce pour chaque classe existante.

Cependant, si nous prenons en considération le facteur « temps » précédemment explicité dans la comparaison entre les phases 1 et 2 du TP, l'IA présente un net avantage sur le plan de la vitesse de traitement des données par rapport à l'être humain. En effet, le point fort d'un ordinateur réside dans la vitesse d'exécution d'un grand nombre d'opérations dans un intervalle de temps très court.

5.2.2. Pertinence pour la formation

Au terme de ce travail, il s'est avéré que la plupart des étudiants étaient d'accord sur le fait que la classification de radiographies par incidences semble plus pertinente pour notre profession au vu des difficultés révélées entre les phases 1 et 2.

Par ailleurs, l'introduction d'un rappel entre les phases 2 et 3 a engendré une nette amélioration des performances témoignant d'une meilleure fiabilité des étudiants par rapport à l'IA, appuyant notre hypothèse initiale sur les gains apportés par l'utilisation d'un support didactique. Ainsi, les phases 2 et 3 ont été jugées plus pertinentes d'un point de vue pédagogique même si, d'un point de vue clinique, la détection pathologique semble avoir beaucoup plus d'applications.

Néanmoins, il en ressort qu'un système de reconnaissance d'incidences radiologiques d'épaule semble plus utile dans la mesure où elle pourrait fournir une aide à l'identification des incidences plus complexes, comme c'est le cas des différents degrés de rotation de l'épaule de face.

Cependant, au vu des variabilités observées parmi les réponses des étudiants, il apparaît que l'utilisation d'une méthode de tagging indirect ne saurait pas entièrement profitable dans le cadre d'un projet de labellisation d'images radiologiques à grande échelle. Et pour causes, l'être humain est par nature susceptible de commettre de nombreuses erreurs (mauvais clics, fatigue, incompréhension, manque de connaissances, etc.). Parmi les solutions, nous pourrions envisager de réaliser cette tâche à plusieurs, de mettre à disposition des supports didactiques ou d'attribuer un indice de confiance à l'opérateur en lui soumettant un échantillon d'images pré-taguées pour vérifier la pertinence de ces réponses avant de lui en soumettre de nouvelles. De plus, l'ergonomie de l'outil de tagging aura son importance pour limiter les pertes de temps liées à sa prise en main.

Ainsi, le format de ce TP a permis d'introduire plusieurs concepts permettant l'élaboration de ces systèmes. De plus, cela aura permis de modifier, voire renforcer, certaines croyances sur l'IA, ces avantages (rapidité) et ces inconvénients (fiabilité).

Après avoir révélé aux étudiants l'importance de ce travail pour notre étude et sa pertinence pour l'avenir de notre formation, ces derniers ont témoigné en faveur du format d'enseignement

conçu pour l'occasion. Ils jugent cette méthode ludique, interactive et plus intéressante pour introduire les concepts gravitant autour du thème de l'IA. De plus, leur sollicitation indirecte dans le cadre d'une tâche de labellisation n'a pas suscité de commentaires négatifs dans l'optique où cela est bénéfique à la recherche. Même si certains estiment que l'honnêteté améliore la motivation et l'assiduité des participants.

Bien que ce procédé semble profiter davantage aux « Data scientist » chargés de traiter des milliers de données en vue de modéliser des solutions informatiques dotées d'intelligence, il semble important de signaler que, dans le cadre de notre étude, la valorisation des données dépend grandement du savoir-faire des personnes compétentes dans le domaine concerné. En conséquence, nous sommes tenus de comprendre les fondamentaux qui nous permettront de nous adapter, voire participer, à l'émergence de ces technologies. En effet, cette étude a également permis de démontrer un réel avantage du tagging indirect en routine clinique, notamment sur la nécessité absolue de recourir à des personnes bénéficiant des connaissances adaptées au traitement préalable des données dont la qualité est déterminante pour le développement de tels systèmes.

5.3. Synthèse

Tout au long de ce travail, nous avons tenté d'expliquer le fonctionnement de l'IA et d'expliquer les différentes étapes de l'entraînement d'un réseau de neurones artificiels en vue de réaliser une tâche de classification de radiographies du membre supérieur en différentes classes (régions anatomiques et incidences d'épaule). Nous avons également comparé différentes approches stratégiques afin d'estimer quels étaient les paramètres les plus importants pour réaliser cette tâche.

Afin de permettre une meilleure interprétation des résultats, nous avons également fourni une appréciation quantitative et qualitative de nos données. Nous avons ainsi pu mettre en évidence certains facteurs limitants.

Au vu des résultats de l'étude 1, nous avons identifié les points forts et les points faibles de chacune des architectures réseau. Le réseau Nasnet_a_large a davantage suscité notre intérêt en démontrant des performances globalement meilleures que le réseau Inception V4 en fonction des choix stratégiques appliqués. Le réseau Inception V4 a toutefois démontré de bonnes performances malgré une stratégie moins riche en nombre de paramètres.

De plus, nous avons isolé l'impact de ces différentes stratégies indépendamment de l'architecture sélectionnée et remarqué leurs avantages et inconvénients en tenant compte des facteurs limitants, et ce pour chacune des deux tâches de classification réalisées. En fonction de quoi, nous avons remarqué que la quantité d'images est un facteur important, notamment si certaines classes minoritaires apparaissent.

L'étude 2 nous a permis d'analyser les performances des étudiants TRM sur différentes tâches de classification d'images radiologiques en vue de les comparer avec les résultats obtenus par notre modèle d'IA. Nous avons remarqué qu'aucun modèle ne semble égaler les étudiants, même si ses performances restent convaincantes pour les incidences les plus faciles.

Toutefois, la comparaison des phases 1 et 2 du TP nous montre que les étudiants semblent proportionnellement démontrer les mêmes difficultés que l'IA en fonction de la complexité de la tâche (régions anatomiques vs incidences d'épaule) et de certaines classes jugées plus complexes (degré de rotation d'épaule).

5.4. Potentiel, limites et perspectives

Afin d'initier la réalisation de ce travail, une grande quantité d'images a dû être préalablement passée en revue afin d'identifier pour chacune d'elle la classe qui s'y réfère. Nous avons très vite constaté que pour cette étude, les données MURA représenteraient un enjeu majeur, tant au niveau de la quantité de données à traiter qu'en termes de qualité d'image. En effet, ce dataset a initialement été conçu pour entraîner un réseau de neurones à détecter la présence de pathologie (fracture et/ou prothèses). De ce fait, près de la moitié des images (toute classe confondue) comporte des éléments artéfactuels nuisibles dans le cadre de cette étude. Ces images rendent la tâche de labellisation fastidieuse dans la mesure où certaines radiographies ont malgré tout été labellisées alors qu'objectivement leur appartenance faisait débat, sans compter les images malencontreusement labellisées à tort.

Ce phénomène a été rendu saillant lors de la réalisation du TP (étude 2), car sur l'échantillon d'images taguées par les étudiants, nous avons constaté une grande variabilité au sein des réponses et des performances récoltées. Ainsi, les performances réalisées par les deux parties (IA vs étudiants) ne reflètent pas nécessairement les performances réelles qui auraient pu être obtenues si les données d'évaluation ne présentaient aucun défaut. Les réponses attendues étant directement liées à l'expertise du tagging réalisé préalablement par nos soins, il aurait été intéressant de réitérer l'expérience sur cet échantillon de données pour constater si, oui ou non, nous reproduisons les mêmes réponses plusieurs mois après. Le cas échéant, cela aurait révélé un biais remettant en question la pertinence de certains résultats. Néanmoins, ce choix arbitraire, qui privilégie la quantité à la qualité, a été justifié par la nécessité de fournir un nombre conséquent d'images à notre réseau de neurones afin qu'il puisse bénéficier d'un entraînement acceptable.

Nous pensons que dans le cadre de futures études mobilisant ce dataset, il sera important de tenir compte de ces limitations afin d'effectuer des choix stratégiques adaptés à l'entraînement d'un modèle de reconnaissance d'images radiologiques.

Malgré tout, cette étude a démontré les capacités des réseaux de neurones à reconnaître des images radiologiques malgré l'occurrence limitée de certaines classes par rapport à d'autres. Afin de surmonter ces limitations, plusieurs modèles ont été comparés, démontrant les forces et faiblesses des stratégies appliquées. Concernant les architectures réseau, le nombre de tests

comparés demeure assez faible pour dégager la tendance d'un modèle à surpasser l'autre. En tenant compte de la qualité et de la quantité de données utilisées pour entraîner et tester ces modèles, il est possible qu'ils aient été sensiblement impactés par la répartition des données d'entraînement, de validation et d'évaluation. En conséquence, nous avons pu observer que certaines stratégies n'ont pas eu les effets bénéfiques escomptés. D'où l'importance d'avoir recours à certaines méthodes pour contourner certaines limitations propres aux données (ex. Cross Validation).

De plus, il apparaît que la disparité des classes (régions anatomiques) entre les jeux de données HEDS et MURA, ait pu biaiser les résultats obtenus pour les tests réalisés sur ces deux dataset, rendant leur comparaison difficile, notamment pour les régions anatomiques. De plus, contrairement à la base de données MURA, les données HEDS n'étaient labellisées que selon le tag de 2^{ème} degré (régions anatomiques), rendant impossible l'évaluation des performances selon le tag de 3^{ème} degré (incidences d'épaule). Il serait donc judicieux d'égaliser le nombre de classes afin d'isoler les variations de performances sur des images issues de deux datasets de provenance et de qualité différentes en fonction de la tâche de classification.

Tenons également compte du fait que pour compléter notre analyse, nous aurions pu proposer de nouvelles stratégies d'entraînement (tableau 30). Toutefois, étant donné le nombre déjà conséquent de données obtenues au travers des différents tests réalisés, nous nous sommes limités à exploiter ces seules données.

| Architecture | Stratégie | | | | |
|---|-------------------|---------------------|---------------|---------|-------------|
| | Data augmentation | Balancing | No validation | k-folds | Paramètres |
| Régions anatomiques | | | | | |
| Inception V4 | X | X | Full | X | bs 18 / e20 |
| Incidences de l'épaule | | | | | |
| NASNet_a_large | Par défaut | Sur-échantillonnage | X | 5 | bs 18 / e20 |
| NASNet_a_large | Par défaut | Sur-échantillonnage | Full | 5 | bs 18 / e20 |
| Régions anatomiques & Incidences de l'épaule | | | | | |
| NASNet_a_large | X | Sur-échantillonnage | Full | 5 | bs 18 / e20 |
| NASNet_a_large | Xray | Sur-échantillonnage | Full | 5 | bs 18 / e20 |
| NASNet_a_large | Xray | Sur-échantillonnage | X | 5 | bs 18 / e20 |
| Inception V4 | Xray | X | X | X | bs 18 / e20 |
| Inception V4 | X | X | X | X | bs 18 / e20 |

Tableau 30 – Proposition de nouvelles stratégies d'IA.

Par ailleurs, le travail pratique (étude 2) nous a permis de récolter de précieuses données sur la fiabilité des étudiants de 3^{ème} à reproduire une tâche de labellisation d'images en différentes catégories (tags). De plus, cela nous a permis de tester un nouveau format d'enseignement dont la pertinence a pu être évaluée en fonction des commentaires récoltés à chaud pendant et après le TP. Sur ce dernier point, il aurait été souhaitable de réaliser un rapide questionnaire anonyme afin d'encourager les participants les plus discrets à partager leur avis sur les thématiques abordées durant cette matinée.

6. Conclusion

Ce travail s'est intéressé au domaine de l'IA pour la reconnaissance d'images radiologiques, dans le cadre de la formation pour les TRM. Tout d'abord, nous avons cherché à rendre compréhensibles certains concepts théoriques liés au fonctionnement de l'IA et à l'entraînement d'un réseau de neurones en vue de réaliser une tâche de classification d'images. En nous basant sur des études récentes, nous avons fourni des exemples d'applications déjà existantes dans le domaine de l'imagerie médicale et des perspectives d'avenir pour son implication dans l'enseignement.

Au travers des deux études réalisées, nous avons cherché à démontrer les performances de l'IA sur une tâche de classification d'images radiologiques en différentes classes de complexité variables (régions anatomiques et incidences) et comparé les résultats obtenus avec ceux récoltés auprès des étudiants dans le cadre d'un travail pratique conçu à cet effet. Nous avons constaté que, dans certains cas, les performances de l'IA se rapprochent de celles de l'humain et évalué la pertinence d'une potentielle implication des TRM, professionnels ou étudiants, dans la valorisation des données destinées au développement de l'IA en radiologie.

À l'heure actuelle, aucun système de reconnaissance d'incidences radiologiques ne semble exister. De plus, contrairement aux radiologues, les TRM sont très peu mentionnés dans la littérature sur le sujet. En ce sens, ce travail constitue une étape préliminaire de sensibilisation des étudiants aux thématiques de l'IA en imagerie médicale et à la valorisation de leurs connaissances dans le développement et l'amélioration de ces systèmes.

Cela nous laisse penser qu'à l'avenir, les TRM pourraient faire partie intégrante de projets de recherche visant à développer des applications cliniques spécifiques aux problématiques du métier.

7. Bibliographie

- Allibhai, E.. (2018). Hold-out vs. Cross-validation in Machine Learning. Accès <https://medium.com/@ejaz/holdout-vs-cross-validation-in-machine-learning-7637112d3f8f>
- Al-Musibli, A. & Cagdas O.. (2019). Intelligence artificielle en imagerie ostéo-articulaire : le rôle du TRM ? (Travail de Master). Haute École de Santé, Genève.
- Amazon SageMake : Manuel du développeur. (2020). Accès https://docs.aws.amazon.com/fr_fr/sagemaker/latest/dg/sagemaker-dg.pdf#IC-HowItWorks
- Audevert, A.. (2017). Machine learning de A à Z. Salon de la data, 19 septembre 2017, Nantes.
- Audevert, A.. (2018). Voyage dans le monde du Deep learning. Salon de la data, 12 septembre 2018, Nantes.
- Bastin, G. & Tubaro. P.. (2018). Le moment big data des sciences sociales. Revue française de sociologie, Vol. 59(3), 375-394. Accès <https://www.cairn.info/revue-francaise-de-sociologie-2018-3-page-375.htm>
- Baya-Laffite, N., Beaudé, B., & Garrigues, J. (2018). Le deep learning au service de la prédiction de l'orientation sexuelle dans l'espace public. *Rezeaux*, n° 211(5), 137-172. Accès
- Benhamou, S. & Janin, L.. (2018). INTELLIGENCE ARTIFICIELLE ET TRAVAIL. Accès <http://pinguet.free.fr/fstraia318.pdf>
- Béranger, J. (2016). La valeur éthique des big data en santé. *Les Cahiers du numérique*, Vol. 12(1), 109-132. Accès <https://www.cairn.info/revue-les-cahiers-du-numerique-2016-1-page-109.htm>
- Bertrand, D., Fluss, J., Billard. C. & Ziegler, J. C.. (2010). Efficacité, sensibilité, spécificité : comparaison de différents tests de lecture. pp. 299-320. Accès <https://www.cairn.info/revue-l-annee-psychologique1-2010-2-page-299.htm>
- Biernat, E., Lutz, M.. (2015). Data Science : Fondamentaux et études de cas - Machine learning avec Python et R. Accès <https://static.fnac-static.com/multimedia/editorial/pdf/9782212142433.pdf>
- Bousquet, M., Pin, R., Sivault, C.. (2019). Le guide complet de l'intelligence artificielle : intelligence humaine, animale et végétale. Editions du sens
- Canon. (2020). Advanced intelligent Clear-IQ Engine (AiCE) - Sharp, clear and distinct images At low dose. Accès <https://us.medical.canon/products/computed-tomography/aice/>

- Cardon, D., Cointet, J. P. & Mazières, A.. (2018). La revanche des neurones. L'invention des machines inductives et la controverse de l'intelligence artificielle. (p. 173-220). Accès <https://www.cairn.info/revue-reseaux-2018-5-page-173.htm>
- Carpentier, M. (2019). Classification fine par réseau de neurones à convolution. Mémoire. Accès <https://corpus.ulaval.ca/jspui/bitstream/20.500.11794/35835/1/35356.pdf>
- COMPUTING MACHINERY AND INTELLIGENCE. A. Turing. (1950). Accès http://www.espace-turing.fr/IMG/pdf/Computing_Machinery_and_Intelligence_A-M-_Turing.pdf
- Degila, K. (2019). La validation croisée en Machine Learning : Tout ce qu'il faut savoir. Accès <https://www.kevindegila.com/la-validation-croisee-en-machine-learning-tout-ce-quil-faut-savoir/>
- Deng, F., Wang, D. J. & al.. (s.d.). Computer aided diagnosis. Accès <https://radiopaedia.org/articles/computer-aided-diagnosis-1>
- Département fédéral de l'économie, de la formation et de la recherche (DEFR) - Secrétariat d'Etat à la formation, à la recherche et à l'innovation (SEFRI). (2019).
- Do, S., Song, K. D., & Chung, J. W. (2020). Basics of Deep Learning : A Radiologist's Guide to Understanding Published Radiology Articles on Deep Learning. Korean Journal of Radiology, 21(1), 33-41. Accès <https://doi.org/10.3348/kjr.2019.0312>
- Dowek. G.. (2011). Les quatre concepts de l'informatique. Sciences et technologies de l'information et de la communication en milieu éducatif : Analyse de pratiques et enjeux didactiques. Accès <https://edutice.archives-ouvertes.fr/edutice-00676169/document>
- Dubois, A. & al.. (2019). ATAWAO CONSULTING. Intelligence artificielle - État de l'art et perspectives pour la France - Rapport final. Accès https://www.entreprises.gouv.fr/files/files/directions_services/etudes-et-statistiques/prospective/Intelligence_artificielle/2019-02-intelligence-artificielle-etat-de-l-art-et-perspectives.pdf
- Dutta, A. & Zisserman, A.. (2019). Visual Geomtry Group (VGG). Department of Engineering Science. University of Oxford. The VGG Image Annotator (VIA). Accès <https://www.groundai.com/project/the-vgg-image-annotator-via/1>
- Fitaire, C. & al.. (2018). Comment l'intelligence artificielle va-t-elle bouleverser la médecine ?. Accès <https://www.revmed.ch/RMS/2018/RMS-N-629/Comment-l-intelligence-artificielle-va-t-elle-bouleverser-la-medecine>
- Frija, G.. (2018). Imagerie : le Big Data est nécessaire pour faire progresser l'IA. [Consulté le 28.03.2020]. Accès <https://www.santeperso.ch/Opinions/Imagerie-le-Big-Data-est-necessaire-pour-faire-progresser-l-IA>
- Géron, A.. (2017). Machine Learning avec Scikit-Learn. Accès <https://www.dunod.com/sites/default/files/atoms/files/feuilleter1.pdf>

- Ghennam, B.. & Smara, S.. (2019). Les réseaux de neurone convolutionnel (CNN) pour la classification des images associées aux places de stationnement d'un parc de véhicule. Mémoire. Accès <http://193.194.80.11/jspui/bitstream/123456789/2997/1/Doc3.pdf>
- Gong, E., Pauly, J.M., Wintermark, M., Zaharchuk, G.. (2018). Deep learning enables reduced gadolinium dose for contrast-enhanced brain MRI. doi : 10.1002/jmri.25970
- Gooßen, A., Deshpande, H., Harder, T., Schwab, E., Baltruschat, I., Mabotuwana, T., Cross, N., & Saalbach, A. (2019). Deep Learning for Pneumothorax Detection and Localization in Chest Radiographs. arXiv:1907.07324 [cs, eess]. Accès <http://arxiv.org/abs/1907.07324>
- Gottfredson, L. S.. (1997). Mainstream science on intelligence : An editorial with 52 signatories, history, and bibliography. *Intelligence*, 24, 25–52. doi:10.1016/S0160-2896(97)90011-8
- Grenier, P., & Neuenschwander, S. (2006). Que doit-on savoir sur les « CAD ». *Journal de Radiologie*, 87(10), 1215. [https://doi.org/10.1016/S0221-0363\(06\)86762-4](https://doi.org/10.1016/S0221-0363(06)86762-4)
- Gugerty, L.. (2006). Newell and Simon's Logic Theorist : Historical Background and Impact on Cognitive Modeling. Accès <https://journals.sagepub.com/doi/abs/10.1177/154193120605000904>
- Gupta, P.. (2017). Cross-Validation in Machine Learning. Accès <https://towardsdatascience.com/cross-validation-in-machine-learning-72924a69872f>
- Heverhagen, J. T., Ebner, L., Mougiakakou, S. G., Dahlweid, F.-M., & Christe, A. (2019). Radiologie : Intelligence artificielle : la radiologie au fil du temps. *Forum Médical Suisse*, 19(0304), 56-58. <https://doi.org/10.4414/fms.2019.08035>
- Holzberger, K. & Kim, W.. (2018). L'intelligence artificielle accessible à tous : la promesse d'une puissante technologie pour le secteur de la radiologie. Accès https://www.nuance.com/content/dam/nuance/fr_fr/collateral/healthcare/white-paper/wp-nuance-sante-livreblan-lintelligence-artificielle-radiologie.pdf
- Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., ... & Wang, Y. (2017). Artificial intelligence in healthcare : past, present and future. *Stroke and vascular neurology*, 2(4), 230-243. Accès <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5829945/>
- Kaenampornpan, N. & al.. (2018). Multi-disciplinary Trends in Artificial Intelligence. Accès https://books.google.fr/books?id=YON3DwAAQBAJ&pg=PA115&lpg=PA115&dq=Nasnet+a+large+google+brain&source=bl&ots=GC9rExZY-x&sig=ACfU3U1Eu7U-pwQQm8kIU13N4oyAkdFiQg&hl=fr&sa=X&ved=2ahUKEwjvqZTNkN7pAhUIKBoKHf_zBTUQ6AEwA3oECAoQAQ#v=onepage&q=Nasnet%20a%20large%20google%20brain&f=false
- Karim, R.. (2019). Illustrated : 10 CNN Architectures : A compiled visualisation of the common Convolutional neural networks. Accès <https://towardsdatascience.com/illustrated-10-cnn-architectures-95d78ace614d#81e0>

- Kaur Gill, J.. (2018). Automatic Log Analysis using Deep Learning and AI. Accès <https://www.xenonstack.com/blog/log-analytics-deep-machine-learning/>
- Kushwaha, A.. (2019). Solving Class Imbalance problem in CNN. Accès <https://medium.com/x8-the-ai-community/solving-class-imbalance-problem-in-cnn-9c7a5231c478>
- Larlus, D., Dorkó G. & Jurie, F.. (2006). Création de vocabulaires visuels efficaces pour la catégorisation d'images. Accès https://hal.archives-ouvertes.fr/file/index/docid/203624/filename/F._JURIE.pdf
- LeCun, Y.. (2016). Les Enjeux de la Recherche en Intelligence Artificielle. 7. Accès https://dataanalyticspost.com/wp-content/uploads/2017/04/ylecun_college_France.pdf
- LeCun, Y.. (2018). La prochaine révolution de l'intelligence artificielle. Unexpected Source of Inspiration (USI), 26 juin 2018, Paris.
- Leonel, J.. (2019). Hyperparameters in Machine/Deep Learning. Accès <https://medium.com/@jorgesleonel/hyperparameters-in-machine-deep-learning-ca69ad10b981>
- Lovis, P. C., Gaudet-Blavignac, C., Chevrier, R., Robert, A., & Issom, D. (2018). BigData, intelligence artificielle, blockchain : Guide pratique. REVUE MÉDICALE SUISSE, 5. Accès <https://www.revmed.ch/RMS/2018/RMS-N-617/BigData-intelligence-artificielle-blockchain-guide-pratique>
- Memmi, D.. (1990). Connexionnisme, Intelligence Artificielle et Modélisation cognitive. Accès https://www.persee.fr/doc/intel_0769-4113_1990_num_9_1_880
- Microsoft. (2020). Optimiser les hyperparamètres de votre modèle avec Azure Machine Learning. Accès <https://docs.microsoft.com/fr-fr/azure/machine-learning/how-to-tune-hyperparameters>
- Moujahid, A.. (2016). A Practical Introduction to Deep Learning with Caffe and Python. Accès <http://adilmoujahid.com/posts/2016/06/introduction-deep-learning-python-caffe/>
- Nabli, L. (2010). Cours Système Expert. Accès https://www.researchgate.net/publication/322740806_Cours_Systeme_Expert
- Open Data Science Conference (ODSC). (2019). Visualizing Your Convolutional Neural Network Predictions With Saliency Maps. Accès <https://medium.com/@ODSC/visualizing-your-convolutional-neural-network-predictions-with-saliency-maps-9604eb03d766>
- Oufella, Y.. (2008). ÉVOLUTION DU CONCEPT DE FRONT ROC ET COMBINAISON DE CLASSIFIEUR. Accès http://clement.chatelain.free.fr/docs/STAGE_M2_OUFELLA.pdf
- Parizeau, M.. (2004). Réseau de neurones. GIF-21140 et GIF-64326. Accès http://frostiebek.free.fr/docs/Reseaux%20de%20neurones/RNF_fzzy.pdf

- Radhakrishnan, P.. (2017). What are Hyperparameters ? and How to tune the Hyperparameters in a Deep Neural Network ?. Accès <https://towardsdatascience.com/what-are-hyperparameters-and-how-to-tune-the-hyperparameters-in-a-deep-neural-network-d0604917584a>
- Raj, B.. (2018). A Simple Guide to the Versions of the Inception Network. Accès <https://towardsdatascience.com/a-simple-guide-to-the-versions-of-the-inception-network-7fc52b863202>
- Ramesh, A. N. & al.. (2004). Artificial intelligence in medicine. Accès www.ncbi.nlm.nih.gov/pmc/articles/PMC1964229/pdf/15333167.pdf
- Rocheleau, J.. (2009) L'apport des théories de l'apprentissage cognitivistes. Accès https://oraprdnt.uqtr.quebec.ca/pls/public/docs/GSC332/F766183874_Approche_cognitiviste_apprentissage2009_10_05.pdf
- Romdhane, N. B., Taouil, K., Boudaya, S., Turki, H., & Bouhleb, M. S. (2007, March). Selection des variables et classification par reseaux de neurones des lesions dermatologiques. In 4th International Conference: Sciences of Electronic, Technologies of Information and Telecommunications, Tunisia. Accès <http://www.setit.rnu.tn/CDs%20SETIT/SETIT%202007/IV/18.pdf>
- Ruizendaal, R.. (2017). Deep Learning #3 : More on CNNs & Handling Overfitting. Accès <https://towardsdatascience.com/deep-learning-3-more-on-cnns-handling-overfitting-2bd5d99abe5d>
- Saha, S.. (2018). A Comprehensive Guide to Convolutional Neural Networks - the ELI5 way. Accès <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>
- Santini G. & al.. (2020). Segmentation automatique des métastases hépatiques en imagerie TEP/TDM basée sur l'apprentissage profond dans le cadre du cancer du sein métastatique. EM-Consulte. Accès <https://www.em-consulte.com/article/1354184>
- Santos, M. K., Ferreira Júnior, J. R., Wada, D. T., Tenório, A. P. M., Barbosa, M. H. N., & Marques, P. M. de A.. (2019). Artificial intelligence, machine learning, computer-aided diagnosis, and radiomics : Advances in imaging towards to precision medicine. Radiologia Brasileira, 52(6), 387-396. Accès <https://doi.org/10.1590/0100-3984.2019.0049>
- Sarkar. D. J.. (2018). A Comprehensive Hands-on Guide to Transfer Learning with Real-World Applications in Deep Learning. Accès <https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning-212bf3b2f27a>
- Scikit-learn 0.23.1 documentation Receiver Operating Characteristic (ROC). (s. d.). Accès https://scikit-learn.org/stable/auto_examples/model_selection/plot_roc.html#area-under-roc-for-the-multiclass-problem

- Seif, G.. (2018). Everything you need to know about AutoML and Neural Architecture Search. Accès <https://towardsdatascience.com/everything-you-need-to-know-about-automl-and-neural-architecture-search-8db1863682bf>
- Shah, T.. (2017). About Train, Validation and Test Sets in Machine Learning. Accès <https://towardsdatascience.com/train-validation-and-test-sets-72cb40cba9e7>
- Simon, F.. (2018). Deep Learning, les fonctions d'activation | SUPINFO, École Supérieure d'Informatique. Accès <https://www.supinfo.com/articles/single/7923-deep-learning-fonctions-activation>
- Song, W., Li, S., Liu, J., Qin, H., Zhang, B., Zhang, S., & Hao, A.. (2019). Multitask Cascade Convolution Neural Networks for Automatic Thyroid Nodule Detection and Recognition. *IEEE Journal of Biomedical and Health Informatics*, 23(3), 1215-1224. <https://doi.org/10.1109/JBHI.2018.2852718>
- Thiéry, O.. (2018). INTELLIGENCE ARTIFICIELLE (IA), ENJEUX ÉTHIQUES. Accès <https://espaces-numeriques.org/wp-content/uploads/2018/04/L111Sp29.pdf>
- Vittet, D.. (2009). Du cognitivisme aux neurosciences, au regard de la psychanalyse. Accès <https://www.cairn.info/revue-cahiers-jungiens-de-psychanalyse-2009-3-page-11.htm>
- Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A.. (2015). Learning Deep Features for Discriminative Localization. arXiv:1512.04150 [cs]. Accès <http://arxiv.org/abs/1512.04150>
- Zhou, X.. (2020). Automatic Segmentation of Multiple Organs on 3D CT Images by Using Deep Learning Approaches. doi : 10.1007/978-3-030-33128-3_9
- Zolynski, C. (2015). Big data : Pour une éthique des données. *I2D Information, données documents*, Volume 52(2), 25-26. Accès <https://www.cairn.info/revue-i2d-information-donnees-et-documents-2015-2-page-25.htm>
- Zoph, B. & al. (2018). Learning Transferable Architectures for Scalable Image Recognition. Accès http://openaccess.thecvf.com/content_cvpr_2018/papers/Zoph_Learning_Transferable_Architectures_CVPR_2018_paper.pdf

8. Annexes

I.Méthode d'évaluation des performances (Courbes ROC, macro, micro average)

Courbe « macro-average »

Tient compte des AUC de chaque classe indépendamment puis effectue une moyenne globale (en traitant donc toutes les classes de la même manière). Cette méthode est utilisée lorsque l'intérêt est porté sur les performances globales, sans tenir compte des classes minoritaires ou lorsque les classes sont équilibrées entre elles.

Courbe « micro-average »

Tient compte des AUC de l'ensemble de classes et de leur contribution puis effectue une moyenne globale (en traitant les classes différemment en fonction de leur occurrence dans le dataset d'évaluation). Cette méthode est utilisée lorsque nous suspectons un déséquilibre des classes au sein des données pouvant altérer les performances globales.

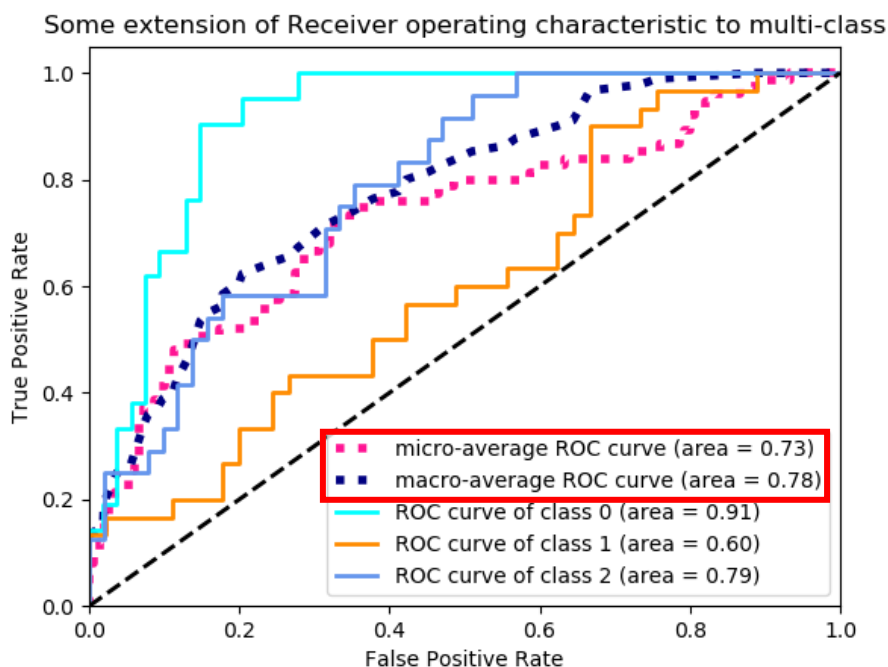


Figure 25. Scikit-learn 0.23.1 documentation Receiver Operating Characteristic (ROC). (s. d.). Exemple de courbe ROC multi-classes.

II.Étude 1 : Résumé des stratégies et performances par modèle

Régions anatomiques

Modèle n°1 : mura_anat_part_e20_bs18_Nasnet_a_large

Stratégie

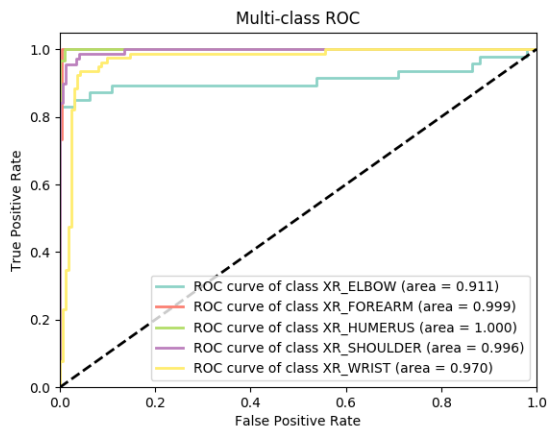
| Architecture | | Stratégie | | | | |
|--------------|----------------|-------------------|-----------|---------------|---------|-------------|
| | | Data augmentation | Balancing | No validation | k-folds | Paramètres |
| 1) | NASNet_a_large | Par défaut | X | X | X | bs 18 / e20 |

Performances (Accuracy) par classe et globales

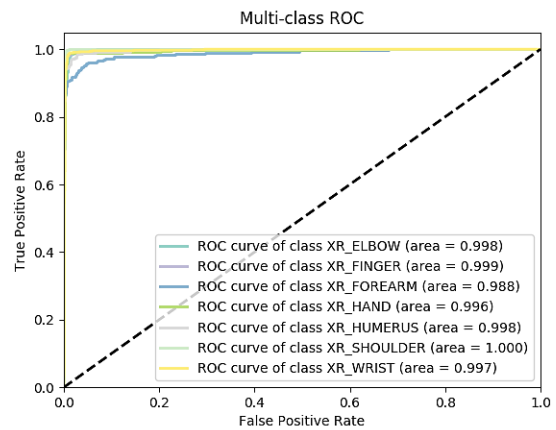
| N° | Accuracy par classe | | | | | | | Accuracy globale |
|----|---------------------|------------|---------|--------|---------|-------|--------|------------------|
| | Coude | Avant-bras | Humérus | Épaule | Poignet | Main | Doigts | |
| 1) | 0.808 | 1 | 1 | 0.956 | 0.884 | X | X | 0.912 |
| | 0.963 | 0.870 | 0.944 | 0.995 | 0.986 | 0.970 | 0.972 | 0.965 |

Courbes ROC

Test HEDS



Test MURA



Modèle n°2 : mura_anat_part_e20_bs18_inceptionv4

Stratégie

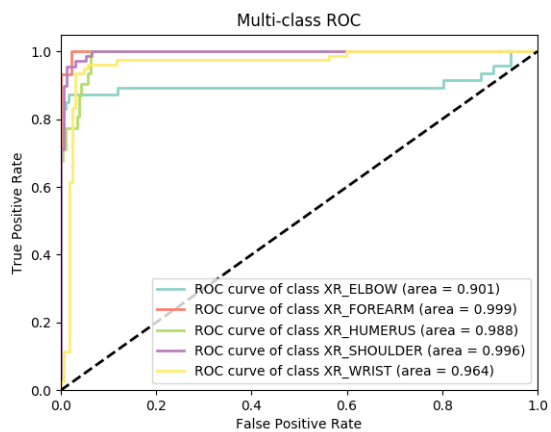
| Architecture | | Stratégie | | | | |
|--------------|--------------|-------------------|-----------|---------------|---------|-------------|
| | | Data augmentation | Balancing | No validation | k-folds | Paramètres |
| 2) | Inception V4 | Par défaut | X | X | X | bs 18 / e20 |

Performances (Accuracy) par classe et globales

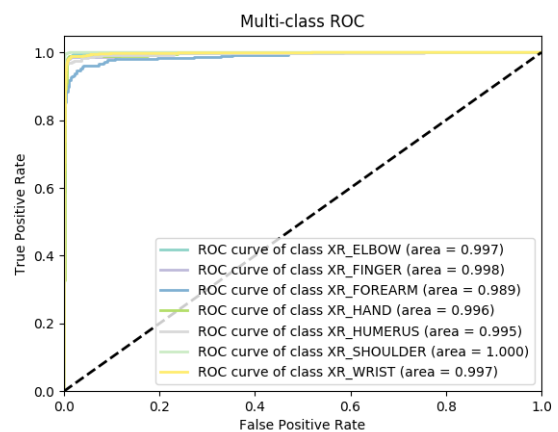
| N° | Accuracy par classe | | | | | | | Accuracy globale |
|----|---------------------|------------|---------|--------|---------|-------|--------|------------------|
| | Coude | Avant-bras | Humérus | Épaule | Poignet | Main | Doigts | |
| 2) | 0.872 | 1 | 0.774 | 0.971 | 0.833 | X | X | 0.883 |
| | 0.978 | 0.867 | 0.938 | 0.995 | 0.983 | 0.974 | 0.970 | 0.966 |

Courbes ROC

Test HEDS



Test MURA



Modèle n°3 : mura_anat_part_no_augment_e20_bs18_Nasnet_a_large

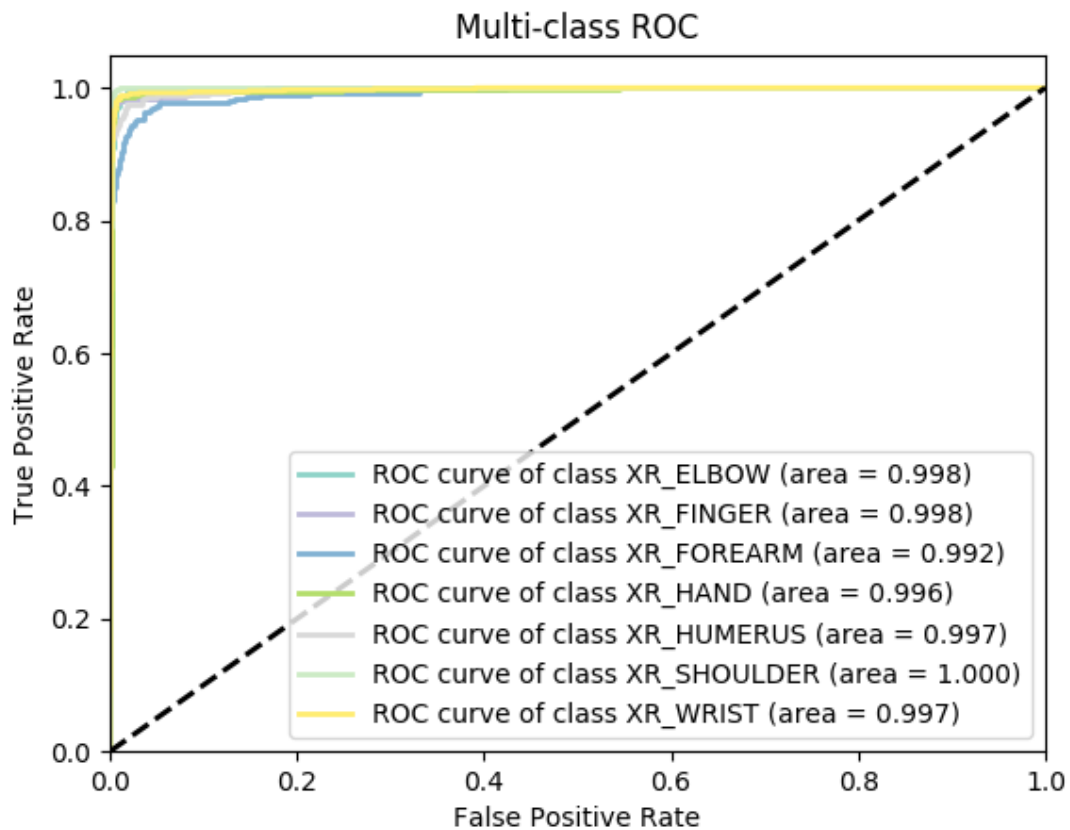
Stratégie

| Architecture | | Stratégie | | | | |
|--------------|----------------|-------------------|-----------|---------------|---------|-------------|
| | | Data augmentation | Balancing | No validation | k-folds | Paramètres |
| 3) | NASNet_a_large | X | X | X | X | bs 18 / e20 |

Performances (Accuracy) par classe et globales

| N° | Accuracy par classe | | | | | | | Accuracy globale |
|----|---------------------|------------|---------|--------|---------|-------|--------|------------------|
| | Coude | Avant-bras | Humérus | Épaule | Poignet | Main | Doigts | |
| 3) | 0.970 | 0.854 | 0.924 | 0.998 | 0.983 | 0.983 | 0.974 | 0.965 |

Courbes ROC



Incidences

Modèle n°4 : mura_shoulder_inc_balanced_e20_bs18_Nasnet_a_large

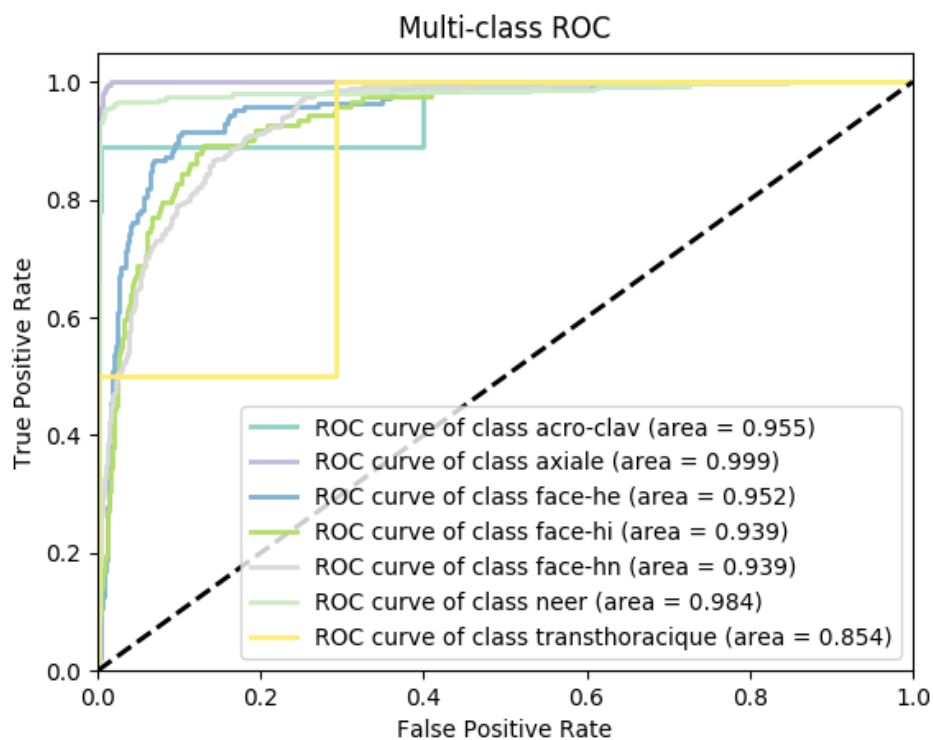
Stratégie

| | | Stratégie | | | | |
|-------------------------------|-------------------|------------|---------------------|---------|------------|-------------|
| Architecture | Data augmentation | Balancing | No validation | k-folds | Paramètres | |
| Incidences de l'épaule | | | | | | |
| 4) | NASNet_a_large | Par défaut | Sur-échantillonnage | X | X | bs 18 / e20 |

Performances (Accuracy) par classe et globales

| N° | Accuracy par classe | | | | | | | Accuracy globale |
|----|---------------------|--------|---------|---------|---------|-------|------------------|------------------|
| | Acro-clav | Axiale | Face HE | Face HI | Face HN | Neer | Trans-thoracique | |
| 4) | 0.778 | 1.00 | 0.734 | 0.795 | 0.639 | 0.952 | 0 | 0.838 |

Courbes ROC



Modèle n°5 : mura_shoulder_inc_e20_bs18_Nasnet_a_large

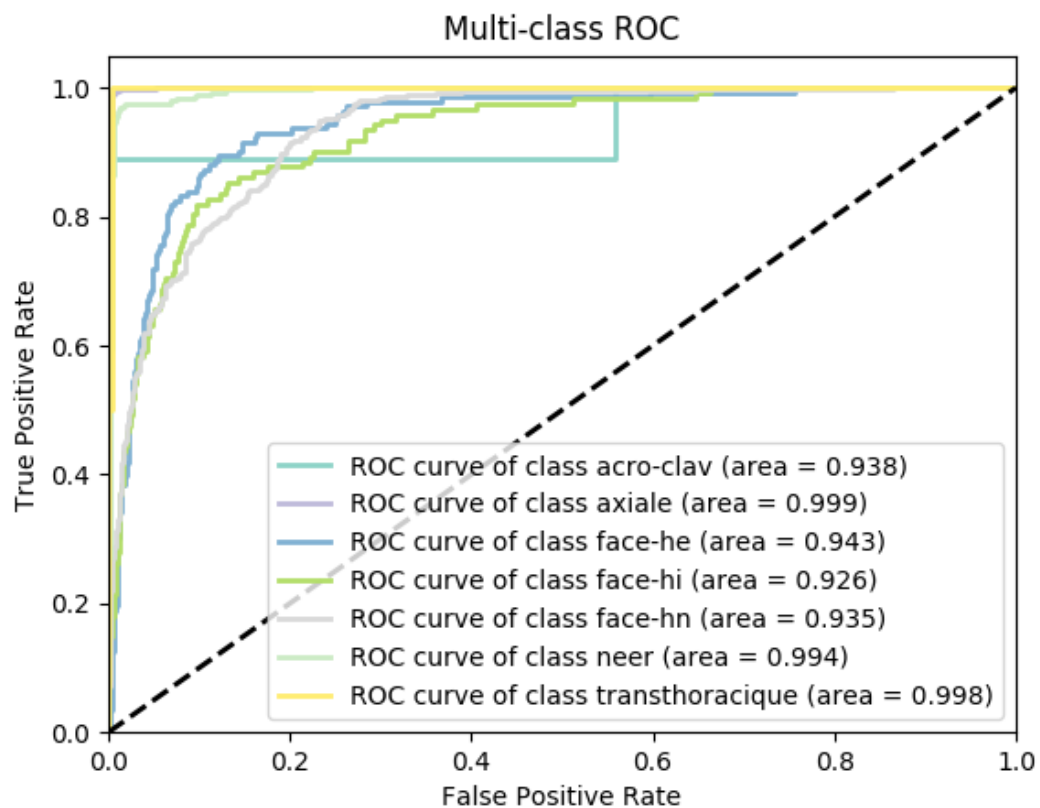
Stratégie

| | | Stratégie | | | | |
|-------------------------------|-------------------|------------|---------------|---------|------------|-------------|
| Architecture | Data augmentation | Balancing | No validation | k-folds | Paramètres | |
| Incidences de l'épaule | | | | | | |
| 5) | NASNet_a_large | Par défaut | X | X | X | bs 18 / e20 |

Performances (Accuracy) par classe et globales

| N° | Accuracy par classe | | | | | | | Accuracy globale |
|----|---------------------|--------|---------|---------|---------|-------|------------------|------------------|
| | Acro-clav | Axiale | Face HE | Face HI | Face HN | Neer | Trans-thoracique | |
| 5) | 0.889 | 0.992 | 0.056 | 0.582 | 0.861 | 0.969 | 0 | 0.813 |

Courbes ROC



Modèle n°6 : mura_shoulder_inc_balanced_e20_bs18_inceptionv4

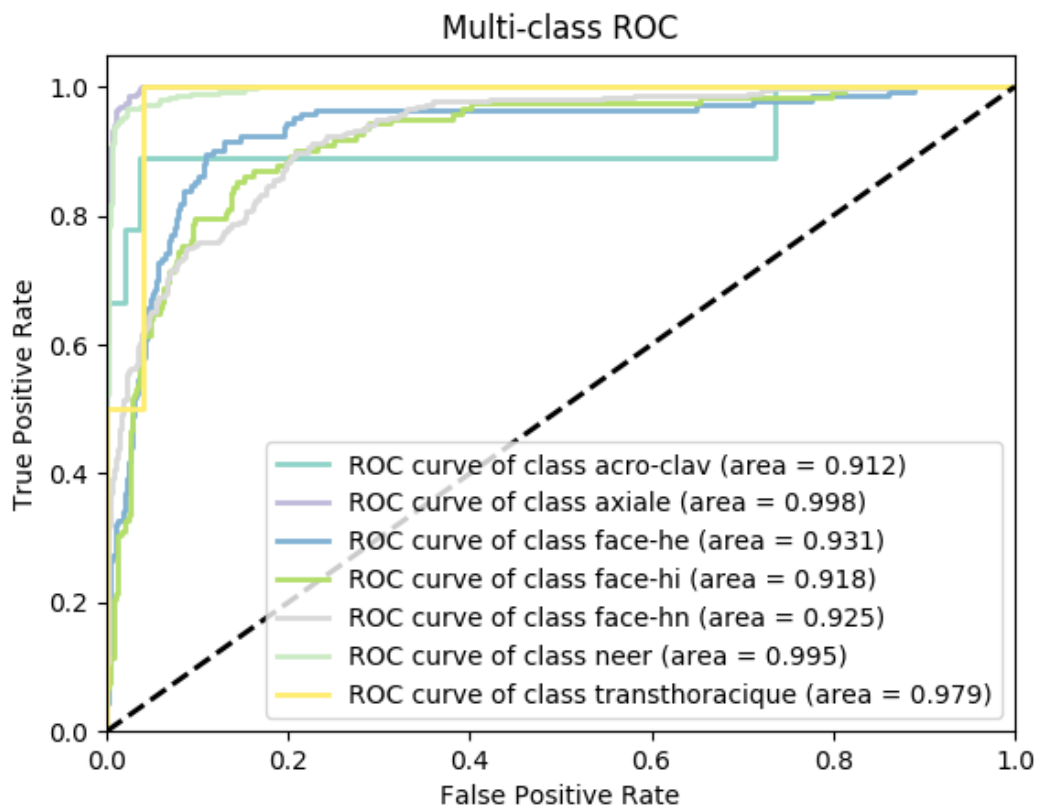
Stratégie

| | | Stratégie | | | | |
|-------------------------------|---------------------|------------|---------------------|---------|------------|-------------|
| Architecture | Data augmentation | Balancing | No validation | k-folds | Paramètres | |
| Incidences de l'épaule | | | | | | |
| 6) | Inception V4 | Par défaut | Sur-échantillonnage | X | X | bs 18 / e20 |

Performances (Accuracy) par classe et globales

| N° | Accuracy par classe | | | | | | | Accuracy globale |
|----|---------------------|--------|---------|---------|---------|-------|------------------|------------------|
| | Acro-clav | Axiale | Face HE | Face HI | Face HN | Neer | Trans-thoracique | |
| 6) | 0.444 | 0.988 | 0.839 | 0.672 | 0.533 | 0.969 | 0 | 0.806 |

Courbe ROC



Modèle n°7 : mura_shoulder_inc_no_augment_e20_bs18_Nasnet_a_large

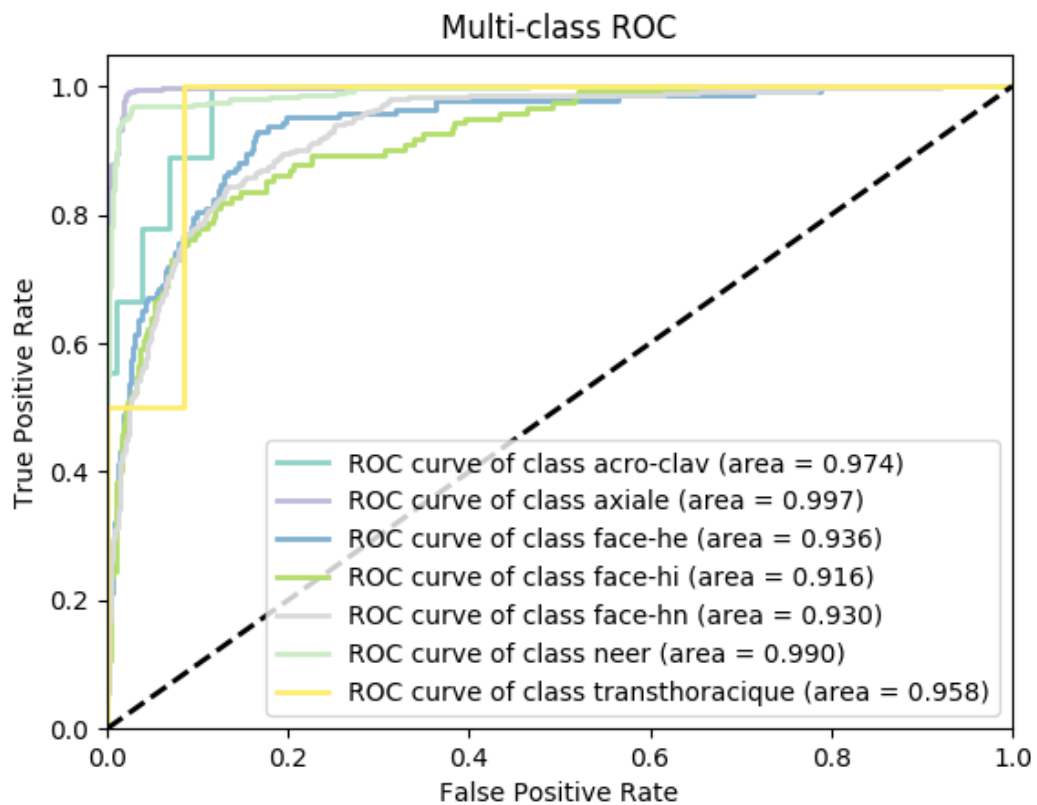
Stratégie

| | | Stratégie | | | | |
|-------------------------------|-------------------|-----------|---------------|---------|------------|-------------|
| Architecture | Data augmentation | Balancing | No validation | k-folds | Paramètres | |
| Incidences de l'épaule | | | | | | |
| 7) | NASNet_a_large | X | X | X | X | bs 18 / e20 |

Performances (Accuracy) par classe et globales

| N° | Accuracy par classe | | | | | | | Accuracy globale |
|----|---------------------|--------|---------|---------|---------|-------|------------------|------------------|
| | Acro-clav | Axiale | Face HE | Face HI | Face HN | Neer | Trans-thoracique | |
| 7) | 0.444 | 0.995 | 0.224 | 0.434 | 0.873 | 0.959 | 0 | 0.816 |

Courbe ROC



Modèle n°8 : mura_shoulder_inc_balanced_aug_xray5_full_e20_bs18_Nasnet_a_large

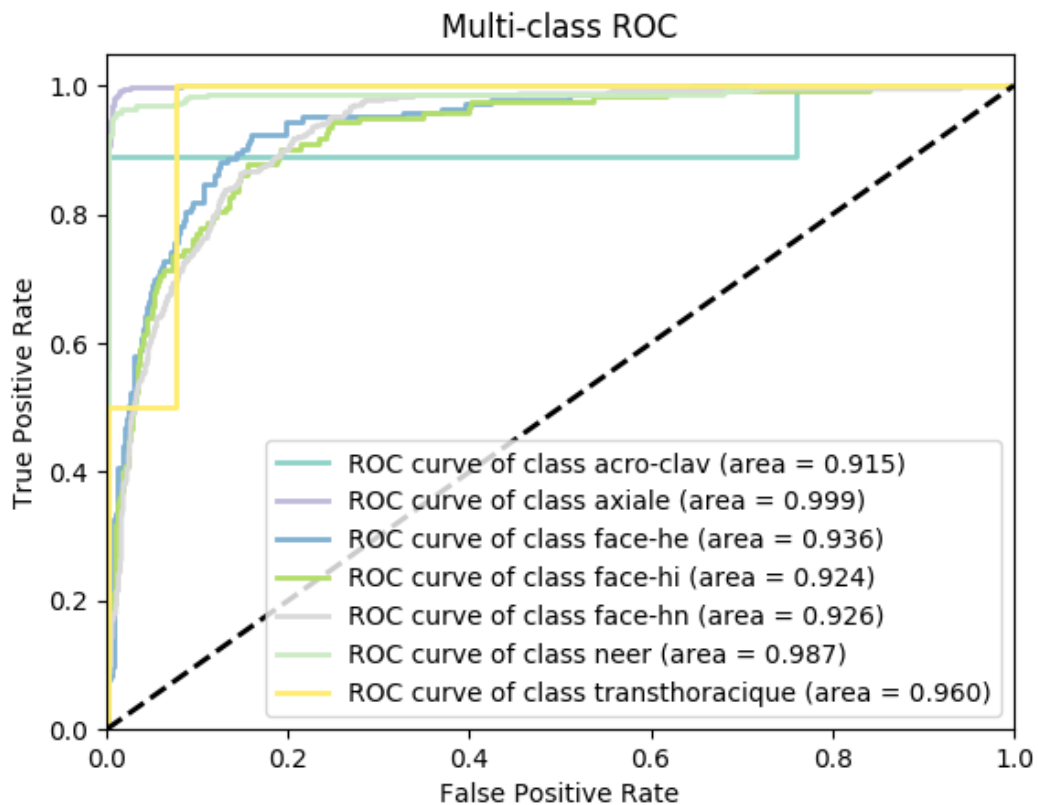
Stratégie

| | | Stratégie | | | | |
|-------------------------------|-------------------|-----------|---------------------|---------|------------|-------------|
| Architecture | Data augmentation | Balancing | No validation | k-folds | Paramètres | |
| Incidences de l'épaule | | | | | | |
| 8) | NASNet_a_large | Xray | Sur-échantillonnage | Full | 5 | bs 18 / e20 |

Performances (Accuracy) par classe et globales

| N° | Accuracy par classe | | | | | | | Accuracy globale |
|----|---------------------|--------|---------|---------|---------|-------|------------------|------------------|
| | Acro-clav | Axiale | Face HE | Face HI | Face HN | Neer | Trans-thoracique | |
| 8) | 0.666 | 0.992 | 0.489 | 0.680 | 0.752 | 0.959 | 0 | 0.833 |

Courbe ROC



Modèle n°9 : mura_shoulder_inc_balanced_aug_xray5_e20_bs18_Nasnet_a_large

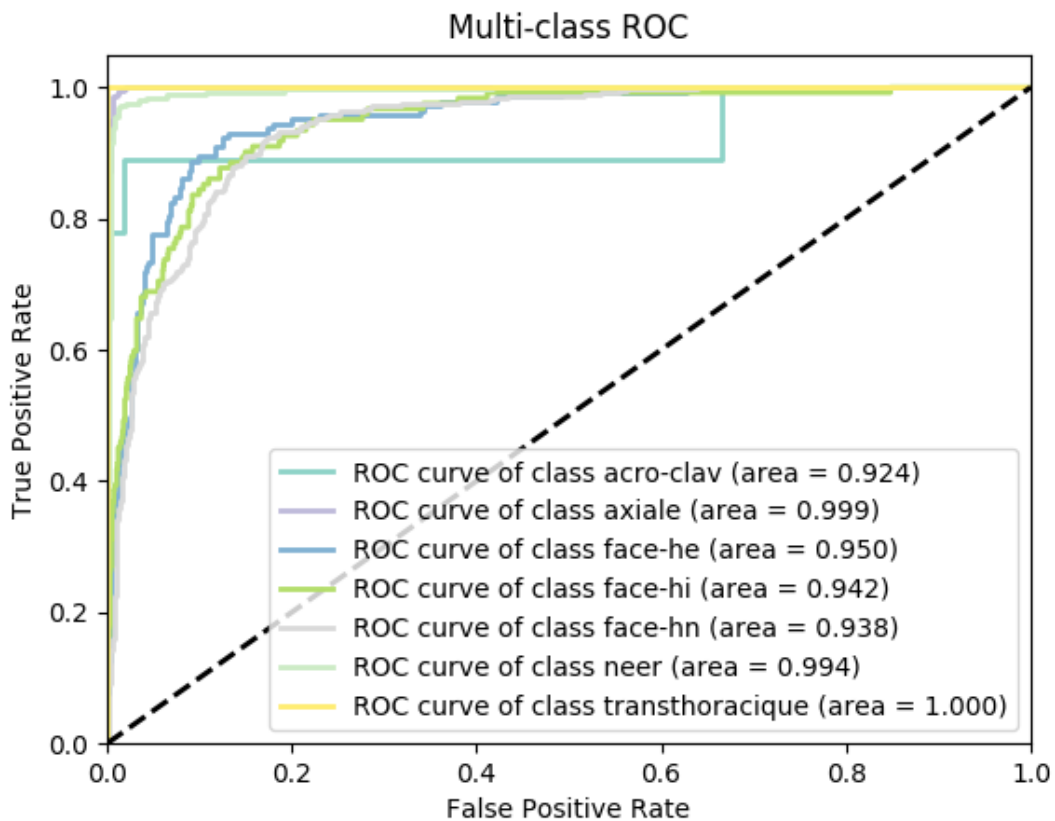
Stratégie

| | | Stratégie | | | | |
|-------------------------------|-------------------|-----------|---------------------|---------|------------|-------------|
| Architecture | Data augmentation | Balancing | No validation | k-folds | Paramètres | |
| Incidences de l'épaule | | | | | | |
| 9) | NASNet_a_large | Xray | Sur-échantillonnage | X | X | bs 18 / e20 |

Performances (Accuracy) par classe et globales

| N° | Accuracy par classe | | | | | | | Accuracy globale |
|----|---------------------|--------|---------|---------|---------|-------|------------------|------------------|
| | Acro-clav | Axiale | Face HE | Face HI | Face HN | Neer | Trans-thoracique | |
| 9) | 0.777 | 0.995 | 0.594 | 0.713 | 0.729 | 0.955 | 0.500 | 0.842 |

Courbes ROC



Modèle n°10 : mura_shoulder_inc_balanced_full_e20_bs18_Nasnet_a_large

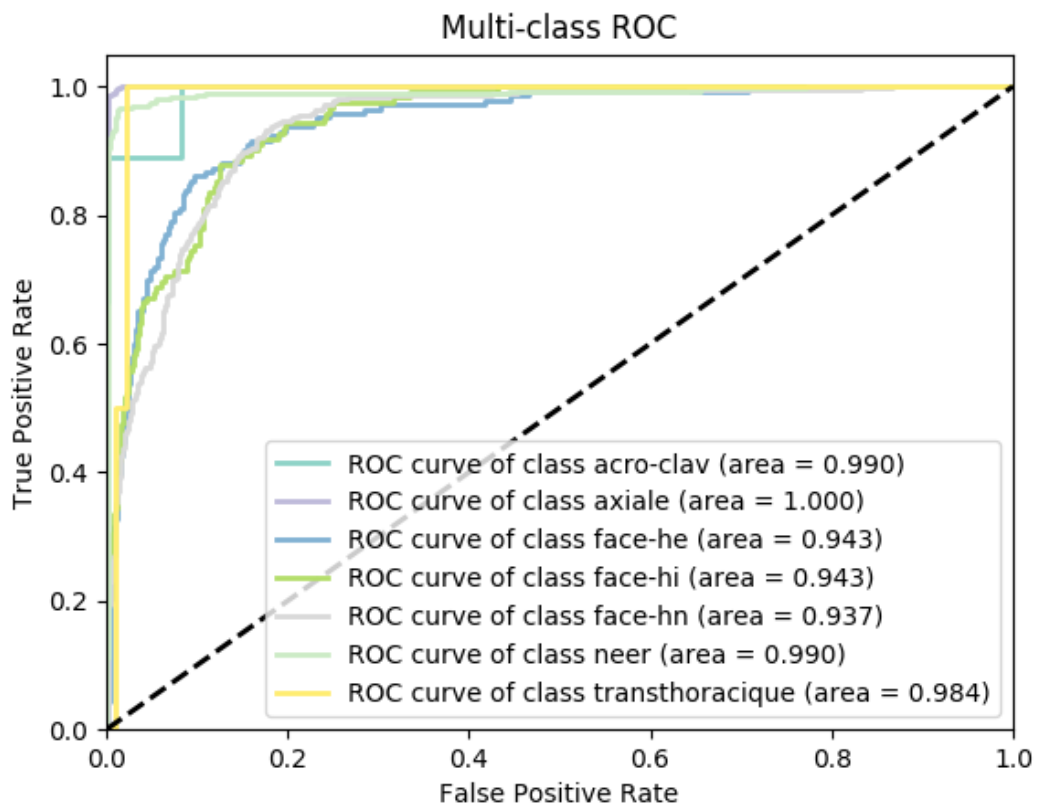
Stratégie

| | | Stratégie | | | | |
|-------------------------------|-------------------|------------|---------------------|---------|------------|-------------|
| Architecture | Data augmentation | Balancing | No validation | k-folds | Paramètres | |
| Incidences de l'épaule | | | | | | |
| 10) | NASNet_a_large | Par défaut | Sur-échantillonnage | Full | X | bs 18 / e20 |

Performances (Accuracy) par classe et globales

| N° | Accuracy par classe | | | | | | | Accuracy globale |
|-----|---------------------|--------|---------|---------|---------|-------|------------------|------------------|
| | Acro-clav | Axiale | Face HE | Face HI | Face HN | Neer | Trans-thoracique | |
| 10) | 0.666 | 0.995 | 0.657 | 0.680 | 0.703 | 0.959 | 0 | 0.837 |

Courbes ROC



Modèle n°11 : mura_shoulder_inc_e20_bs18_inceptionv4

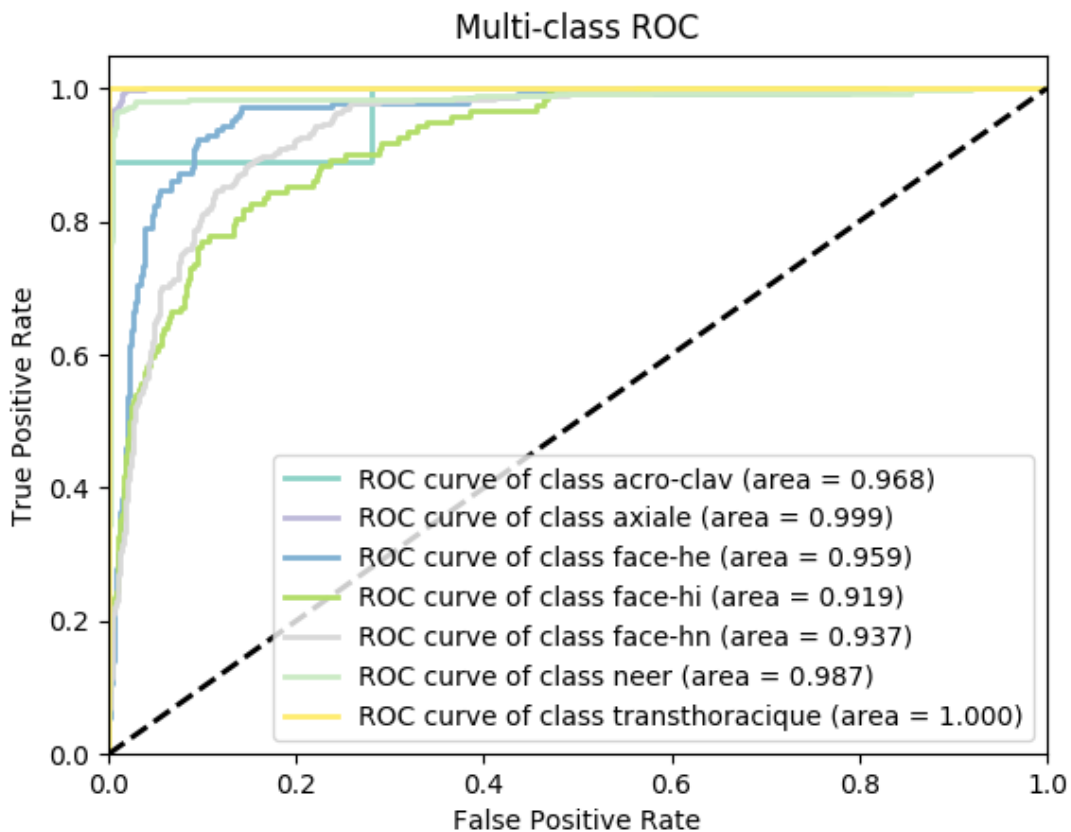
Stratégie

| | | Stratégie | | | | |
|-------------------------------|---------------------|------------|---------------|---------|------------|-------------|
| Architecture | Data augmentation | Balancing | No validation | k-folds | Paramètres | |
| Incidences de l'épaule | | | | | | |
| 11) | Inception V4 | Par défaut | X | X | X | bs 18 / e20 |

Performances (Accuracy) par classe et globales

| N° | Accuracy par classe | | | | | | | Accuracy globale |
|-----|---------------------|--------|---------|---------|---------|-------|------------------|------------------|
| | Acro-clav | Axiale | Face HE | Face HI | Face HN | Neer | Trans-thoracique | |
| 11) | 0.666 | 0.992 | 0.706 | 0.606 | 0.734 | 0.966 | 0 | 0.845 |

Courbes ROC



III.Résumé des erreurs par modèle et par classe

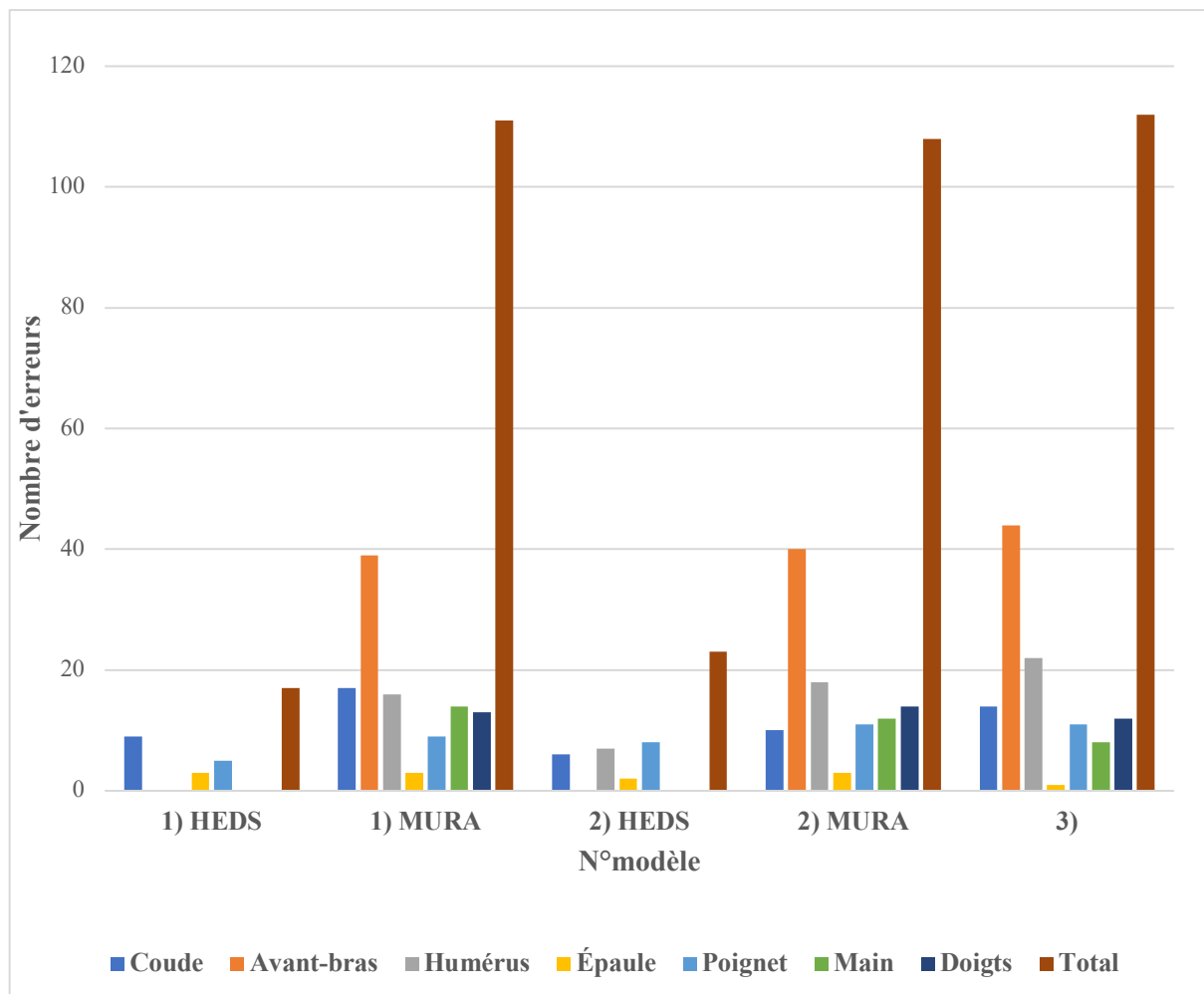
Régions anatomiques

Tableau – Erreurs par modèle et par classe

Test HEDS Test MURA

| N° | Erreurs | | | | | | | Total |
|----|---------|------------|---------|--------|---------|------|--------|-------|
| | Coude | Avant-bras | Humérus | Épaule | Poignet | Main | Doigts | |
| 1) | 9 | 0 | 0 | 3 | 9 | X | X | 21 |
| | 17 | 39 | 16 | 3 | 9 | 14 | 13 | 111 |
| 2) | 6 | 0 | 7 | 2 | 13 | X | X | 28 |
| | 10 | 40 | 18 | 3 | 11 | 12 | 14 | 108 |
| 3) | 14 | 44 | 22 | 1 | 11 | 8 | 12 | 112 |

Graphique – Erreurs par modèle et par classe

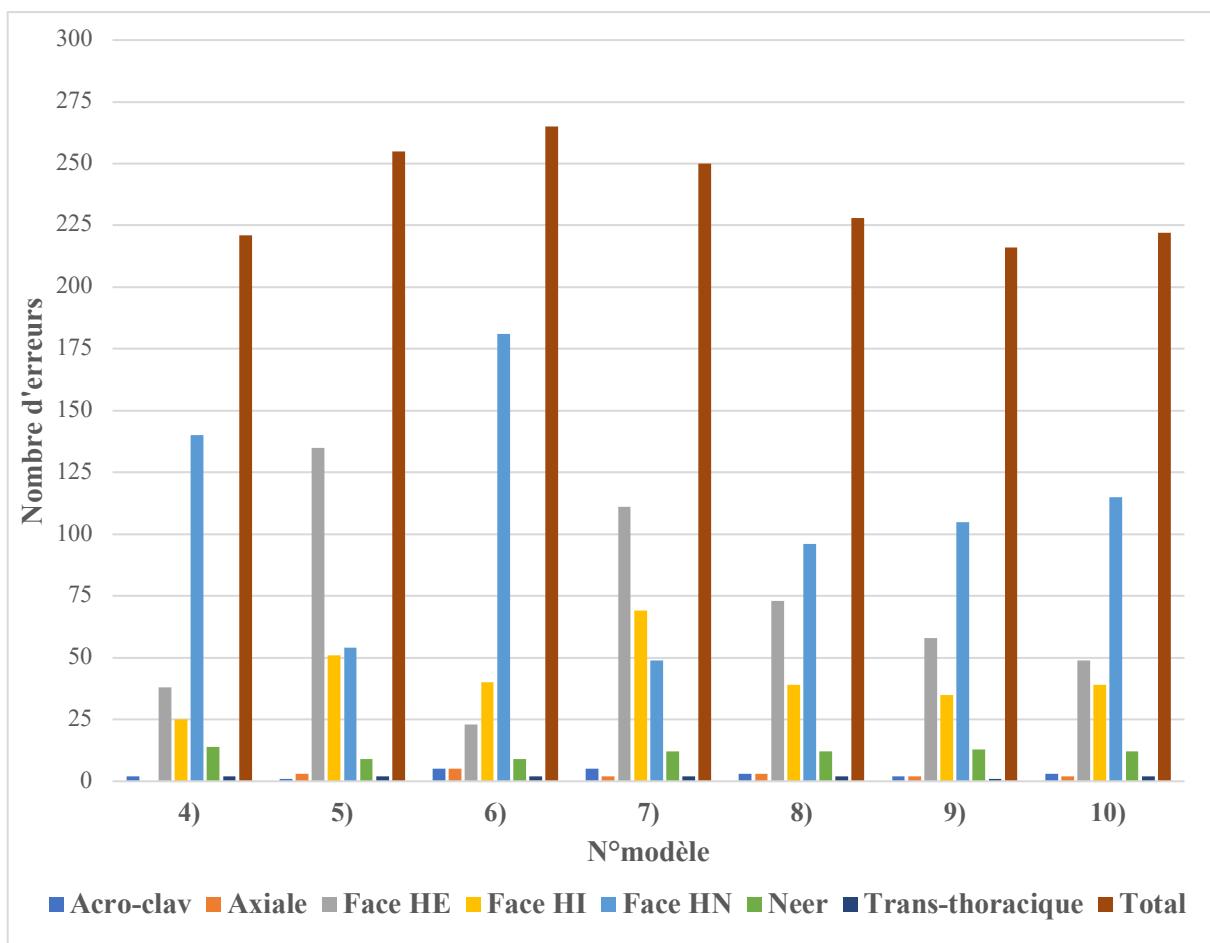


Incidences

Tableau – Erreurs par modèle et par classe

| N° | Erreurs | | | | | | | Total |
|-----|-----------|--------|---------|---------|---------|------|------------------|-------|
| | Acro-clav | Axiale | Face HE | Face HI | Face HN | Neer | Trans-thoracique | |
| 4) | 2 | 0 | 38 | 25 | 140 | 14 | 2 | 221 |
| 5) | 1 | 3 | 135 | 51 | 54 | 9 | 2 | 255 |
| 6) | 5 | 5 | 23 | 40 | 181 | 9 | 2 | 265 |
| 7) | 5 | 2 | 111 | 69 | 49 | 12 | 2 | 250 |
| 8) | 3 | 3 | 73 | 39 | 96 | 12 | 2 | 228 |
| 9) | 2 | 2 | 58 | 35 | 105 | 13 | 1 | 216 |
| 10) | 3 | 2 | 49 | 39 | 115 | 12 | 2 | 222 |
| 11) | 3 | 3 | 42 | 48 | 103 | 10 | 2 | 211 |

Graphique – Erreurs par modèle et par classe



IV. Matrice de confusion des classes par modèle

Régions anatomiques

Modèle n°1 : mura_anat_part_e20_bs18_Nasnet_a_large

| | Erreurs | | | | | | |
|------------|---------|------------|---------|--------|---------|------|--------|
| Réponse | Coude | Avant-bras | Humérus | Épaule | Poignet | Main | Doigts |
| Coude | | | 1 | 2 | 5 | | 1 |
| Avant-bras | | | | | | | |
| Humérus | | | | | | | |
| Épaule | | 2 | 1 | | | | |
| Poignet | | | | 4 | | | 5 |
| Main | | | | | | | |
| Doigts | | | | | | | |
| | Erreurs | | | | | | |
| N° | Coude | Avant-bras | Humérus | Épaule | Poignet | Main | Doigts |
| Coude | | | 11 | 5 | | | |
| Avant-bras | 18 | | 5 | 1 | 15 | | |
| Humérus | 6 | 4 | | 5 | | | 1 |
| Épaule | 1 | | 2 | | | | |
| Poignet | | 4 | | 1 | | 3 | |
| Main | | | | | 11 | | 3 |
| Doigts | | | | 1 | 2 | 10 | |

Modèle n°2 : mura_anat_part_e20_bs18_inceptionv4

| Erreurs | | | | | | | |
|------------|-------|------------|---------|--------|---------|------|--------|
| Réponse | Coude | Avant-bras | Humérus | Épaule | Poignet | Main | Doigts |
| Coude | | | | 1 | 4 | | 1 |
| Avant-bras | | | | | | | |
| Humérus | | 7 | | | | | |
| Épaule | 1 | | 1 | | | | |
| Poignet | 3 | | | 4 | | 1 | 5 |
| Main | | | | | | | |
| Doigts | | | | | | | |

| Erreurs | | | | | | | |
|------------|-------|------------|---------|--------|---------|------|--------|
| N° | Coude | Avant-bras | Humérus | Épaule | Poignet | Main | Doigts |
| Coude | | 3 | 3 | 2 | 2 | | |
| Avant-bras | 14 | | 3 | 3 | 19 | | 1 |
| Humérus | 7 | 2 | | 7 | | | 2 |
| Épaule | | 2 | | | 1 | | |
| Poignet | | 3 | | | | 5 | 3 |
| Main | | 1 | | | 9 | | 1 |
| Doigts | | | | 1 | 1 | 12 | |

Modèle n°3 : mura_anat_part_no_augment_e20_bs18_Nasnet_a_large

| Erreurs | | | | | | | |
|------------|-------|------------|---------|--------|---------|------|--------|
| N° | Coude | Avant-bras | Humérus | Épaule | Poignet | Main | Doigts |
| Coude | | 4 | 5 | 5 | | | |
| Avant-bras | 13 | | 4 | 8 | 16 | 1 | |
| Humérus | 6 | 8 | | 5 | 1 | | 1 |
| Épaule | | | 1 | | | | |
| Poignet | | 2 | | 1 | | 7 | 1 |
| Main | | | | | 7 | | 1 |
| Doigts | | | | 1 | 1 | 10 | |

Incidences

Modèle n°4 : mura_shoulder_inc_balanced_e20_bs18_Nasnet_a_large

| N° | Erreurs | | | | | | |
|--------|---------|--------|----|----|----|------|--------|
| | Acro. | Axiale | HE | HI | HN | Neer | Trans. |
| Acro. | | 1 | | 1 | | | |
| Axiale | | | | | | | |
| HE | | 1 | | 6 | 29 | 2 | |
| HI | 1 | 5 | 1 | | 16 | 2 | |
| HN | | 13 | 38 | 86 | | 3 | |
| Neer | | | 7 | 5 | 2 | | |
| Trans. | | | | | | 2 | |

Modèle n°5: mura_shoulder_inc_e20_bs18_Nasnet_a_large

| N° | Erreurs | | | | | | |
|--------|---------|--------|----|----|-----|------|--------|
| | Acro. | Axiale | HE | HI | HN | Neer | Trans. |
| Acro. | | | | 1 | | | |
| Axiale | | | | | 3 | | |
| HE | | | | 4 | 126 | 5 | |
| HI | | 4 | | | 41 | 6 | |
| HN | 4 | 5 | 3 | 38 | | 4 | |
| Neer | | | 2 | 2 | 5 | | |
| Trans. | | | | | | 2 | |

Modèle n°6 : mura_shoulder_inc_balanced_e20_bs18_inceptionv4

| N° | Erreurs | | | | | | |
|--------|---------|--------|----|----|----|------|--------|
| | Acro. | Axiale | HE | HI | HN | Neer | Trans. |
| Acro. | | 1 | | 1 | 1 | 2 | |
| Axiale | | | | 1 | | 4 | |
| HE | | 5 | | 4 | 8 | 6 | |
| HI | | 3 | 13 | | 13 | 11 | |
| HN | | 16 | 89 | 61 | | 15 | |
| Neer | | 1 | 3 | 4 | 1 | | |
| Trans. | | | | | | 2 | |

Modèle n°7 : mura_shoulder_inc_no_augment_e20_bs18_Nasnet_a_large

| Erreurs | | | | | | | |
|---------|-------|--------|----|----|----|------|--------|
| N° | Acro. | Axiale | HE | HI | HN | Neer | Trans. |
| Acro. | | 2 | | | 3 | | |
| Axiale | | | | | 1 | 1 | |
| HE | | | | 3 | 99 | 9 | |
| HI | | 7 | 3 | | 54 | 5 | |
| HN | | 20 | 5 | 14 | | 10 | |
| Neer | | 1 | | 3 | 8 | | |
| Trans. | | | | | | 2 | |

Modèle n°8 : mura_shoulder_inc_balanced_aug_xray5_full_e20_bs18_Nasnet_a_large

| Erreurs | | | | | | | |
|---------|-------|--------|----|----|----|------|--------|
| N° | Acro. | Axiale | HE | HI | HN | Neer | Trans. |
| Acro. | | 1 | | 1 | 1 | | |
| Axiale | | | | 1 | 2 | | |
| HE | | | | 3 | 64 | 6 | |
| HI | | 3 | 3 | | 28 | 5 | |
| HN | 1 | 10 | 22 | 60 | | 3 | |
| Neer | | | 5 | 1 | 5 | | |
| Trans. | | | | | | 2 | |

Modèle n°9 : mura_shoulder_inc_balanced_aug_xray5_e20_bs18_Nasnet_a_large

| Erreurs | | | | | | | |
|---------|-------|--------|----|----|----|------|--------|
| N° | Acro. | Axiale | HE | HI | HN | Neer | Trans. |
| Acro. | | 1 | | 1 | | | |
| Axiale | | | | 1 | 1 | | |
| HE | | | | 8 | 46 | 4 | |
| HI | | 4 | 1 | | 28 | 2 | |
| HN | | 11 | 29 | 62 | | 3 | |
| Neer | | | 6 | 3 | 4 | | |
| Trans. | | | | | | 1 | |

Modèle n°10 : mura_shoulder_inc_balanced_full_e20_bs18_Nasnet_a_large

| Erreurs | | | | | | | |
|---------------|-------|--------|----|----|----|------|--------|
| N° | Acro. | Axiale | HE | HI | HN | Neer | Trans. |
| Acro. | | 1 | | 2 | | | |
| Axiale | | | | | 2 | | |
| HE | | | | 4 | 42 | 3 | |
| HI | | 5 | 2 | | 30 | 2 | |
| HN | | 7 | 41 | 59 | | 8 | |
| Neer | | | 6 | 3 | 3 | | |
| Trans. | | | | | 1 | 1 | |

Modèle n°11 : mura_shoulder_inc_e20_bs18_inceptionv4

| Erreurs | | | | | | | |
|---------------|-------|--------|----|----|----|------|--------|
| N° | Acro. | Axiale | HE | HI | HN | Neer | Trans. |
| Acro. | | 1 | | 1 | 1 | | |
| Axiale | | | | | 3 | | |
| HE | | | | 2 | 35 | 5 | |
| HI | | 5 | 3 | | 36 | 4 | |
| HN | | 10 | 35 | 55 | | 3 | |
| Neer | | | 4 | 4 | 2 | | |
| Trans. | | | | | | 2 | |

Étude 2 – Travail pratique IA

V. Résumé des stratégies et performances par modèle

Incidences

Stratégie

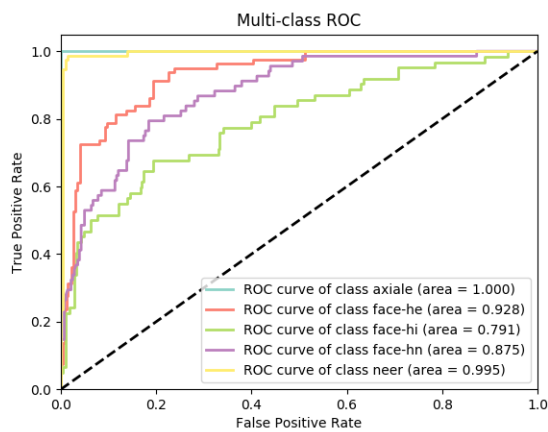
| Architecture | | Stratégie | | | | |
|-------------------------------|-----------------------|-------------------|---------------------|---------------|---------|-------------|
| | | Data augmentation | Balancing | No validation | k-folds | Paramètres |
| Incidences de l'épaule | | | | | | |
| 4) | NASNet_a_large | Par défaut | Sur-échantillonnage | X | X | bs 18 / e20 |
| 9) | NASNet_a_large | Xray | Sur-échantillonnage | X | X | bs 18 / e20 |

Performances IA

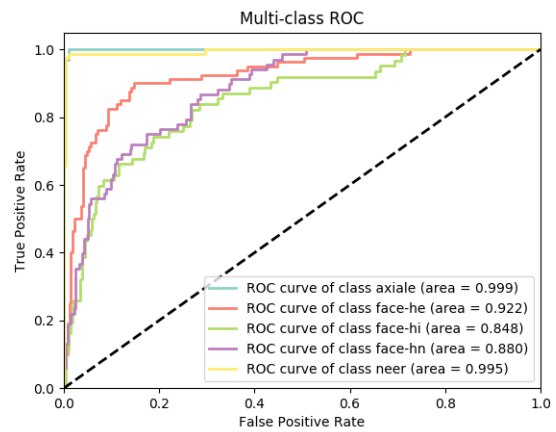
| N° | Accuracy par classe | | | | | Accuracy globale |
|----|---------------------|---------|---------|---------|-------|------------------|
| | Axiale | Face HE | Face HI | Face HN | Neer | |
| 4) | 0.984 | 0.525 | 0.258 | 0.853 | 0.973 | 0.723 |
| 9) | 1.00 | 0.625 | 0.322 | 0.764 | 0.987 | 0.745 |

Courbes ROC

Modèle n°4



Modèle n°9



VI.Résumé d'erreurs par modèle et par classe

TP - Incidences d'épaule

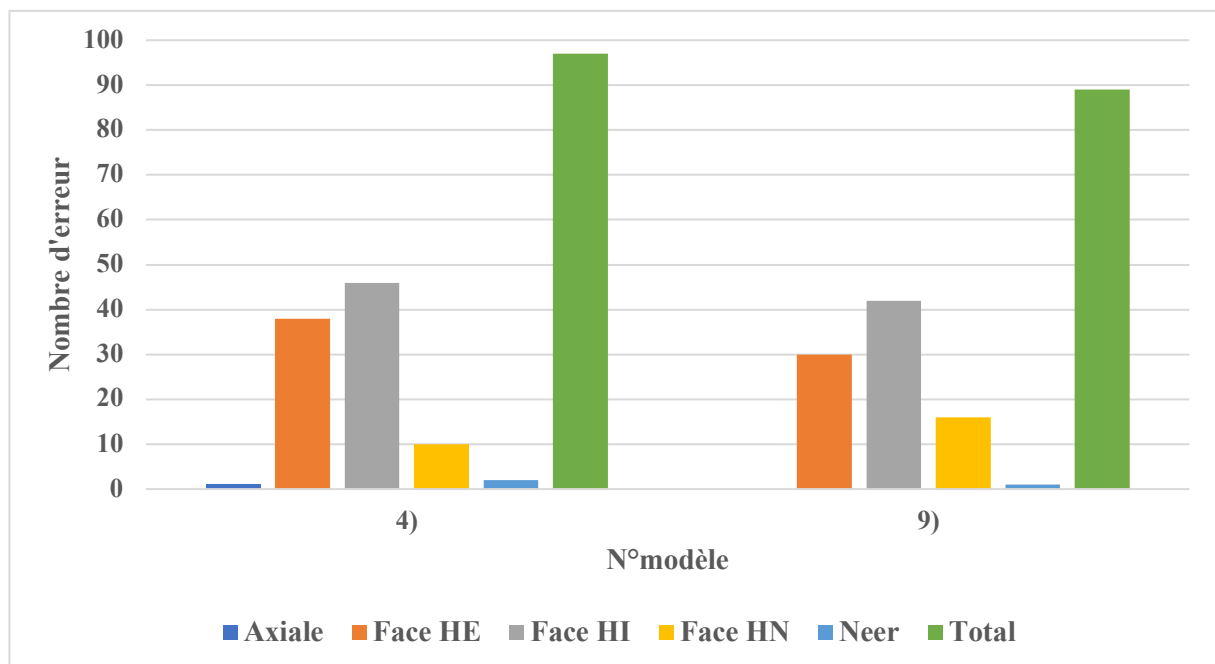
Tableau – Performances (accuracy)

| N° | Accuracy par classe | | | | | Accuracy Globale |
|----|---------------------|---------|---------|---------|-------|------------------|
| | Axiale | Face HE | Face HI | Face HN | Neer | |
| 4) | 0.984 | 0.525 | 0.258 | 0.853 | 0.973 | 0.723 |
| 9) | 1.00 | 0.625 | 0.322 | 0.764 | 0.987 | 0.745 |

Tableau - Erreurs

| N° | Erreurs | | | | | Total |
|----|---------|---------|---------|---------|------|-------|
| | Axiale | Face HE | Face HI | Face HN | Neer | |
| 4) | 1 | 38 | 46 | 10 | 2 | 97 |
| 9) | 0 | 30 | 42 | 16 | 1 | 89 |

Graphique - Erreurs par modèle et par classe



VII.TP IA – Protocole

TP IA 8 janvier 2020 3ème année

Durée totale : 90 minutes idéalement (Temps TP : 70 minutes)

Participant : 21 personnes

Salle : 314

Répartition salle : 1 poste / personne

Phase 0 : Préchauffe

Description : Taguer des images de félins

Équipe : 10 équipes de 2 et 1 personne individuelle

Base de données : Images de félins

Quantité : 32

But : S'approprier le logiciel en vue des étapes suivantes ;

Consigne générale : Taguer toutes les images correctement

Déroulement :

1. « Charger le fichier config.json en allant le chercher dans un répertoire accessible sur les ordinateurs des salles » ;
2. Taguer les images en cochant la (les) bonne(s) case(s);
3. Sauvegarder le fichier « tagging » dans un dossier distinct ;

Durée : à effectuer à la fin du cours théorique – 15 minutes

Phase 1 :

Description : Taguer des images par région anatomique selon les instructions contenues dans l'enveloppe ;

Équipe : Garder les mêmes équipes + 2 groupes IA forte / faible ;

Base de données : MURA - Membres supérieurs (épaules, coudes et mains uniquement)

Quantité : Toutes les images (TRAIN) (18'853 images ; Dataset équilibré) ;

But :

- 1) Faire une expérience plus radiologique faisant plus de sens pour les étudiants ;
- 2) Montrer dans la phase de retours avec les étudiants que l'IA peut être excellente ou non si pas assez de données (à définir lorsque nous allons créer cette IA faible) ;
- 3) Détecter ceux qui ont fait tout faux ou partiellement, de ceux qui ont fait tout juste.

Ceci présente l'idée du tagging indirect, ex. fait taguer des images par d'autres personnes (à leur insu), mais pour lesquels il faut s'assurer de la qualité de tagging en leur présentant des images dont nous connaissons la vérité ;

Consigne générale : Taguer le maximum d'image.

Déroulement :

1. Donner les enveloppes avec les consignes
 - Taguer tout correctement ;
 - Taguer tout faux ;
 - Taguer partiellement faux 1/3 ;
2. « Charger le fichier config.json en allant le chercher dans un répertoire accessible sur les ordinateurs des salles » ;
3. Taguer les images selon les consignes attribuées ;
4. Sauvegarder le fichier « tagging » dans un dossier distinct ;

Durée : 15 minutes

Phase 2 :

Description : Taguer des images d'incidences d'épaules selon la consigne ;

Équipe : Garder les mêmes équipes ;

Base de données : Incidence de l'articulation de l'épaule (Images TRAIN MURA) ;

Quantité : 350 images ; Tout le monde a les mêmes images ; 5 Incidences – Épaule de face (neutre – interne – externe) – Neer – Axiale ; Dataset équilibré

But :

1) Montrer un exemple plus complexe, mais faisant encore plus de sens d'un point de vue apprentissage, du moins c'est notre hypothèse ;

2) Montrer que le tagging devient plus difficile ;

Consigne générale : Taguer le maximum d'images correctement

Déroulement :

1. « Charger le fichier config.json en allant le chercher dans un répertoire accessible sur les ordinateurs des salles » ;

2. Consigne : Taguer les images correctement ;

3. Sauvegarder le fichier « tagging » ;

Durée : 20 minutes

Fin de la phase 2 : Distribution des fiches de rappel d'incidences (épaule face rot. HN / HI / HE, Axiale et Neer)

Phase 3 :

Description : Taguer des images d'épaules selon la consigne + feuille de rappel sur les incidences d'épaules ;

Équipe : Individuel ;

Base de données : Incidence de l'articulation de l'épaule (Images TRAIN MURA) ;

Quantité : 350 images ; Tout le monde a les mêmes images ; 5 Incidences – Épaule de face (neutre – interne – externe) – Neer – Axiale ; Dataset équilibré

But :

1) Voir si les performances sont meilleures avec la fiche de rappel ;

Consigne générale : Taguer correctement.

Déroulement :

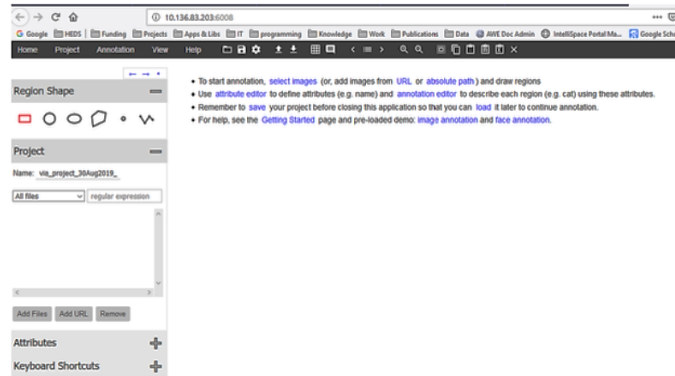
1. « Charger le fichier config.json en allant le chercher dans un répertoire accessible sur les ordinateurs des salles » ;
2. Consigne : Taguer les images correctement en s'appuyant sur la feuille avec critères de réussites ;
3. Sauvegarder le fichier « tagging » ;

Durée : 20 – 25 minutes

VIII.TP IA - Consignes

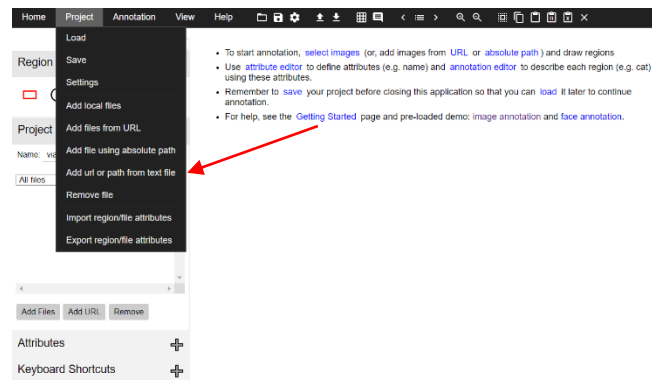
Phase 0

1. Ouvrir un navigateur internet.
2. Accédez à l'outil VIA. Adresse : <http://10.136.83.203:6008/>



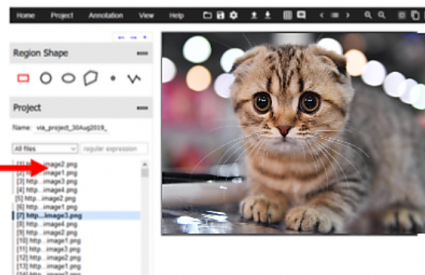
3. Préparation de l'outil de tagging : Cet outil va vous permettre d'associer un ou plusieurs labels (chat, lion, etc.) à vos images.

- Vous devez charger 2 choses :
 - Les images de félins ;
 - La liste des labels (attributs) ;

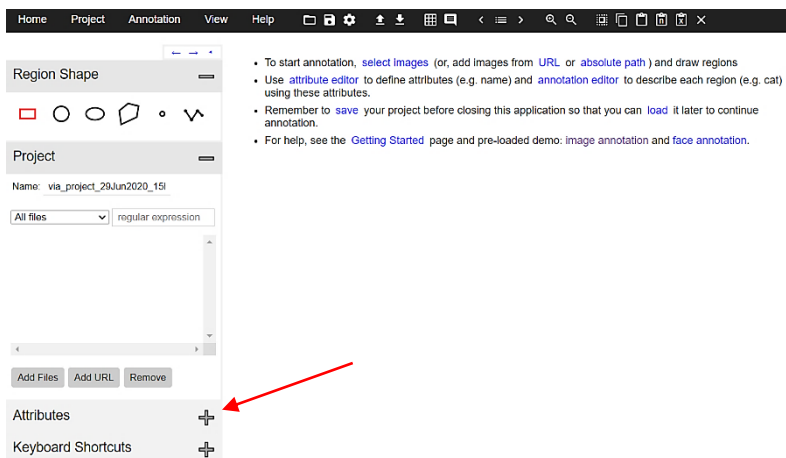


Après le chargement, vous devriez avoir ceci :

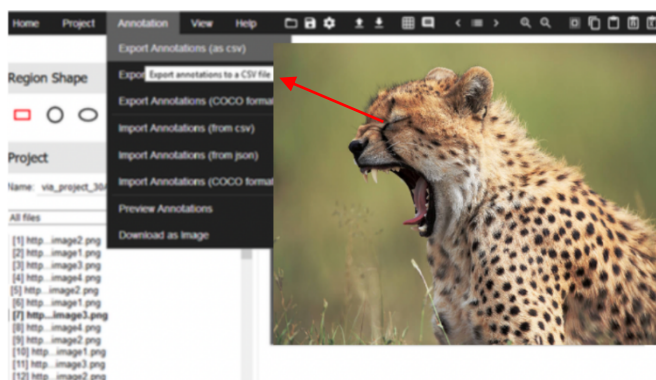
Liste d'images: un clic dessus ouvre l'image. Si ça ne marche pas, l'étape 1 n'a pas été faite correctement



- Si le chargement des labels s’est bien passé, vous devriez voir dans le menu « Attributes » et dans la catégorie « File Attributes », le menu de l’image ci-jointe qui vous montre une zone de choix qui correspond à vos 5 catégories de félins.



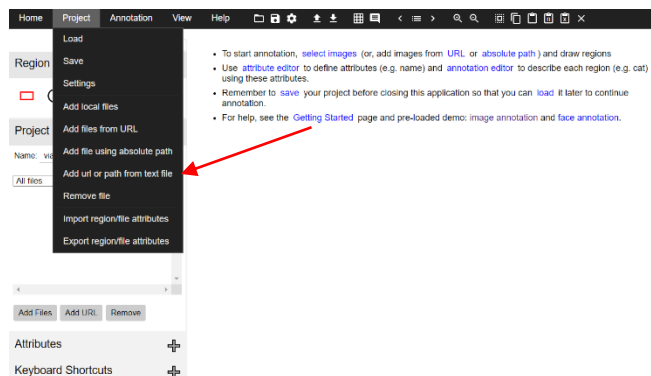
- Appuyez sur la barre d’espace pour que les cases à cocher (labels) apparaissent
4. Taguez les images en vous référant aux consignes dans l’enveloppe ;
 5. Une fois terminé, sauvegardez le fichier « tagging » dans un dossier distinct ;



Phase 1

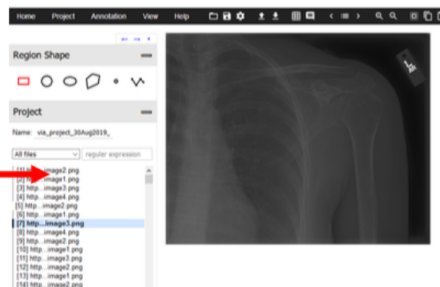
1. Préparation de l’outil de tagging : Cet outil va vous permettre d’associer un ou plusieurs labels à vos images.

- Vous devez charger 2 choses :
 - **Les images radiologiques ;**
 - **La liste des labels ;**

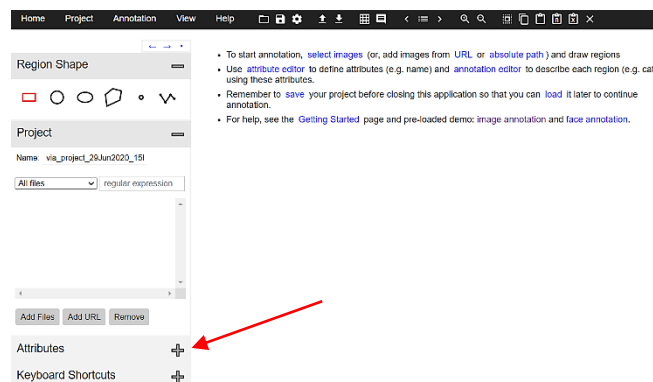


- Après le chargement, vous devriez avoir ceci :

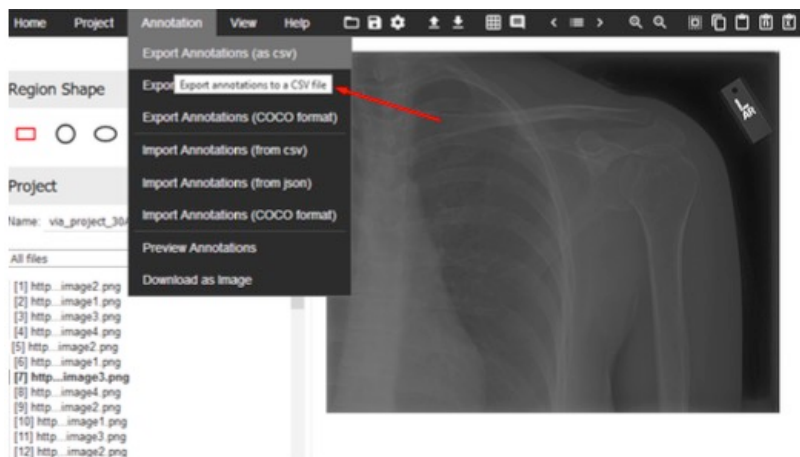
Liste d'images: un clic dessus ouvre l'image. Si ça ne marche pas, l'étape 1 n'a pas été faite correctement



- Si le chargement des labels s'est bien passé, vous devriez voir dans le menu « Attributes » et dans la catégorie « File Attributes », le menu de l'image ci-jointe qui vous montre une zone de choix qui correspond à vos 5 catégories de félins.



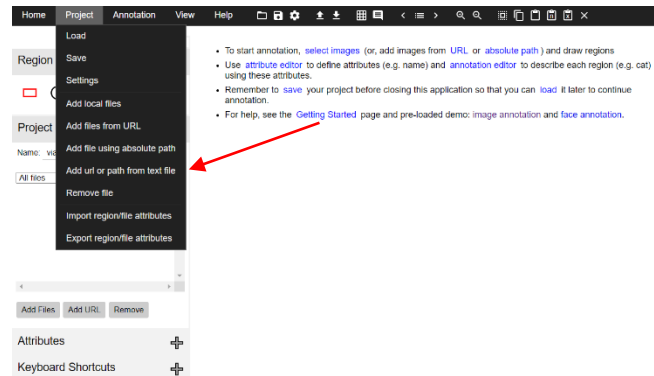
- Appuyez sur la barre d'espace pour que les cases à cocher (labels) apparaissent
2. Taguez les images en vous référant aux consignes dans l'enveloppe ;
 3. Une fois terminé, sauvegardez le fichier « tagging » dans un dossier distinct ;



Phase 2

1. Préparation de l’outil de tagging : Cet outil va vous permettre d’associer un ou plusieurs labels à vos images.

- Vous devez charger 2 choses :
 - **Les images radiologiques ;**
 - **La liste des labels ;**

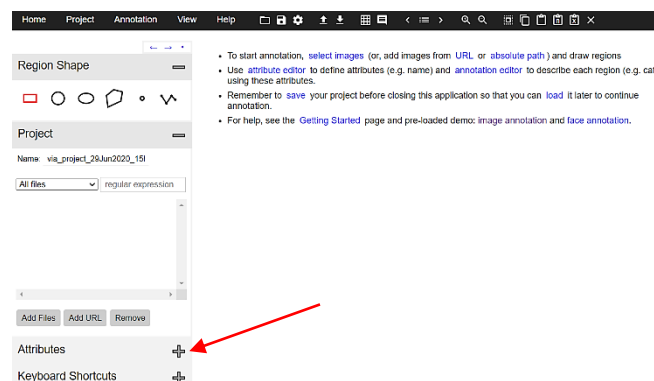


- Après le chargement, vous devriez avoir ceci :

Liste d’images: un clic dessus ouvre l’image. Si ça ne marche pas, l’étape 1 n’a pas été faite correctement

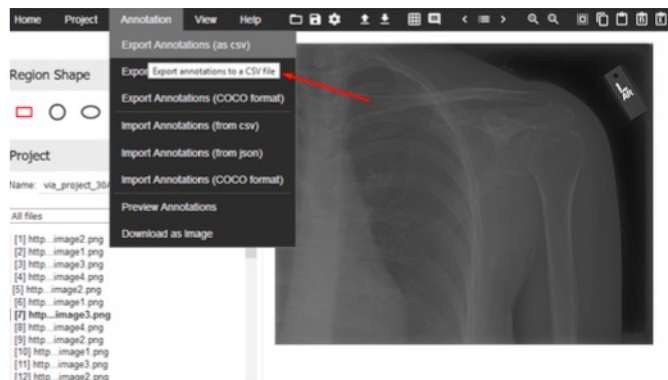


- Si le chargement des labels s’est bien passé, vous devriez voir dans le menu « Attributes » et dans la catégorie «File Attributes», le menu de l’image ci-jointe qui vous montre une zone de choix qui correspond à vos 5 incidences.



- Appuyez sur la barre d’espace pour que les cases à cocher (labels) apparaissent ;

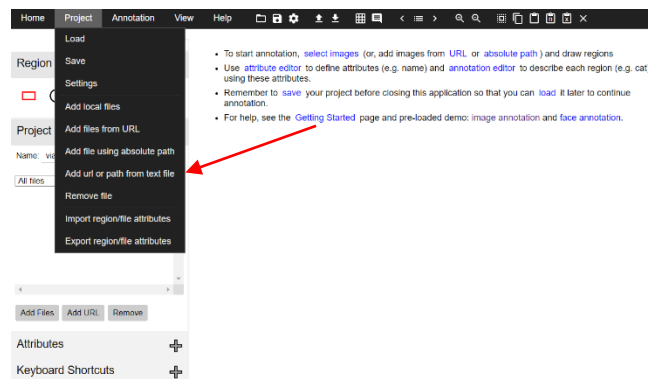
2. Taguez les images en vous référant aux consignes dans l'enveloppe ;
3. Une fois terminé, sauvegardez le fichier « tagging » dans un dossier distinct ;



Phase 3

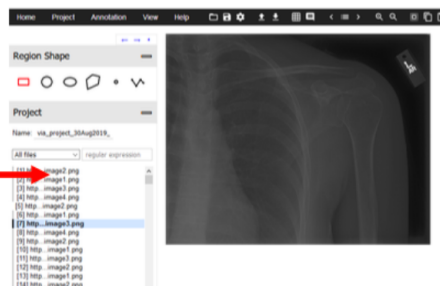
1. Préparation de l'outil de tagging : Cet outil va vous permettre d'associer un ou plusieurs labels à vos images.

- Vous devez charger 2 choses :
 - **Les images radiologiques ;**
 - **La liste des labels ;**

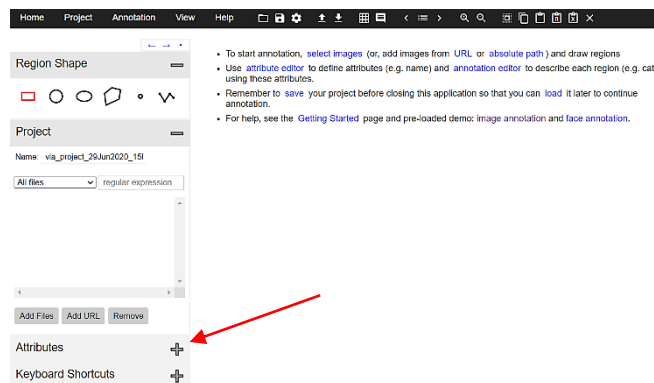


- Après le chargement, vous devriez avoir ceci :

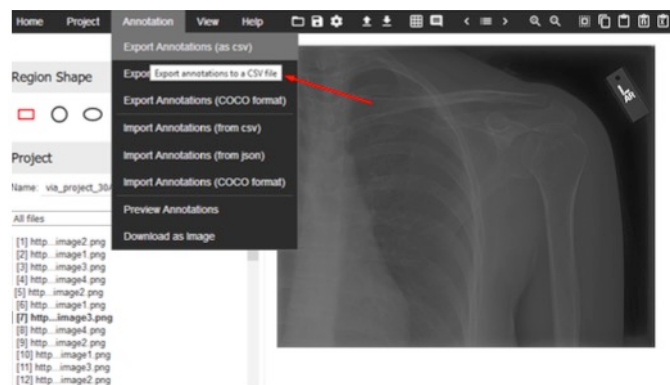
Liste d'images: un clic dessus ouvre l'image. Si ça ne marche pas, l'étape 1 n'a pas été faite correctement







- Si le chargement des labels s'est bien passé, vous devriez voir dans le menu « Attributes » et dans la catégorie «File Attributes», le menu de l'image ci-jointe qui vous montre une zone de choix qui correspond à vos 5 incidences.



- Appuyez sur la barre d'espace pour que les cases à cocher (labels) apparaissent ;
2. Taguez les images en vous référant aux consignes dans l'enveloppe ;
 3. Une fois terminé, sauvegardez le fichier « tagging » dans un dossier distinct ;



IX.TP IA - Équipes

| Logo | Équipe | Membres |
|---|------------------------|-----------------------|
|  | <p>Avengers</p> | <p>Iron Man</p> |
| | | <p>Captain Marvel</p> |
|  | <p>Supernanas</p> | <p>Belle</p> |
| | | <p>Rebelle</p> |
|  | <p>Teen Titans</p> | <p>Changelin</p> |
| | | <p>Starfire</p> |
|  | <p>Totally Spies</p> | <p>Clover</p> |
| | | <p>Sam</p> |
|  | <p>Indestructibles</p> | <p>Elastigirl</p> |
| | | <p>Violette</p> |

| | | |
|---|----------------|-----------------|
|  | X-Men | Wolverine |
| | | Magneto |
|  | Power-Rangers | Rouge |
| | | Noir |
|  | Minions | Un œil |
| | | Deux yeux |
|  | 4 Fantastiques | Torche humaine |
| | | Surfer d'argent |
|  | Tortues Ninja | Donatello |
| | | Leonardo |
|  | Pokémon | Tortank |

X.TP IA - Résultats des étudiants (Excel)

| Participant | Images | TP | TN | FP | FN | TPR | TNR | Precision | FPR | ACC |
|------------------------|--------|----|----|----|----|------|------|-----------|------|------|
| Phase 0 - Félin | | | | | | | | | | |
| Chat | | | | | | | | | | |
| IRON_MAN | 32 | 6 | 25 | 1 | 0 | 1.00 | 0.04 | 0.86 | 0.04 | 0.97 |
| WOLVERINE | 32 | 6 | 25 | 1 | 0 | 1.00 | 0.04 | 0.86 | 0.04 | 0.97 |
| 1_OEIL | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| 2_YEUX | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| BELLE | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| CAPTAIN_MARVEL | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| CHANGELIN | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| CLOVER | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| DONATELLO | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| ELASTIGIRL | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| LEONARDO | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| MAGNETO | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| NOIR | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| REBELLE | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| ROUGE | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| SAM | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| STARFIRE | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| SURFEUR_ARGENT | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| TORCHE_HUMAINE | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| TORTANK | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| Guépard | | | | | | | | | | |
| 2_YEUX | 32 | 3 | 27 | 0 | 2 | 0.60 | 0.00 | 1.00 | 0.00 | 0.94 |
| 1_OEIL | 32 | 4 | 27 | 0 | 1 | 0.80 | 0.00 | 1.00 | 0.00 | 0.97 |
| BELLE | 32 | 4 | 27 | 0 | 1 | 0.80 | 0.00 | 1.00 | 0.00 | 0.97 |
| CAPTAIN_MARVEL | 32 | 4 | 27 | 0 | 1 | 0.80 | 0.00 | 1.00 | 0.00 | 0.97 |
| CLOVER | 32 | 4 | 27 | 0 | 1 | 0.80 | 0.00 | 1.00 | 0.00 | 0.97 |
| DONATELLO | 32 | 4 | 27 | 0 | 1 | 0.80 | 0.00 | 1.00 | 0.00 | 0.97 |
| IRON_MAN | 32 | 5 | 26 | 1 | 0 | 1.00 | 0.04 | 0.83 | 0.04 | 0.97 |
| LEONARDO | 32 | 4 | 27 | 0 | 1 | 0.80 | 0.00 | 1.00 | 0.00 | 0.97 |
| MAGNETO | 32 | 4 | 27 | 0 | 1 | 0.80 | 0.00 | 1.00 | 0.00 | 0.97 |
| NOIR | 32 | 4 | 27 | 0 | 1 | 0.80 | 0.00 | 1.00 | 0.00 | 0.97 |
| REBELLE | 32 | 4 | 27 | 0 | 1 | 0.80 | 0.00 | 1.00 | 0.00 | 0.97 |
| ROUGE | 32 | 4 | 27 | 0 | 1 | 0.80 | 0.00 | 1.00 | 0.00 | 0.97 |
| SAM | 32 | 4 | 27 | 0 | 1 | 0.80 | 0.00 | 1.00 | 0.00 | 0.97 |
| WOLVERINE | 32 | 5 | 26 | 1 | 0 | 1.00 | 0.04 | 0.83 | 0.04 | 0.97 |
| CHANGELIN | 32 | 5 | 27 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| ELASTIGIRL | 32 | 5 | 27 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| STARFIRE | 32 | 5 | 27 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| SURFEUR_ARGENT | 32 | 5 | 27 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| TORCHE_HUMAINE | 32 | 5 | 27 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| TORTANK | 32 | 5 | 27 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |

| Lion | | | | | | | | | | |
|----------------|----|---|----|---|---|------|------|------|------|------|
| 2_YEUX | 32 | 5 | 24 | 0 | 3 | 0.63 | 0.00 | 1.00 | 0.00 | 0.91 |
| WOLVERINE | 32 | 6 | 23 | 1 | 2 | 0.75 | 0.04 | 0.86 | 0.04 | 0.91 |
| MAGNETO | 32 | 7 | 23 | 1 | 1 | 0.88 | 0.04 | 0.88 | 0.04 | 0.94 |
| STARFIRE | 32 | 6 | 24 | 0 | 2 | 0.75 | 0.00 | 1.00 | 0.00 | 0.94 |
| SURFEUR_ARGENT | 32 | 6 | 24 | 0 | 2 | 0.75 | 0.00 | 1.00 | 0.00 | 0.94 |
| 1_OEIL | 32 | 8 | 23 | 1 | 0 | 1.00 | 0.04 | 0.89 | 0.04 | 0.97 |
| BELLE | 32 | 8 | 23 | 1 | 0 | 1.00 | 0.04 | 0.89 | 0.04 | 0.97 |
| CAPTAIN_MARVEL | 32 | 7 | 24 | 0 | 1 | 0.88 | 0.00 | 1.00 | 0.00 | 0.97 |
| CHANGELIN | 32 | 7 | 24 | 0 | 1 | 0.88 | 0.00 | 1.00 | 0.00 | 0.97 |
| DONATELLO | 32 | 8 | 23 | 1 | 0 | 1.00 | 0.04 | 0.89 | 0.04 | 0.97 |
| IRON_MAN | 32 | 8 | 23 | 1 | 0 | 1.00 | 0.04 | 0.89 | 0.04 | 0.97 |
| NOIR | 32 | 8 | 23 | 1 | 0 | 1.00 | 0.04 | 0.89 | 0.04 | 0.97 |
| REBELLE | 32 | 8 | 23 | 1 | 0 | 1.00 | 0.04 | 0.89 | 0.04 | 0.97 |
| CLOVER | 32 | 8 | 24 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| ELASTIGIRL | 32 | 8 | 24 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| LEONARDO | 32 | 8 | 24 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| ROUGE | 32 | 8 | 24 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| SAM | 32 | 8 | 24 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| TORCHE_HUMAINE | 32 | 8 | 24 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| TORTANK | 32 | 8 | 24 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| Lynx | | | | | | | | | | |
| CAPTAIN_MARVEL | 32 | 4 | 26 | 1 | 1 | 0.80 | 0.04 | 0.80 | 0.04 | 0.94 |
| IRON_MAN | 32 | 3 | 27 | 0 | 2 | 0.60 | 0.00 | 1.00 | 0.00 | 0.94 |
| CLOVER | 32 | 5 | 26 | 1 | 0 | 1.00 | 0.04 | 0.83 | 0.04 | 0.97 |
| LEONARDO | 32 | 5 | 26 | 1 | 0 | 1.00 | 0.04 | 0.83 | 0.04 | 0.97 |
| ROUGE | 32 | 5 | 26 | 1 | 0 | 1.00 | 0.04 | 0.83 | 0.04 | 0.97 |
| SAM | 32 | 5 | 26 | 1 | 0 | 1.00 | 0.04 | 0.83 | 0.04 | 0.97 |
| 1_OEIL | 32 | 5 | 27 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| 2_YEUX | 32 | 5 | 27 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| BELLE | 32 | 5 | 27 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| CHANGELIN | 32 | 5 | 27 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| DONATELLO | 32 | 5 | 27 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| ELASTIGIRL | 32 | 5 | 27 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| MAGNETO | 32 | 5 | 27 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| NOIR | 32 | 5 | 27 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| REBELLE | 32 | 5 | 27 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| STARFIRE | 32 | 5 | 27 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| SURFEUR_ARGENT | 32 | 5 | 27 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| TORCHE_HUMAINE | 32 | 5 | 27 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| TORTANK | 32 | 5 | 27 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| WOLVERINE | 32 | 5 | 27 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |

| Tigre | | | | | | | | | | |
|--------------------------------------|-------|------|-------|-----|-----|------|------|------|------|------|
| 2_YEUX | 32 | 3 | 26 | 0 | 3 | 0.50 | 0.00 | 1.00 | 0.00 | 0.91 |
| WOLVERINE | 32 | 6 | 24 | 2 | 0 | 1.00 | 0.08 | 0.75 | 0.08 | 0.94 |
| CAPTAIN_MARVEL | 32 | 6 | 25 | 1 | 0 | 1.00 | 0.04 | 0.86 | 0.04 | 0.97 |
| CHANGELIN | 32 | 5 | 26 | 0 | 1 | 0.83 | 0.00 | 1.00 | 0.00 | 0.97 |
| DONATELLO | 32 | 5 | 26 | 0 | 1 | 0.83 | 0.00 | 1.00 | 0.00 | 0.97 |
| IRON_MAN | 32 | 6 | 25 | 1 | 0 | 1.00 | 0.04 | 0.86 | 0.04 | 0.97 |
| MAGNETO | 32 | 5 | 26 | 0 | 1 | 0.83 | 0.00 | 1.00 | 0.00 | 0.97 |
| 1_OEIL | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| BELLE | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| CLOVER | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| ELASTIGIRL | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| LEONARDO | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| NOIR | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| REBELLE | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| ROUGE | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| SAM | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| STARFIRE | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| SURFEUR_ARGENT | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| TORCHE_HUMAINE | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| TORTANK | 32 | 6 | 26 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| Phase 1 – Régions anatomiques | | | | | | | | | | |
| Épaule | | | | | | | | | | |
| TOTALLY_SPIES | 285 | 0 | 92 | 93 | 100 | 0.00 | 0.50 | 0.00 | 0.50 | 0.32 |
| X-MEN | 318 | 71 | 170 | 40 | 37 | 0.66 | 0.19 | 0.64 | 0.19 | 0.76 |
| Thanos | 18853 | 8295 | 10369 | 105 | 84 | 0.99 | 0.01 | 0.99 | 0.01 | 0.99 |
| INDESTRUCTIBLES | 439 | 146 | 291 | 1 | 1 | 0.99 | 0.00 | 0.99 | 0.00 | 1.00 |
| MINIONS | 360 | 119 | 240 | 0 | 1 | 0.99 | 0.00 | 1.00 | 0.00 | 1.00 |
| TORTUES_NINJAS | 421 | 139 | 281 | 0 | 1 | 0.99 | 0.00 | 1.00 | 0.00 | 1.00 |
| POWER_RANGERS | 422 | 140 | 281 | 0 | 1 | 0.99 | 0.00 | 1.00 | 0.00 | 1.00 |
| AVENGERS | 449 | 151 | 297 | 1 | 0 | 1.00 | 0.00 | 0.99 | 0.00 | 1.00 |
| SUPER_NANAS | 453 | 150 | 302 | 0 | 1 | 0.99 | 0.00 | 1.00 | 0.00 | 1.00 |
| TEEN_TITANS | 480 | 157 | 322 | 0 | 1 | 0.99 | 0.00 | 1.00 | 0.00 | 1.00 |
| LES_4_FANTASTIQUES | 486 | 161 | 324 | 0 | 1 | 0.99 | 0.00 | 1.00 | 0.00 | 1.00 |
| POKEMON | 444 | 149 | 295 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| Coude | | | | | | | | | | |
| TOTALLY_SPIES | 285 | 0 | 109 | 70 | 106 | 0.00 | 0.39 | 0.00 | 0.39 | 0.38 |
| X-MEN | 318 | 79 | 175 | 27 | 37 | 0.68 | 0.13 | 0.75 | 0.13 | 0.80 |
| Thanos | 18853 | 4734 | 13365 | 557 | 197 | 0.96 | 0.04 | 0.89 | 0.04 | 0.96 |
| POWER_RANGERS | 422 | 150 | 270 | 1 | 1 | 0.99 | 0.00 | 0.99 | 0.00 | 1.00 |
| INDESTRUCTIBLES | 439 | 154 | 283 | 1 | 1 | 0.99 | 0.00 | 0.99 | 0.00 | 1.00 |
| MINIONS | 360 | 128 | 231 | 1 | 0 | 1.00 | 0.00 | 0.99 | 0.00 | 1.00 |
| TORTUES_NINJAS | 421 | 150 | 270 | 1 | 0 | 1.00 | 0.00 | 0.99 | 0.00 | 1.00 |
| AVENGERS | 449 | 155 | 293 | 0 | 1 | 0.99 | 0.00 | 1.00 | 0.00 | 1.00 |
| SUPER_NANAS | 453 | 160 | 292 | 1 | 0 | 1.00 | 0.00 | 0.99 | 0.00 | 1.00 |
| TEEN_TITANS | 480 | 166 | 313 | 1 | 0 | 1.00 | 0.00 | 0.99 | 0.00 | 1.00 |
| LES_4_FANTASTIQUES | 486 | 168 | 317 | 1 | 0 | 1.00 | 0.00 | 0.99 | 0.00 | 1.00 |
| POKEMON | 444 | 153 | 291 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |

| Main | | | | | | | | | | |
|---|-------|------|-------|-----|-----|------|------|------|------|------|
| TOTALLY_SPIES | 285 | 0 | 84 | 122 | 79 | 0.00 | 0.59 | 0.00 | 0.59 | 0.29 |
| X-MEN | 318 | 74 | 197 | 27 | 20 | 0.79 | 0.12 | 0.73 | 0.12 | 0.85 |
| Thanos | 18853 | 5377 | 12911 | 399 | 166 | 0.97 | 0.03 | 0.93 | 0.03 | 0.97 |
| POWER_RANGERS | 422 | 130 | 291 | 1 | 0 | 1.00 | 0.00 | 0.99 | 0.00 | 1.00 |
| AVENGERS | 449 | 142 | 307 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| INDESTRUCTIBLES | 439 | 137 | 302 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| LES_4_FANTASTIQUES | 486 | 156 | 330 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| MINIONS | 360 | 112 | 248 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| POKEMON | 444 | 142 | 302 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| SUPER_NANAS | 453 | 142 | 311 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| TEEN_TITANS | 480 | 156 | 324 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| TORTUES_NINJAS | 421 | 131 | 290 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| Phase 2 – Incidences d'épaule (sans fiche de rappel) | | | | | | | | | | |
| Axiale | | | | | | | | | | |
| MINIONS | 180 | 24 | 152 | 0 | 4 | 0.86 | 0.00 | 1.00 | 0.00 | 0.98 |
| TOTALLY_SPIES | 166 | 23 | 141 | 0 | 2 | 0.92 | 0.00 | 1.00 | 0.00 | 0.99 |
| TORTUES_NINJAS | 282 | 52 | 227 | 1 | 2 | 0.96 | 0.00 | 0.98 | 0.00 | 0.99 |
| AVENGERS | 210 | 35 | 173 | 0 | 2 | 0.95 | 0.00 | 1.00 | 0.00 | 0.99 |
| INDESTRUCTIBLES | 258 | 46 | 210 | 0 | 2 | 0.96 | 0.00 | 1.00 | 0.00 | 0.99 |
| SUPER_NANAS | 135 | 20 | 114 | 0 | 1 | 0.95 | 0.00 | 1.00 | 0.00 | 0.99 |
| LES_4_FANTASTIQUES | 213 | 36 | 176 | 0 | 1 | 0.97 | 0.00 | 1.00 | 0.00 | 1.00 |
| POKEMON | 106 | 17 | 89 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| POWER_RANGERS | 197 | 35 | 162 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| TEEN_TITANS | 107 | 19 | 88 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| X-MEN | 129 | 20 | 109 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| Face HE | | | | | | | | | | |
| MINIONS | 180 | 5 | 119 | 15 | 41 | 0.11 | 0.11 | 0.25 | 0.11 | 0.69 |
| TORTUES_NINJAS | 282 | 21 | 178 | 36 | 47 | 0.31 | 0.17 | 0.37 | 0.17 | 0.71 |
| TOTALLY_SPIES | 166 | 11 | 108 | 15 | 32 | 0.26 | 0.12 | 0.42 | 0.12 | 0.72 |
| AVENGERS | 210 | 1 | 150 | 7 | 52 | 0.02 | 0.04 | 0.13 | 0.04 | 0.72 |
| INDESTRUCTIBLES | 258 | 9 | 185 | 9 | 55 | 0.14 | 0.05 | 0.50 | 0.05 | 0.75 |
| X-MEN | 129 | 15 | 85 | 11 | 18 | 0.45 | 0.11 | 0.58 | 0.11 | 0.78 |
| TEEN_TITANS | 107 | 6 | 77 | 6 | 18 | 0.25 | 0.07 | 0.50 | 0.07 | 0.78 |
| SUPER_NANAS | 135 | 12 | 95 | 6 | 22 | 0.35 | 0.06 | 0.67 | 0.06 | 0.79 |
| LES_4_FANTASTIQUES | 213 | 23 | 153 | 6 | 31 | 0.43 | 0.04 | 0.79 | 0.04 | 0.83 |
| POKEMON | 106 | 17 | 73 | 9 | 7 | 0.71 | 0.11 | 0.65 | 0.11 | 0.85 |
| POWER_RANGERS | 197 | 25 | 143 | 5 | 24 | 0.51 | 0.03 | 0.83 | 0.03 | 0.85 |
| Face HI | | | | | | | | | | |
| TORTUES_NINJAS | 282 | 11 | 166 | 65 | 40 | 0.22 | 0.28 | 0.14 | 0.28 | 0.63 |
| MINIONS | 180 | 12 | 128 | 19 | 21 | 0.36 | 0.13 | 0.39 | 0.13 | 0.78 |
| TOTALLY_SPIES | 166 | 7 | 123 | 14 | 22 | 0.24 | 0.10 | 0.33 | 0.10 | 0.78 |
| TEEN_TITANS | 107 | 3 | 81 | 12 | 11 | 0.21 | 0.13 | 0.20 | 0.13 | 0.79 |
| X-MEN | 129 | 8 | 94 | 11 | 16 | 0.33 | 0.10 | 0.42 | 0.10 | 0.79 |
| AVENGERS | 210 | 6 | 161 | 12 | 31 | 0.16 | 0.07 | 0.33 | 0.07 | 0.80 |
| SUPER_NANAS | 135 | 5 | 103 | 7 | 20 | 0.20 | 0.06 | 0.42 | 0.06 | 0.80 |
| INDESTRUCTIBLES | 258 | 13 | 194 | 15 | 36 | 0.27 | 0.07 | 0.46 | 0.07 | 0.80 |
| POKEMON | 106 | 5 | 82 | 3 | 16 | 0.24 | 0.04 | 0.63 | 0.04 | 0.82 |
| POWER_RANGERS | 197 | 17 | 155 | 7 | 18 | 0.49 | 0.04 | 0.71 | 0.04 | 0.87 |

| Face HN | | | | | | | | | | |
|---|-----|----|-----|----|----|------|------|------|------|------|
| AVENGERS | 210 | 36 | 99 | 71 | 4 | 0.90 | 0.42 | 0.34 | 0.42 | 0.64 |
| SUPER_NANAS | 135 | 17 | 74 | 36 | 8 | 0.68 | 0.33 | 0.32 | 0.33 | 0.67 |
| MINIONS | 180 | 19 | 103 | 43 | 15 | 0.56 | 0.29 | 0.31 | 0.29 | 0.68 |
| INDESTRUCTIBLES | 258 | 42 | 134 | 75 | 7 | 0.86 | 0.36 | 0.36 | 0.36 | 0.68 |
| X-MEN | 129 | 12 | 80 | 24 | 13 | 0.48 | 0.23 | 0.33 | 0.23 | 0.71 |
| TEEN_TITANS | 107 | 12 | 65 | 21 | 9 | 0.57 | 0.24 | 0.36 | 0.24 | 0.72 |
| TOTALLY_SPIES | 166 | 23 | 98 | 36 | 9 | 0.72 | 0.27 | 0.39 | 0.27 | 0.73 |
| LES_4_FANTASTIQUES | 213 | 29 | 133 | 40 | 11 | 0.73 | 0.23 | 0.42 | 0.23 | 0.76 |
| TORTUES_NINJAS | 282 | 17 | 206 | 23 | 36 | 0.32 | 0.10 | 0.43 | 0.10 | 0.79 |
| POWER_RANGERS | 197 | 32 | 124 | 35 | 6 | 0.84 | 0.22 | 0.48 | 0.22 | 0.79 |
| POKEMON | 106 | 14 | 71 | 16 | 5 | 0.74 | 0.18 | 0.47 | 0.18 | 0.80 |
| Neer | | | | | | | | | | |
| MINIONS | 180 | 37 | 135 | 6 | 2 | 0.95 | 0.04 | 0.86 | 0.04 | 0.96 |
| SUPER_NANAS | 135 | 29 | 102 | 3 | 1 | 0.97 | 0.03 | 0.91 | 0.03 | 0.97 |
| POWER_RANGERS | 197 | 39 | 155 | 2 | 1 | 0.98 | 0.01 | 0.95 | 0.01 | 0.98 |
| TORTUES_NINJAS | 282 | 54 | 224 | 2 | 2 | 0.96 | 0.01 | 0.96 | 0.01 | 0.99 |
| LES_4_FANTASTIQUES | 213 | 42 | 168 | 2 | 1 | 0.98 | 0.01 | 0.95 | 0.01 | 0.99 |
| TOTALLY_SPIES | 166 | 36 | 128 | 1 | 1 | 0.97 | 0.01 | 0.97 | 0.01 | 0.99 |
| INDESTRUCTIBLES | 258 | 47 | 208 | 2 | 1 | 0.98 | 0.01 | 0.96 | 0.01 | 0.99 |
| TEEN_TITANS | 107 | 28 | 78 | 0 | 1 | 0.97 | 0.00 | 1.00 | 0.00 | 0.99 |
| X-MEN | 129 | 27 | 101 | 1 | 0 | 1.00 | 0.01 | 0.96 | 0.01 | 0.99 |
| AVENGERS | 210 | 42 | 167 | 0 | 1 | 0.98 | 0.00 | 1.00 | 0.00 | 1.00 |
| POKEMON | 106 | 25 | 81 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| Phase 3 – Incidences d'épaule (avec fiche de rappel) | | | | | | | | | | |
| Axiale | | | | | | | | | | |
| LES_4_FANTASTIQUES | 9 | 0 | 4 | 3 | 2 | 0.00 | 0.43 | 0.00 | 0.43 | 0.44 |
| SUPER_NANAS | 98 | 15 | 80 | 0 | 3 | 0.83 | 0.00 | 1.00 | 0.00 | 0.97 |
| MINIONS | 180 | 24 | 152 | 0 | 4 | 0.86 | 0.00 | 1.00 | 0.00 | 0.98 |
| TEEN_TITANS | 81 | 15 | 65 | 1 | 0 | 1.00 | 0.02 | 0.94 | 0.02 | 0.99 |
| TORTUES_NINJAS | 124 | 19 | 104 | 0 | 1 | 0.95 | 0.00 | 1.00 | 0.00 | 0.99 |
| AVENGERS | 147 | 22 | 124 | 0 | 1 | 0.96 | 0.00 | 1.00 | 0.00 | 0.99 |
| INDESTRUCTIBLES | 156 | 24 | 132 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| POKEMON | 61 | 8 | 53 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| POWER_RANGERS | 160 | 24 | 136 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| TOTALLY_SPIES | 124 | 20 | 104 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| X-MEN | 121 | 19 | 102 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |

| Face HE | | | | | | | | | | |
|--------------------|-----|----|-----|----|----|------|------|------|------|------|
| MINIONS | 180 | 5 | 119 | 15 | 41 | 0.11 | 0.11 | 0.25 | 0.11 | 0.69 |
| LES_4_FANTASTIQUES | 9 | 1 | 6 | 1 | 1 | 0.50 | 0.14 | 0.50 | 0.14 | 0.78 |
| TOTALLY_SPIES | 124 | 12 | 85 | 9 | 18 | 0.40 | 0.10 | 0.57 | 0.10 | 0.78 |
| TORTUES_NINJAS | 124 | 17 | 82 | 12 | 13 | 0.57 | 0.13 | 0.59 | 0.13 | 0.80 |
| X-MEN | 121 | 14 | 84 | 8 | 15 | 0.48 | 0.09 | 0.64 | 0.09 | 0.81 |
| POKEMON | 61 | 8 | 42 | 5 | 6 | 0.57 | 0.11 | 0.62 | 0.11 | 0.82 |
| INDESTRUCTIBLES | 156 | 24 | 105 | 10 | 17 | 0.59 | 0.09 | 0.71 | 0.09 | 0.83 |
| POWER_RANGERS | 160 | 22 | 112 | 6 | 20 | 0.52 | 0.05 | 0.79 | 0.05 | 0.84 |
| AVENGERS | 147 | 18 | 107 | 3 | 19 | 0.49 | 0.03 | 0.86 | 0.03 | 0.85 |
| TEEN_TITANS | 81 | 7 | 62 | 3 | 9 | 0.44 | 0.05 | 0.70 | 0.05 | 0.85 |
| SUPER_NANAS | 98 | 14 | 72 | 5 | 7 | 0.67 | 0.06 | 0.74 | 0.06 | 0.88 |
| Face HI | | | | | | | | | | |
| LES_4_FANTASTIQUES | 9 | 0 | 7 | 1 | 1 | 0.00 | 0.13 | 0.00 | 0.13 | 0.78 |
| MINIONS | 180 | 12 | 128 | 19 | 21 | 0.36 | 0.13 | 0.39 | 0.13 | 0.78 |
| TOTALLY_SPIES | 124 | 13 | 87 | 13 | 11 | 0.54 | 0.13 | 0.50 | 0.13 | 0.81 |
| AVENGERS | 147 | 3 | 118 | 3 | 23 | 0.12 | 0.02 | 0.50 | 0.02 | 0.82 |
| X-MEN | 121 | 10 | 90 | 7 | 14 | 0.42 | 0.07 | 0.59 | 0.07 | 0.83 |
| TORTUES_NINJAS | 124 | 8 | 96 | 4 | 16 | 0.33 | 0.04 | 0.67 | 0.04 | 0.84 |
| POKEMON | 61 | 4 | 48 | 1 | 8 | 0.33 | 0.02 | 0.80 | 0.02 | 0.85 |
| INDESTRUCTIBLES | 156 | 6 | 127 | 1 | 22 | 0.21 | 0.01 | 0.86 | 0.01 | 0.85 |
| SUPER_NANAS | 98 | 6 | 78 | 3 | 11 | 0.35 | 0.04 | 0.67 | 0.04 | 0.86 |
| POWER_RANGERS | 160 | 14 | 127 | 5 | 14 | 0.50 | 0.04 | 0.74 | 0.04 | 0.88 |
| TEEN_TITANS | 81 | 6 | 67 | 2 | 6 | 0.50 | 0.03 | 0.75 | 0.03 | 0.90 |
| Face HN | | | | | | | | | | |
| LES_4_FANTASTIQUES | 9 | 0 | 5 | 1 | 3 | 0.00 | 0.17 | 0.00 | 0.17 | 0.56 |
| MINIONS | 180 | 19 | 103 | 43 | 15 | 0.56 | 0.29 | 0.31 | 0.29 | 0.68 |
| INDESTRUCTIBLES | 156 | 20 | 91 | 36 | 9 | 0.69 | 0.28 | 0.36 | 0.28 | 0.71 |
| AVENGERS | 147 | 28 | 78 | 40 | 1 | 0.97 | 0.34 | 0.41 | 0.34 | 0.72 |
| TOTALLY_SPIES | 124 | 13 | 80 | 19 | 12 | 0.52 | 0.19 | 0.41 | 0.19 | 0.75 |
| X-MEN | 121 | 16 | 75 | 22 | 8 | 0.67 | 0.23 | 0.42 | 0.23 | 0.75 |
| TORTUES_NINJAS | 124 | 17 | 77 | 22 | 8 | 0.68 | 0.22 | 0.44 | 0.22 | 0.76 |
| POKEMON | 61 | 7 | 40 | 11 | 3 | 0.70 | 0.22 | 0.39 | 0.22 | 0.77 |
| POWER_RANGERS | 160 | 26 | 101 | 28 | 5 | 0.84 | 0.22 | 0.48 | 0.22 | 0.79 |
| TEEN_TITANS | 81 | 12 | 53 | 13 | 3 | 0.80 | 0.20 | 0.48 | 0.20 | 0.80 |
| SUPER_NANAS | 98 | 16 | 63 | 16 | 3 | 0.84 | 0.20 | 0.50 | 0.20 | 0.81 |
| Neer | | | | | | | | | | |
| LES_4_FANTASTIQUES | 9 | 0 | 6 | 2 | 1 | 0.00 | 0.25 | 0.00 | 0.25 | 0.67 |
| MINIONS | 180 | 37 | 135 | 6 | 2 | 0.95 | 0.04 | 0.86 | 0.04 | 0.96 |
| AVENGERS | 147 | 29 | 114 | 1 | 3 | 0.91 | 0.01 | 0.97 | 0.01 | 0.97 |
| INDESTRUCTIBLES | 156 | 33 | 120 | 2 | 1 | 0.97 | 0.02 | 0.94 | 0.02 | 0.98 |
| POWER_RANGERS | 160 | 34 | 124 | 1 | 1 | 0.97 | 0.01 | 0.97 | 0.01 | 0.99 |
| TEEN_TITANS | 81 | 22 | 58 | 0 | 1 | 0.96 | 0.00 | 1.00 | 0.00 | 0.99 |
| POKEMON | 61 | 17 | 44 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| SUPER_NANAS | 98 | 23 | 75 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| TORTUES_NINJAS | 124 | 25 | 99 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| TOTALLY_SPIES | 124 | 25 | 99 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| X-MEN | 121 | 25 | 96 | 0 | 0 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |

XI.Phase 0 – Images de félins

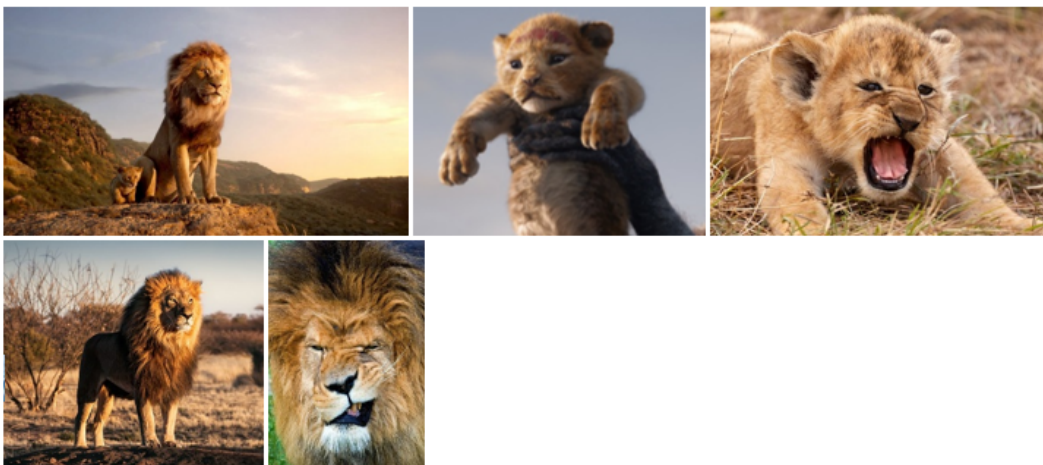
Chat



Guépard



Lion



Lynx



Tigre



Doubles classes



Images piégeuses (pas de classe)



XII.Phase 3 - Fiche de rappel (Rotations de l'humérus)

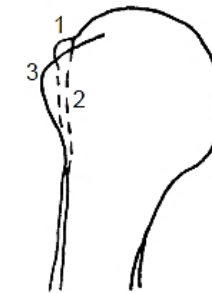
POSITIONNEMENT DU PATIENT

FICHE N° : B 6

ORGANE : EPAULE

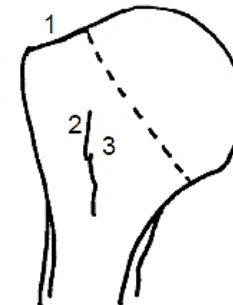
* Rotation externe

- . Le col anatomique est très oblique, en bas et en dedans, par rapport à l'axe de la diaphyse humérale
- . la tête humérale est très asymétrique par rapport à l'axe de la diaphyse humérale,
- . la coulisse bicipitale est rejetée à l'extérieur, par rapport à l'axe de la diaphyse humérale,
- . la partie antérieure de la face supérieure du trochiter (point d'insertion de la partie antérieure du sus-épineux) est visible.



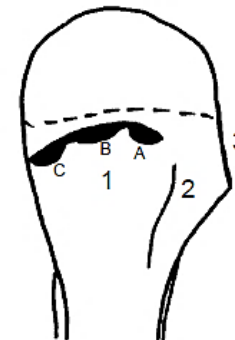
* Face (neutre)

- . Le col anatomique a une obliquité moyenne, par rapport à l'axe de la diaphyse humérale,
- . la face supérieure du trochiter est bien dégagée,
- . la coulisse bicipitale est paramédiane externe, par rapport à l'axe de la diaphyse humérale,
- . le trochin est vu de face.



* Rotation interne

- . Le col anatomique est presque perpendiculaire par rapport à l'axe de la diaphyse humérale,
- . la tête humérale coiffe la diaphyse humérale de façon symétrique,
- . le trochiter est vu à plat par sa face externe,
- . la coulisse bicipitale est rejetée en dedans, par rapport à l'axe de la diaphyse humérale,
- . le trochin est vu de profil, continuant le bord interne de la métaphyse humérale, (sous-scapulaire)



1. Trochiter

2. La coulisse bicipitale

3. Trochin

A. Point d'insertion du sus-épineux (facette supérieure)

B. Point d'insertion du sous-épineux (facette moyenne)

C. Point d'insertion du petit rond (facette postérieure).