Année 2021/2022

Thèse

Pour le

DOCTORAT EN MEDECINE

Diplôme d'État

par

Rayane BENHENNEDA

Né le 29 décembre 1993 à Colombes (92)

TITRE

L'intelligence artificielle peut-elle aider à la décision en arthroscopie ?

Partie 1 : Analyse de concordance inter-observateur pour l'évaluation arthroscopique diagnostique du tendon du long biceps dans les petites ruptures de la coiffe des rotateurs.

Partie 2 : Le modèle IA-RTRO : aide à la décision diagnostique pour le statut du tendon du long biceps dans les petites ruptures de la coiffe des rotateurs.

Présentée et soutenue publiquement le **27 octobre 2022** devant un jury composé de :

Président du Jury :

Professeur Luc FAVARD, Chirurgie orthopédique, Faculté de médecine - Tours

Membres du Jury :

Professeur Phillipe ROSSET, Chirurgie orthopédique, Faculté de médecine – Tours

Monsieur Thierry BROUARD, Département d'informatique, École Polytechnique - Tours

Professeur Julien BERHOUET, Chirurgie orthopédique, Faculté de médecine - Tours

Docteur Louis-Romée LE NAIL, Chirurgie orthopédique, PH, CHU Trousseau – Tours

Directeur de thèse : Professeur Julien BERHOUET

UNIVERSITE DE TOURS
FACULTE DE MEDECINE DE TOURS

DOYEN
Pr Patrice DIOT

VICE-DOYEN
Pr Henri MARRET

ASSESSEURS
Pr Denis ANGOULVANT, Pédagogie
Pr Mathias BUCHLER, Relations internationales
Pr Theodora BEJAN-ANGOULVANT, Moyens – relations avec l'Université
Pr Clarisse DIBAO-DINA, Médecine générale
Pr François MAILLOT, Formation Médicale Continue
Pr Patrick VOURC'H, Recherche

RESPONSABLE ADMINISTRATIVE
Mme Carole ACCOLAS

****

DOYENS HONORAIRES
Pr Emile ARON (†) – 1962-1966
Directeur de l'Ecole de Médecine - 1947-1962
Pr Georges DESBUQUOIS (†) - 1966-1972
Pr André GOUAZE (†) - 1972-1994
Pr Jean-Claude ROLLAND – 1994-2004
Pr Dominique PERROTIN – 2004-2014

PROFESSEURS EMERITES
Pr Daniel ALISON
Pr Philippe COLOMBAT
Pr Etienne DANQUECHIN-DORVAL
Pr Pascal DUMONT
Pr Bernard FOUQUET
Pr Yves GRUEL
Pr Gérard LORETTE
Pr Dominique PERROTIN
Pr Philippe ROSSET

PROFESSEURS HONORAIRE
P. ANTHONIOZ – P. ARBEILLE – A. AUDURIER – A. AUTRET – P. BAGROS – C. BARTHELEMY – J.L. BAULIEU – C. BERGER – JC. BESNARD – P. BEUTTER – C. BONNARD – P. BONNET – P. BOUGNOUX – P. BURDIN – L. CASTELLANI – J. CHANDENIER – A. CHANTEPIE – B. CHARBONNIER – P. CHOUTET – T. CONSTANS – C. COUET – L. DE LA LANDE DE CALAN – J.P. FAUCHIER – F. FETISSOF – J. FUSCIARDI – P. GAILLARD – G. GINIES – D. GOGA – A. GOUDEAU – J.L. GUILMOT – O. HAILLOT – N. HUTEN – M. JAN – J.P. LAMAGNERE – F. LAMISSE – Y. LANSON – O. LE FLOCH – Y. LEBRANCHU – E. LECA – P. LECOMTE – AM. LEHR-DRYLEWICZ – E. LEMARIE – G. LEROY – M. MARCHAND – C. MAURAGE – C. MERCIER – J. MOLINE – C. MORAINE – J.P. MUH – J. MURAT – H. NIVET – L. POURCELOT – R. QUENTIN – P. RAYNAUD – D. RICHARD-LENOBLE – A. ROBIER – J.C. ROLLAND – D. ROYERE - A. SAINDELLE – E. SALIBA – J.J. SANTINI – D. SAUVAGE – D. SIRINELLI – J. WEILL

RESUME

L'intelligence artificielle peut-elle aider à la décision en arthroscopie ?

Partie 1 : Analyse de Concordance Inter-Observateur pour l'Évaluation Arthroscopique Diagnostique du Tendon du Long Biceps Dans Les Petites Ruptures de la Coiffe des Rotateurs

Introduction : Les lésions du tendon du long biceps (TLB) sont fréquentes et leur diagnostic clinique ou par imagerie reste imprécis. L'arthroscopie est considérée comme le gold-standard pour l'analyse diagnostique du TLB, mais celle-ci reste parfois difficile. Sa fiabilité et sa reproductibilité n'ont par ailleurs pas été évaluées jusqu'à présent. L'intelligence artificielle (IA) pourrait apporter une aide pour l'analyse arthroscopique du TLB. L'objectif principal de cette étude était d'évaluer la concordance inter-observateur pour l'analyse spécifique du TLB, selon un protocole vidéo arthroscopique précis. L'objectif secondaire était de définir une base de données vidéos, dite « vérité terrain », destinée à créer et à entraîner une IA pour l'analyse du TLB.

Hypothèse : L'hypothèse était que la concordance d'analyse inter-observateur était suffisamment forte pour utiliser les vidéos comme « vérité terrain » d'une solution d'IA de diagnostic arthroscopique du TLB.

Matériels et méthode : Cent quatre-vingt-dix-neuf vidéos arthroscopiques protocolisées pour l'exploration du TLB ont été évaluées par 3 observateurs indépendants. Chacun devait caractériser le caractère sain ou pathologique du tendon, en précisant le type de lésion, dite « intrinsèque » : rupture partielle, hypertrophie avec sablier, instabilité, fissuration, désinsertion

(SLAP 2), ou « extrinsèque » : chondral print et poulie pathologique sans instabilité. Les niveaux de concordance inter-observateur ont été mesurées au moyen du coefficient Kappa (K) de Cohen et du Kappa Accuracy.

Résultats : La force de concordance était modérée à forte selon les observateurs (Kappa 0,537 à 0,701 et KappaAcc de 86 à 92%), pour déterminer le caractère sain ou pathologique du TLB. Lorsque le tendon était pathologique, la force de concordance était modérée à forte lorsqu'il s'agissait d'une rupture partielle (Kappa 0,494 à 0,708 et KappaAcc de 85 à 92%), d'une fissuration (Kappa -0,498 à 0,698 et Kappa Acc de 36 à 93%) ou d'une désinsertion (0,543 à 0,882 et Kappa Acc de 90 à 97%). Elle était faible pour une lésion d'instabilité (Kappa 0,039 à 0,248 et KappaAcc de 36 à 88%).

Conclusion : L'analyse du TLB, selon un protocole vidéo arthroscopique précis, était hautement concordante pour le diagnostic de son caractère sain ou pathologique. Le taux de concordance diminuait cependant pour le diagnostic de lésions tendineuses rares ou dynamiques. Une IA qui serait entraînée par une analyse humaine serait donc confrontée à ces mêmes difficultés si l'on se limitait à une analyse arthroscopique seule. L'intégration de données cliniques et para-cliniques est nécessaire à l'amélioration du diagnostic arthroscopique des lésions du TLB. Elle semble aussi être un prérequis indispensable à la constitution d'une base de données dite de « vérité terrain » pour la construction d'une solution d'IA performante.

Mots-clés : arthroscopie d'épaule ; intelligence artificielle ; analyse inter-observateur ; long biceps ; coiffe des rotateurs

Type d'étude : série prospective inter-observateur

Niveau de preuve : III

RESUME

L'intelligence artificielle peut-elle aider à la décision en arthroscopie ?

Partie 2 : Le modèle IA-RTRO : aide à la décision diagnostique pour le statut du tendon du long biceps dans les ruptures distales du tendon supra-épineux

Introduction : Les applications possibles de l'intelligence artificielle (IA) en chirurgie orthopédique sont prometteuses. Le deep learning est utilisable en chirurgie arthroscopique grâce au signal vidéo exploitable par computer vision. La gestion per-opératoire du tendon du long biceps (TLB) est l'objet d'une controverse ancienne. L'objectif principal de cette étude était de modéliser une IA diagnostique capable de statuer sur le caractère sain ou pathologique du TLB sur des images arthroscopiques. L'objectif secondaire était de créer un deuxième modèle IA diagnostique en fonction d'images arthroscopiques et des données médicales de chaque patient, cliniques et imagerie, pour statuer sur le caractère sain ou pathologique du TLB.

Hypothèse : L'hypothèse de l'étude était qu'il était possible de construire un modèle IA à partir d'images opératoires d'arthroscopie d'aide au diagnostic du caractère sain ou pathologique du TLB.

Matériels et Méthodes : Les données prospectives cliniques et d'imagerie de 199 patients ont été collectées et associées aux images issues d'une analyse vidéo arthroscopique protocolisée validée, dite « vérité terrain », faite par le chirurgien opérateur. Un modèle basé sur un réseau neuronal convolutif (RNC) modélisé via un *transfer learning* sur le modèle Inception V3 a été construit pour l'analyse des images arthroscopiques. Ce modèle a ensuite été couplé à un

Perceptron MultiCouche (*MultiLayer Perceptron, MLP*) intégrant les données cliniques et imageries. Chaque modèle a été entraîné et testé selon un apprentissage supervisé.

Résultats : La précision du RNC sur le diagnostic du statut sain ou pathologique du TLB était de 93,7% en apprentissage et 80,66% en généralisation. Couplé aux données cliniques de chaque patient, la précision du modèle assemblant le RNC et un MLP étaient respectivement de 77% et 58% en apprentissage et en généralisation.

Conclusion : Le modèle IA construit à partir d'un RNC parvient à déterminer le caractère sain ou pathologique du TLB avec un taux de précision de 80,66%. Ce modèle n'est pas suffisamment pertinent pour une application en pratique clinique. Une augmentation des données d'entrée pour limiter l'over-fitting, et l'automatisation de la phase de détection par un *Mask-R-CNN* sont des voies d'amélioration du modèle. Cette étude est la première à évaluer la capacité d'une IA à analyser des images arthroscopiques, et ses résultats doivent être confirmés par d'autres études sur le sujet.

Mots-clés : arthroscopie ; intelligence artificielle ; réseau neuronal convolutif ; tendon du long biceps ; apprentissage profond

Type d'étude : Étude diagnostique

Niveau de preuve : III

ABSTRACT

Can Artificial Intelligence Help Decision Making in Arthroscopy ?

Part 1 : Inter-Observer Analysis for the Arthroscopic Assessment of the Long Head of the Biceps in Small Rotator Cuff Tears

Introduction : The lesions of the long head of the biceps (LHB) are frequent while their clinical or imaging diagnosis remains imprecise. Arthroscopy is considered to be the gold standard for the LHB analysis. However its reliability and reproducibility have not been evaluated so far. Artificial intelligence (AI) could help during arthroscopy for analyzing the LHB. The main objective of the current study was to assess the inter-observer analysis concordance for the LHB diagnosis, according to a precise arthroscopic video protocol. The secondary objective was to define a video database, called "ground truth", in aim to create and train an AI solution for analyzing the LHB.

Hypothesis : The inter-observer analysis concordance was strong enough to use the videos as the “ground truth” for an AI solution for arthroscopic diagnosis of the LHB.

Materials & Method : One hundred and ninety-nine protocolized arthroscopic videos for the exploration of the LHB were evaluated by 3 independent observers. Each had to characterize the healthy or pathological nature of the tendon, specifying the type of lesion, called "intrinsic": partial rupture, hypertrophy with hourglass figure, instability, cracking, disinsertion (SLAP 2), or "extrinsic": chondral print and pathological pulley without instability. Inter-observer agreement levels were measured using Cohen's Kappa (K) coefficient and Kappa Accuracy.

Results : The strength of concordance was moderate to strong according to the observers (Kappa 0,537 to 0,701 and KappaAcc from 86 to 92%), to determine the healthy or pathological character of the LHB. In case of pathological aspect, the concordance force was moderate to strong when it was a partial rupture (Kappa 0,494 to 0,708 and KappaAcc from 85 to 92%), a fissuration (Kappa -0,498 to 0,698 and KappaAcc from 36 to 93%), or a disinsertion (Kappa 0,543 to 0,882 and KappaAcc from 90 to 97%),, and low for an unstable lesion (Kappa 0,039 to 0,248 and KappaAcc from 36 to 88%).

Conclusion : Analyzing the LHB according to a precise arthroscopic video protocol was highly concordant for the diagnosis of its healthy or pathological status. However, the concordance rate decreased for the diagnosis of rare or dynamic tendon lesions. An AI that would be trained by a human analysis would therefore face the same difficulties if it was limited to a solely arthroscopic analysis. Additional clinical and para-clinical datas are necessary to improve the arthroscopic diagnosis of LHB lesions. It also seems to be an essential prerequisite for the constitution of a so-called “ground truth” database for the construction of a high-performance AI solution.

Keywords: shoulder arthroscopy; artificial intelligence ; inter-observer analysis; long biceps; rotator cuff

Study type : inter-observer prospective study

Level of evidence : III

ABSTRACT

Can Artificial Intelligence Help Decision Making in Arthroscopy ?

Part 2 : The IA-RTRO model : help decision making for the diagnosis of the long head of the biceps status in small rotator cuff tears

Introduction : Artificial Intelligence (AI) is a promising asset in orthopedic surgery. Deep learning can be used in arthroscopic surgery due to the video signal that can be used by computer vision. The intraoperative management of the long biceps tendon (LTB) is the subject of a long-standing controversy. The main objective of this study was to model a diagnostic AI capable of deciding on the healthy or pathological character of LTB on an arthroscopic image. The secondary objective was to create a second diagnostic AI model based on arthroscopic images and the medical data of each patient, clinical and imaging, to decide on the healthy or pathological nature of the LTB.

Hypothesis : The hypothesis of the study was that it was possible to build an AI model from operative arthroscopic images to help diagnose the healthy or pathological nature of LTB.

Materials & Methods : The clicnal and imaging data of 199 patients were collected prospectively and associated with images from a validated protocolized arthroscopic video analysis, called “ground truth”, made by the operating surgeon. A model based on a convolutional neural network (CNN) modeled via transfer learning on the Inception V3 model was built for the analysis of arthroscopic images. This model was then coupled to a MultiLayer Perceptron (MLP) integrating clinical and imaging data. Each model has been trained and tested using supervised learning.

Results : The accuracy of the CNN on the diagnosis of the healthy or pathological status of the LTB was 93.7% in learning and 80.66% in generalization. Coupled with the clinical data of each patient, the accuracy of the model assembling the CNN and the MLP were respectively 77% and 58% in learning and in generalization.

Conclusion : The AI model built from a CNN manages to determine the healthy or pathological character of the LTB with an accuracy rate of 80.66%. This model is not sufficiently relevant for application in clinical practice. An increase in input data to limit overfitting, and the automation of the detection phase by a Mask-R-CNN are ways of improving the model. This study is the first to assess the ability of an AI to analyze arthroscopic images, and its results need to be confirmed by other studies on the subject.

Keywords: shoulder arthroscopy; artificial intelligence ; convolutional neural network; long tendon biceps; deep learning

Study type : diagnostic study

Level of evidence : III

LISTE DES ABREVIATIONS

1D : 1 Dimension

2D : 2 Dimensions

Acc : Accuracy

C : Chirurgien opérateur

CART : Classification And Regression Trees

CNN (RCN) : Convolutionnal Neural Network (Réseau Neuronal Convolutif)

DL : Deep Learning

eCRF : e-Case Report Form

IA : Intelligence artificielle

IRM : Imagerie par résonance magnétique

ML : Machine Learning

MLP : MultiLayer Perceptron

Nb Cas : Nombre de cas

O1 : Observateur 1 (junior)

O2 : Observateur 2 (sénior)

ROI : Region Of Interest

SFA : Société Francophone d'Arthroscopie

SLAP : Superior Labrum Anterior-Posterior tears

TLB : Tendon du long biceps

%Ag : Pourcentage de cas où les observateurs sont en accord

%Ch : Pourcentage de cas où les observateurs pourraient tomber en accord de façon aléatoire

SOMMAIRE

Glossaire **27**

Intelligence artificielle (IA) 27

Machine Learning (ML) 27

Deep Learning (DL), ou apprentissage profond 29

Computer vision (vision par ordinateur) 29

Bounding box 29

Réseau de neurones artificiels 30

Convolutionnal neural network (CNN), ou réseau neuronal convolutif 31

MultiLayer Perceptron (MLP) 33

Learning rate 33

Transfer Learning (TL) 33

Apprentissage supervisé 34

Under-fitting et Over-fitting 35

Introduction générale **36**

Partie 1 : Analyse de Concordance Inter-Observateur pour l'Evaluation Arthroscopique Diagnostique du Tendon du Long Biceps Dans Les Petites Ruptures de la Coiffe des Rotateurs **40**

Résumé **41**

Introduction **45**

Matériel et méthodes **47**

Patients 47

Vidéos arthroscopiques 48

Analyse vidéo inter-observateur du TLB 49

Analyse statistique 52

Résultats **54**

Analyse descriptive : occurrence des lésions du TLB selon le diagnostic du « chirurgien opérateur » (C) 54

Analyses des concordances inter-observateurs 54

Discussion **57**

Conclusion **62**

Tableaux **64**

Légendes des figures **67**

Partie 2 : Le Modèle IA-RTHRO. Aide à la Décision Diagnostique pour le Statut du Tendon du Long Biceps Dans Les Petites Ruptures de la Coiffe des Rotateurs **68**

Résumé **69**

Introduction **73**

Matériel et méthodes **75**

Base de données pour le modèle d'IA expérimental : la « vérité terrain » 75

Construction du modèle d'IA 77

Entraînement des modèles d'IA 79

Modèle n°1 : Méthode d'apprentissage du CNN 79

Modèle n°2 : Méthode d'apprentissage du MLP 82

Modèle n°3 : Association des deux modèles IA 83

Résultats **87**

Précision du modèle n°1 *(Figure 8)* 87

Précision du modèle n°2 *(Figure 9)* 88

Précision du modèle n°3 *(Figure 10)* 89

Discussion **90**

Conclusion **95**

Tableaux **97**

Légendes **99**

Bibliographie **101**

TABLE DES FIGURES

Figure 1 : Images correspondant aux 4 zones d'analyse du TLB : 1) insertion à la glène, 2) portion intra-articulaire proximale, 3) portion intra-articulaire distale, 4) gouttière bicipitale, avant (A, B, C et D) et après (E, F, G, H) l'étape de segmentation (bounding box en vert).

Figure 2 : Image avec bounding box sur une fissuration, pouvant être considérée comme un TLB sain ou pathologique selon la subjectivité de l'observateur.

Figure 3 : Image avec hypertrophie en sablier, pouvant être considérée comme un TLB sain ou pathologique selon la subjectivité de l'observateur.

Figure 4 : Architecture du réseau neuronal convolutif construit.

Figure 5 : Architecture du modèle construit, basé sur le CNN et le MLP. F = Fonction d'activation ReLU. TLB = Tendon du long biceps.

Figure 6 : Exemple de transformations subies par l'une des images de la base d'apprentissage.

Figure 7 : Résumé des 3 modèles d'IA conçus. IA = Intelligence Artificielle ; *CNN = Convolutional Neural Network ; MLP = MultiLayer Perceptron* ; TLB = Tendon du Long Biceps

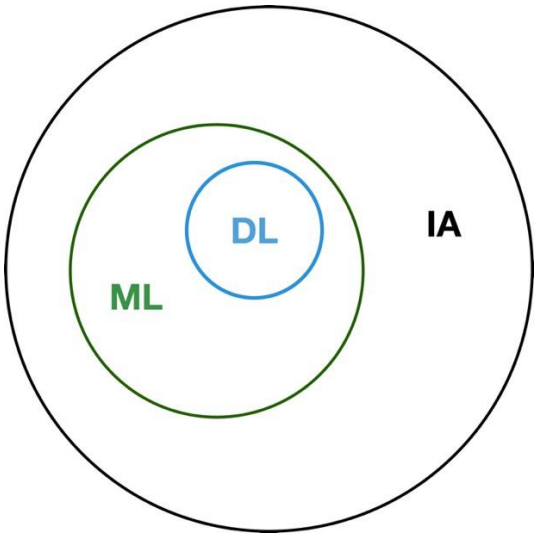
Figure 8 : Précision du modèle n°1 (*CNN*). En bleu, courbe d'apprentissage (training set) et en orange, courbe de test (testing set). *Model accuracy* : précision du modèle, *Accuracy* : précision, *epoch* : époque.

Figure 9 : Précision du modèle n°2 (*MLP*). En bleu, courbe d'apprentissage (training set) et en orange, courbe de test (testing set). *Model accuracy* : précision du modèle, *Accuracy* : précision, *epoch* : époque.

Figure 10 : Précision du modèle n°3 (assemblage des modèles 1 et 2). En bleu, courbe d'apprentissage (training set) et en orange, courbe de test (testing set). *Model accuracy* : précision du modèle, *Accuracy* : précision, *epoch* : époque.

Figure 11 : Segmentation automatique d'images par un Mask-R-CNN.

Glossaire



IA : Intelligence artificielle – ML : Machine Learning – DL : Deep Learning

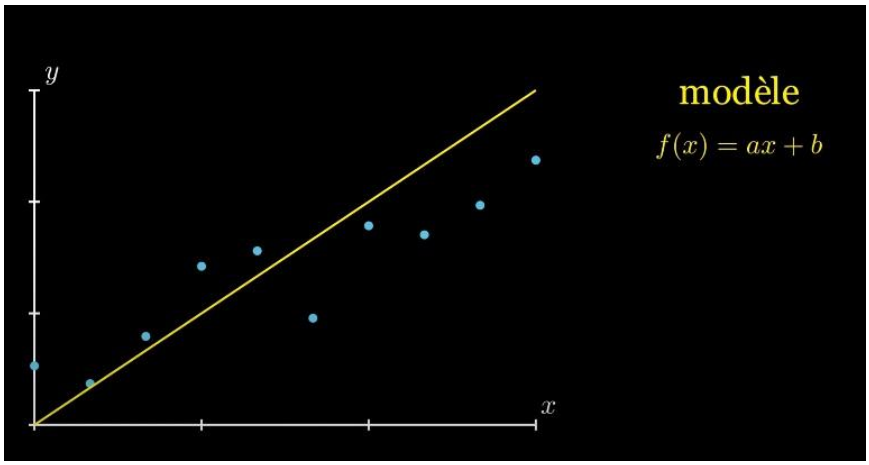
Intelligence artificielle (IA) :

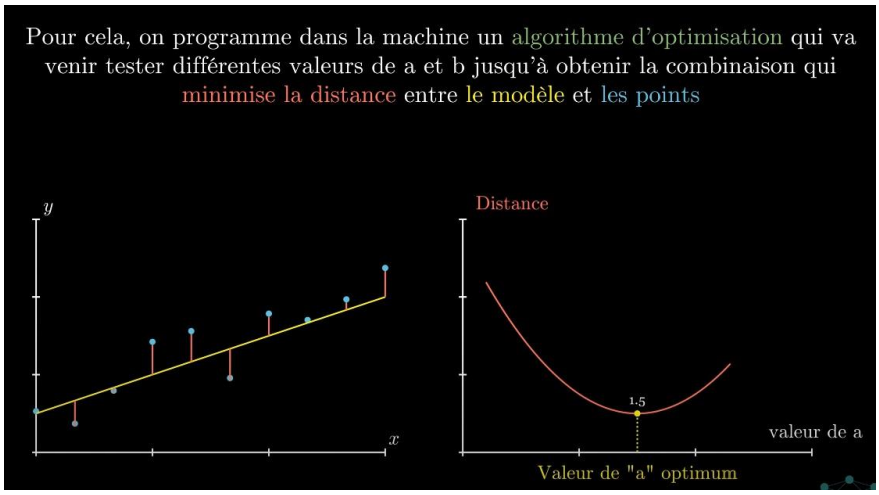
Capacité d'une machine à simuler l'intelligence humaine.

Machine Learning (ML) :

Domaine de l'IA consistant à programmer une machine pour que celle-ci apprenne à réaliser des tâches après en avoir étudié des exemples.

D'un point de vue mathématique, ces exemples sont des **données** que la machine utilise pour développer un **modèle** (par exemple un fonction f(x) = ax + b). L'objectif est de faire varier les paramètres a et b, afin d'obtenir le meilleur modèle possible : celui qui s'ajuste le mieux aux données.





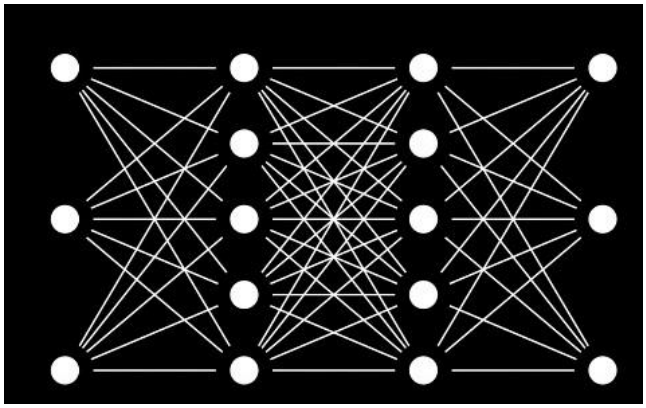
Au total, le ML revient à développer un modèle en se servant d'un algorithme d'optimisation pour minimiser les erreurs entre le modèle et nos données.

Plusieurs modèles existent : modèles linéaires, arbres de décision, support vector machines… chacun avec un algorithme d'optimisation ad hoc (respectivement : descente de gradients, algorithme CART, marge maximum…).

Deep Learning (DL), ou apprentissage profond :

Domaine du ML dans lequel le modèle utilisé est celui d'un réseau de neurones artificiels, dont le principe de fonctionnement est le même : données, modèle et algorithme d'optimisation.

Ici, le modèle est un réseau de fonction connectées les unes aux autres (= un réseau de neurones). Plus ces réseaux sont profonds (= plus il y a de fonctions à l'intérieur), plus la machine est capable d'accomplir des tâches complexes. D'où le terme d'apprentissage « profond ».



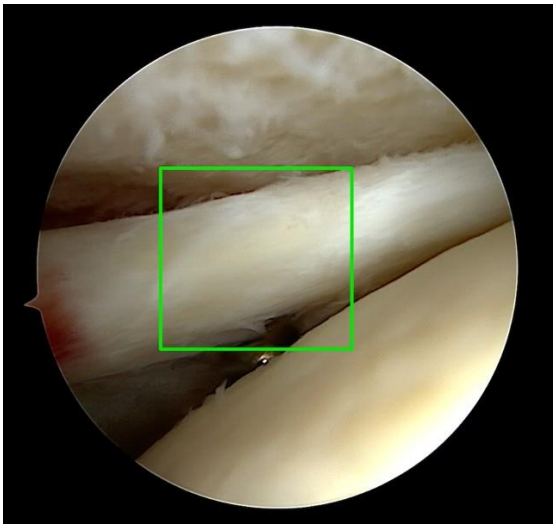
Réseau de neurones

Computer vision (vision par ordinateur) :

La computer vision est la capacité d'une machine à acquérir une compréhension de haut niveau à partir d'images ou de vidéos numériques.

Bounding box :

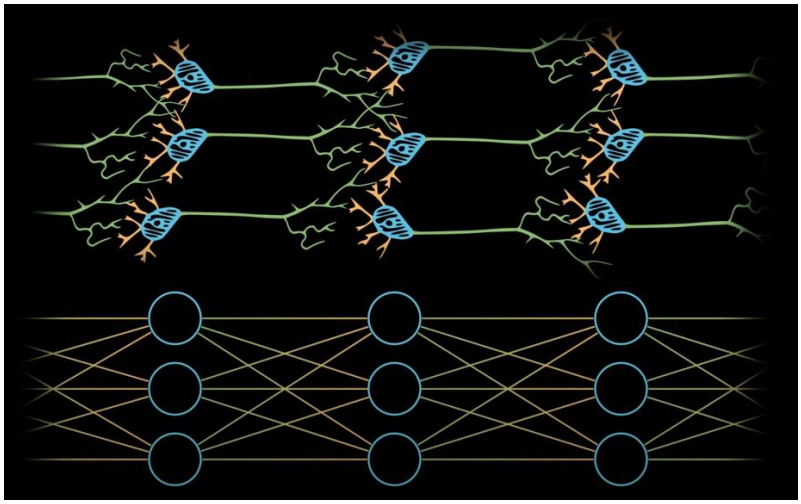
Une bounding box est un rectangle imaginaire qui sert de point de référence pour la détection d'objet sur une image.



Exemple de Bounding Box sur la portion intra-articulaire d'un tendon du long biceps sain

Réseau de neurones artificiels :

C'est un système dont la conception est inspirée du fonctionnement des neurones biologiques.



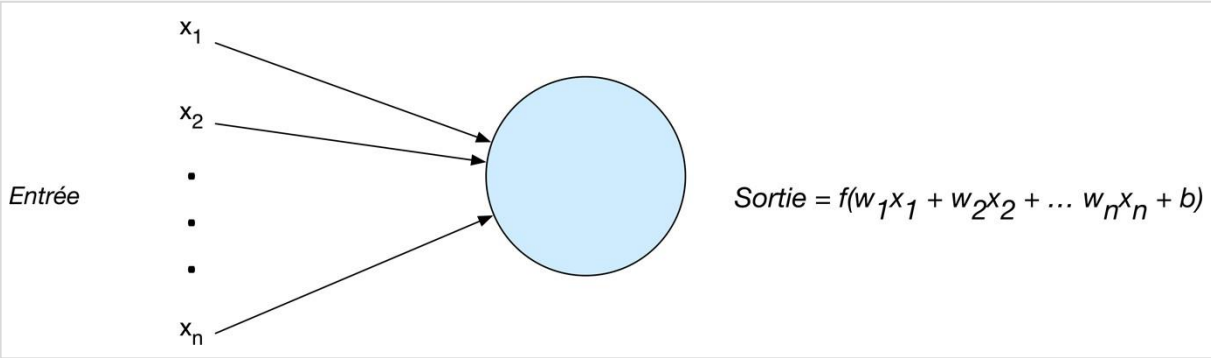
Neurones biologiques et neurones artificiels

Les réseaux de neurones offrent un moyen d'approximer une fonction inconnue lorsque les

données de sortie d'un grand nombre de données d'entrées sont connues, mais que la relation entre elles ne peut être déduite par les approches mathématiques conventionnelles.

L'unité fondamentale du réseau de neurones est le neurone artificiel, qui est utilisé pour transformer plusieurs données d'entrée (x_1, x_2, …, x_n) en une seule donnée de sortie. Les 3 composantes principales d'un neurone artificiel sont :

- Un ensemble de poids (w_1, w_2, …, w_n)
- Un biais (b)
- Une fonction d'activation (f)



Neurone artificiel

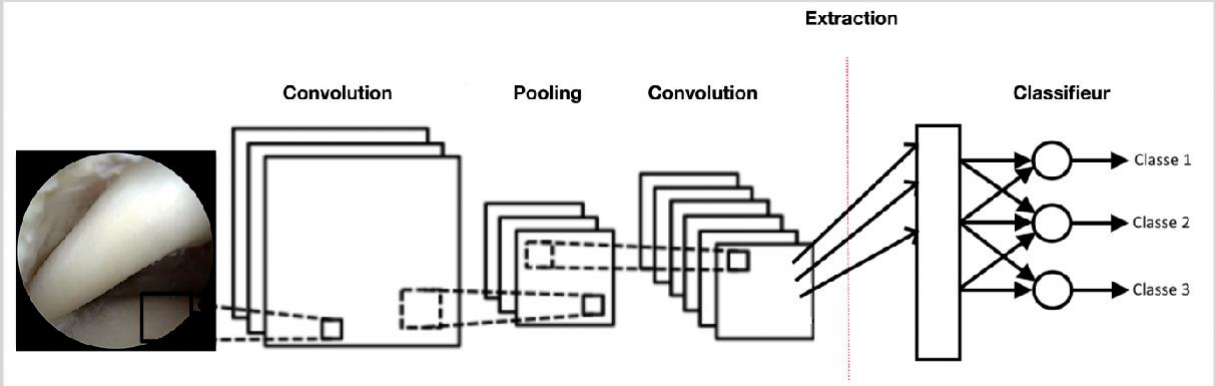
Convolutionnal neural network (CNN), ou réseau neuronal convolutif :

Un réseau neuronal convolutif est un type de réseau de neurones constitué de noyaux convolutifs conçu pour traiter des images en données d'entrée.

Ce modèle est composé de plusieurs couches :

- Une couche de convolution : elle prend en entrée les pixels de l'image, chaque neurone étant exposé à une partie de l'image. En fonction des poids de chaque pixel (réglages prédéfinis), la matrice de convolution donne une valeur qui active ou non la fonction d'activation du neurone.

- Une couche de pooling : par des transformations mathématiques, le pooling a pour but de réduire la taille des données d'entrée pour faciliter leur traitement.
- Une couche d'extraction : c'et la couche finale du modèle, qui présentera les données condensées au classifieur.

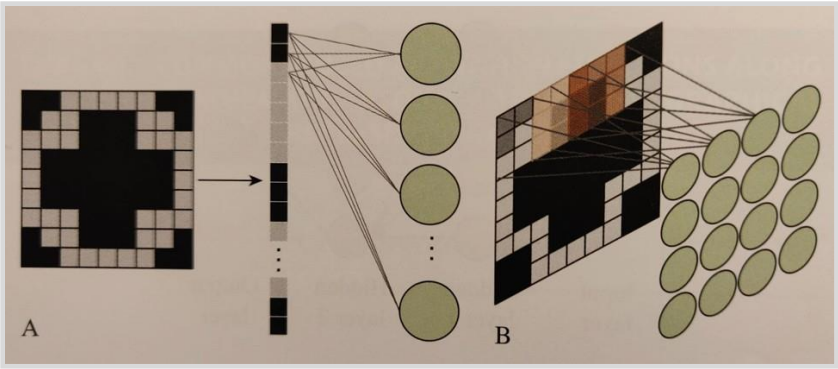


Réseau neuronal convolutif

L'utilisation d'un réseau de neurone en une dimension (1D) présente 2 problématiques pour le traitement d'image :

- La première est que les pixels avoisinants d'une image 2D originale peuvent apparaître à des distances irréalistes l'un de l'autre lors de la transformation vectorielle en 1D.
- La deuxième est qu'il nécessite un grand nombre de paramètre, de l'ordre du million, pour traiter une image de haute définition.

Ces deux limitations sont résolues par l'utilisation de réseaux neuronaux convolutifs.



Représentation de neurones et leurs entrées.

A = réseau de neurones connectés. B = réseau de neurones convolutifs

MultiLayer Perceptron (MLP) :

Le Perceptron MultiCouche (ou Multilayer Perceptron) est un type de réseau neuronal artificiel organisé en plusieurs couches. Il fonctionne de la même manière que les réseaux de neurones artificiels « classiques » définis précédemment.

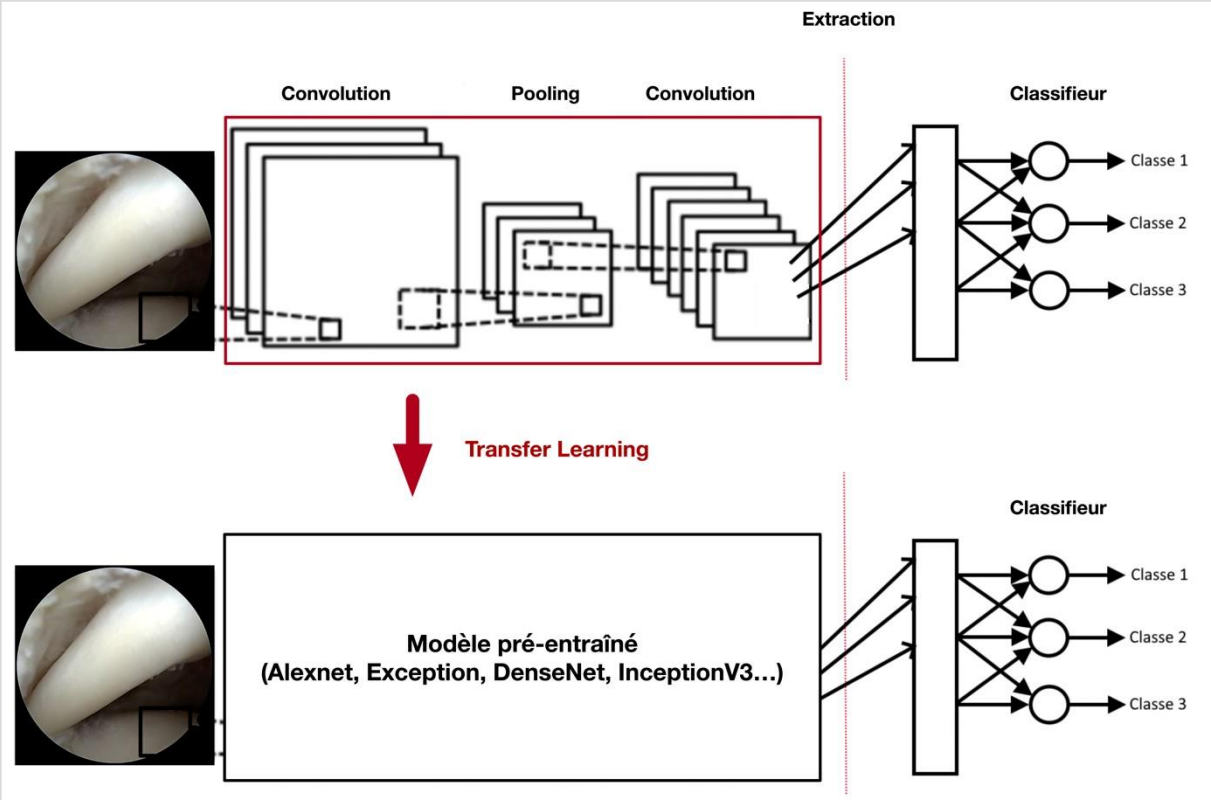
Learning rate :

Le taux d'apprentissage alpha (*learning rate*) est la vitesse à laquelle le modèle va converger vers le résultat souhaité lors de la descente de gradient. Cette descente de gradient consiste à ajuster les paramètres w et b de façon à minimiser les erreurs du modèles (c'est-à-dire à minimiser la fonction coût, ou *loss function*).

Transfer Learning (TL) :

C'est une stratégie qui consiste à utiliser un réseau de neurones pré-entraîné sur une tâche alternative mais similaire à celle que l'on souhaite modéliser.

En effet, entraîner une couche de convolution est une opération qui nécessite des temps de calcul élevés, des ressources importantes, et une base de données d'apprentissage conséquente (de l'ordre de la centaine de milliers, voire millions d'image). On utilise donc des réseaux de neurones pré-entraînés sur des bases de données massives (comme le modèle AlexNet entraîné sur la librairie libre ImageNet par exemple).



Transfer Learning

Apprentissage supervisé :

L'apprentissage supervisé est une méthode d'apprentissage d'un modèle IA en DL, par opposition à la méthode non supervisée.

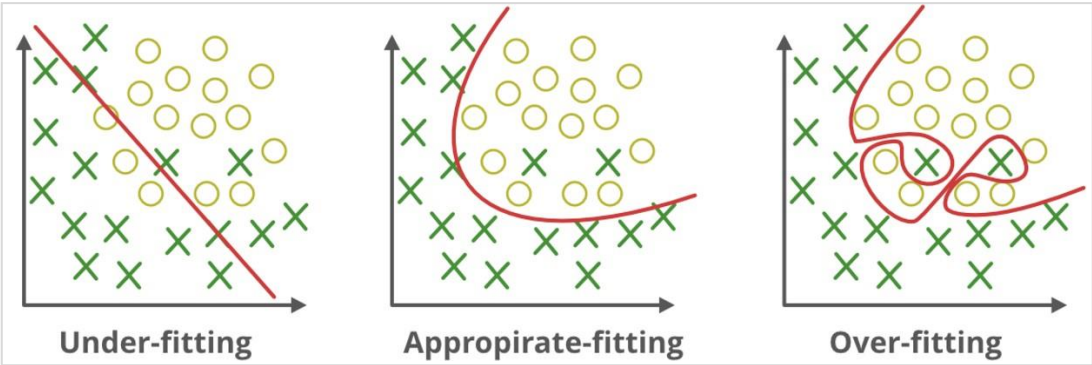
En apprentissage supervisé, les données d'entrée (des images par exemple) sont étiquetées, c'est-à-dire que l'on donne à la machine la valeur de sortie correspondant à ces données. Ainsi,

le modèle en disposant des données d'entrée et de sortie, aura pour but de trouver la fonction qui s'adapte le mieux à ces « contraintes ».

En apprentissage non supervisé, les données d'entrée ne sont pas étiquetées, et c'est au modèle de reconnaître de façon automatisée des « patterns » parmi les données (des contours, couleurs, structures… dans le cas d'images) sans aucun guidage humain.

Under-fitting et Over-fitting :

En cas de biais élevé et de variance faible, on expose le modèle à de l'under-fitting : la fonction est trop « simple » pour correspondre avec précision à la réalité. En cas de biais faible et de variance élevée, on risque de l'over-fitting : la fonction sera trop complexe pour être généralisée à des données non apprises (le modèle est capable d'apprendre par cœur, mais ne saura pas analyser des informations qu'il n'a jamais vues).



Schématisation des concepts d'under-et d'over-fitting

Introduction générale

L'intelligence artificielle (IA) désigne l'ensemble des théories et techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine. Elle représente un domaine des sciences informatiques en pleine expansion, dont les application actuelles et futures concernent tous les champs d'activité de la vie moderne.

La notion d'IA remonte à l'Antiquité. Dans la mythologie grecque, Talos était un automate créé par Héphaïstos, Dieu de la forge et des volcans, pour protéger la Crète. Les premiers neurones artificiels ont été élaborés en 1943, par Warren Mc Culloch et Walter Pitts (1), qui s'inspirèrent des neurones biologiques pour programmer ces premiers neurones artificiels. Le terme d'IA a été pour la première fois utilisé en 1956 par John Mc Carthy et son équipe (2). En 1958, Frank Rosenblatt a créé le premier algorithme d'apprentissage basé sur un réseau de neurones, le Perceptron. Ce réseau permettait pour la première fois, à l'image des réseaux de neurones modernes, d'approximer une fonction inconnue lorsque les données d'entrée et de sortie étaient connues, mais que le lien entre elles ne pouvait être déterminé par les méthodes conventionnelles.

Aujourd'hui, l'un des domaines d'application les plus prometteurs de l'IA est la santé. Après la radiologie (3,4) et la médecine diagnostique (5), c'est désormais à la chirurgie de profiter du potentiel grandissant de l'IA. L'acte chirurgical, à fortiori en orthopédie, est une tâche complexe qui implique à la fois des compétences techniques et de prise de décision en temps réel. Les évènements indésirables évitables lors des actes opératoires sont estimées à 5,2% selon une revue systématique (6). Le premier point d'impact de l'IA dans notre spécialité est la formation chirurgicale, à travers divers moyens :

- Le feedback per opératoire : outil très puissant, un feedback au cours de l'intervention peut accélérer la courbe de progression par une évaluation de la performance en temps réel ;
- l'analyse post opératoire (7) : des vidéos enregistrées de l'intervention peuvent être une ressource pédagogique précieuse ;
- l'entraînement par simulation (8–12).

Le deuxième point d'impact de l'IA en chirurgie est l'intervention elle-même, autour de trois axes :

- La planification pré opératoire ;
- la navigation ou l'assistance (suivi et contrôle) per opératoire ;
- la prédiction du résultat post opératoire.

Le but de ce travail de thèse a été d'étudier la faisabilité d'un programme d'IA capable de donner un feedback per opératoire, lors de chirurgies arthroscopiques d'épaule. Ce choix de l'arthroscopie a été justifié :

- d'une part, par le fait qu'il s'agit d'une technique chirurgicale de réalisation difficile en pratique et dont la courbe d'apprentissage est longue et fastidieuse (13–15). Ces difficultés s'expliquent notamment par une anatomie spécifique et différente du ciel ouvert (16), une gestuelle subtile de triangulation instrumentale (17), et l'évolutivité rapide d'une discipline, finalement relativement « jeune » (18). La fiabilité de l'évaluation diagnostique arthroscopique per opératoire peut être alors parfois discutable car opérateur et expérience dépendants (19) (20). Une prise de décision adaptée devient difficile et potentiellement source d'erreurs ou de complications évitables (21).

- D’autre part, le signal vidéo produit par la caméra d’arthroscopie constitue une donnée d’analyse à part entière, pouvant être exploitée par une solution d’IA dans un but diagnostique et d’aide à la décision. Cette mise en relation entre des données visuelles (images, vidéos), et un apprentissage machine est l’une des applications du computer vision (22).

Ce « support arthroscopique » étant choisi, il a fallu ensuite identifier une problématique simple autour de laquelle articuler le projet. La gestion du tendon du long biceps (TLB) dans les petites ruptures de la coiffe des rotateurs reste une source de controverse : faut-il le conserver ? le ténotomiser ? faire une ténodèse ? Celle-ci relève davantage du dogme que d’un algorithme décisionnel précis (23). Bien que simple d’apparence, cette problématique soulève déjà de nombreuses questions, dont une fondamentale : « qu’est-ce qu’un long biceps normal ? ». Son anatomie, avec une portion intra-articulaire horizontale et une portion extra-articulaire verticale, ainsi que la fréquence des lésions associées, rendent le diagnostic clinique des lésions du TLB difficile. L’échographie et l’imagerie en coupes constituent une aide diagnostique (24–27), mais semblent insuffisantes pour établir un diagnostic préopératoire fiable (28). L’arthroscopie est alors considérée comme le gold-standard pour l’analyse du TLB (29,30). Dans une étude du service (acceptée pour publication très récemment), il a été néanmoins montré qu’en l’absence de protocole d’analyse précis, la concordance inter-observateur pour l’analyse des lésions du TLB était faible. Ainsi, en l’absence de données d’entrée consensuelles, une IA ne saurait guère faire mieux que l’analyse humaine.

C’est à partir de ces différentes observations que ce projet de thèse a été articulé en 2 parties. La première a consisté à réfléchir comment créer et valider une banque de données d’entrées arthroscopiques fiables pour son implémentation dans un programme d’IA. Une

analyse de concordance inter-observateur pour l'évaluation diagnostique du TLB a alors été conduite à partir d'images arthroscopiques issues de vidéos répondant un protocole précis. La deuxième partie a consisté à créer, entraîner, et évaluer une solution d'IA, dont le but était d'analyser le TLB sur des vidéos arthroscopiques enregistrées dans le cadre de petites ruptures de la coiffe des rotateurs. En associant les données images préalablement validées à des données médicales individuelles, il a été fait l'hypothèse qu'une IA pouvait être en mesure de statuer précisément sur l'aspect sain ou pathologique du TLB.

Article original

L'Intelligence Artificielle Peut Elle Aider à La Décision en Arthroscopie ?

Partie 1 : Analyse de Concordance Inter-Observateur pour l'Evaluation Arthroscopique Diagnostique du Tendon du Long Biceps Dans Les Petites Ruptures de la Coiffe des Rotateurs

Titre anglais : Can Artificial Intelligence Help Decision Making in Arthroscopy ?

Inter-Observer Analysis for the Arthroscopic Assessment of the Long Head of the Biceps in Small Rotator Cuff Tears

Rayane Benhenneda [a,*], Thierry Brouard [b], Julien Berhouet [a,b]

a Service de Chirurgie Orthopédique, Hôpital Trousseau, CHRU de Tours, Faculté de Médecine, Université de Tours Centre-Val de Loire, France.

b LIFAT (EA6300), Ecole Polytechnique Universitaire de Tours, 64 avenue Jean Portalis, Tours, 37200, France.

* Auteur correspondant : Rayane Benhenneda, Service de Chirurgie Orthopédique, Hôpital Trousseau, CHRU de Tours, avenue de la République, Chambray-lès-Tours, 37044 Tours cedex 9, France. Faculté de Médecine, université de Tours, France.

Mail : rayane.benhenneda@gmail.com

Résumé

Introduction : Les lésions du tendon du long biceps (TLB) sont fréquentes et leur diagnostic clinique ou par imagerie reste imprécis. L'arthroscopie est considérée comme le gold-standard pour l'analyse diagnostique du TLB, mais celle-ci reste parfois difficile. Sa fiabilité et sa reproductibilité n'ont par ailleurs pas été évaluées jusqu'à présent. L'intelligence artificielle (IA) pourrait apporter une aide pour l'analyse arthroscopique du TLB. L'objectif principal de cette étude était d'évaluer la concordance inter-observateur pour l'analyse spécifique du TLB, selon un protocole vidéo arthroscopique précis. L'objectif secondaire était de définir une base de données vidéos, dite « vérité terrain », destinée à créer et à entraîner une IA pour l'analyse du TLB.

Hypothèse : L'hypothèse était que la concordance d'analyse inter-observateur était suffisamment forte pour utiliser les vidéos comme « vérité terrain » d'une solution d'IA de diagnostic arthroscopique du TLB.

Matériels et méthode : Cent quatre-vingt-dix-neuf vidéos arthroscopiques protocolisées pour l'exploration du TLB ont été évaluées par 3 observateurs indépendants. Chacun devait caractériser le caractère sain ou pathologique du tendon, en précisant le type de lésion, dite « intrinsèque » : rupture partielle, hypertrophie avec sablier, instabilité, fissuration, désinsertion (SLAP 2), ou « extrinsèque » : chondral print et poulie pathologique sans instabilité. Les niveaux de concordance inter-observateur ont été mesurées au moyen du coefficient Kappa (K) de Cohen et du Kappa Accuracy.

Résultats : La force de concordance était modérée à forte selon les observateurs (Kappa 0,537 à 0,701 et KappaAcc de 86 à 92%), pour déterminer le caractère sain ou pathologique du TLB. Lorsque le tendon était pathologique, la force de concordance était modérée à forte lorsqu'il

s'agissait d'une rupture partielle (Kappa 0,494 à 0,708 et KappaAcc de 85 à 92%), d'une fissuration (Kappa -0,498 à 0,698 et Kappa Acc de 36 à 93%) ou d'une désinsertion (0,543 à 0,882 et Kappa Acc de 90 à 97%). Elle était faible pour une lésion d'instabilité (Kappa 0,039 à 0,248 et KappaAcc de 36 à 88%).

Conclusion : L'analyse du TLB, selon un protocole vidéo arthroscopique précis, était hautement concordante pour le diagnostic de son caractère sain ou pathologique. Le taux de concordance diminuait cependant pour le diagnostic de lésions tendineuses rares ou dynamiques. Une IA qui serait entraînée par une analyse humaine serait donc confrontée à ces mêmes difficultés si l'on se limitait à une analyse arthroscopique seule. L'intégration de données cliniques et para-cliniques est nécessaire à l'amélioration du diagnostic arthroscopique des lésions du TLB. Elle semble aussi être un prérequis indispensable à la constitution d'une base de données dite de « vérité terrain » pour la construction d'une solution d'IA performante.

Mots-clés : arthroscopie d'épaule ; intelligence artificielle ; analyse inter-observateur ; long biceps ; coiffe des rotateurs

Type d'étude : série prospective inter-observateur

Niveau de preuve : III

Abstract

Introduction : The lesions of the long head of the biceps (LHB) are frequent while their clinical or imaging diagnosis remains imprecise. Arthroscopy is considered to be the gold standard for the LHB analysis. However its reliability and reproducibility have not been evaluated so far. Artificial intelligence (AI) could help during arthroscopy for analyzing the LHB. The main objective of the current study was to assess the inter-observer analysis concordance for the LHB diagnosis, according to a precise arthroscopic video protocol. The secondary objective was to define a video database, called "ground truth", in aim to create and train an AI solution for analyzing the LHB.

Hypothesis : The inter-observer analysis concordance was strong enough to use the videos as the “ground truth” for an AI solution for arthroscopic diagnosis of the LHB.

Materials & Method : One hundred and ninety-nine protocolized arthroscopic videos for the exploration of the LHB were evaluated by 3 independent observers. Each had to characterize the healthy or pathological nature of the tendon, specifying the type of lesion, called "intrinsic": partial rupture, hypertrophy with hourglass figure, instability, cracking, disinsertion (SLAP 2), or "extrinsic": chondral print and pathological pulley without instability. Inter-observer agreement levels were measured using Cohen's Kappa (K) coefficient and Kappa Accuracy.

Results : The strength of concordance was moderate to strong according to the observers (Kappa 0,537 to 0,701 and KappaAcc from 86 to 92%), to determine the healthy or pathological character of the LHB. In case of pathological aspect, the concordance force was moderate to strong when it was a partial rupture (Kappa 0,494 to 0,708 and KappaAcc from 85 to 92%), a fissuration (Kappa -0,498 to 0,698 and KappaAcc from 36 to 93%), or a disinsertion (Kappa 0,543 to 0,882 and KappaAcc from 90 to 97%),, and low for an unstable lesion (Kappa 0,039 to 0,248 and KappaAcc from 36 to 88%).

Conclusion : Analyzing the LHB according to a precise arthroscopic video protocol was highly concordant for the diagnosis of its healthy or pathological status. However, the concordance rate decreased for the diagnosis of rare or dynamic tendon lesions. An AI that would be trained by a human analysis would therefore face the same difficulties if it was limited to a solely arthroscopic analysis. Additional clinical and para-clinical datas are necessary to improve the arthroscopic diagnosis of LHB lesions. It also seems to be an essential prerequisite for the constitution of a so-called “ground truth” database for the construction of a high-performance AI solution.

Keywords: shoulder arthroscopy; artificial intelligence ; inter-observer analysis; long biceps; rotator cuff

Study type : diagnostic study

Level of proof : III

Introduction

Le tendon du long biceps (TLB) est une source fréquente de douleurs antérieures d'épaule lorsqu'il est lésé (31,32). Son atteinte est souvent associée à des lésions de la coiffe des rotateurs ou des lésions du bourrelet glénoïdien (33–35). Son anatomie, avec une portion intra-articulaire horizontale et une portion extra-articulaire verticale, ainsi que la fréquence des lésions associées, rendent le diagnostic clinique des lésions du TLB difficile. L'échographie et l'imagerie en coupes constituent une aide diagnostique (24–27), mais semblent insuffisantes pour établir un diagnostic pré opératoire fiable (28). L'arthroscopie est alors considérée comme le gold-standard pour l'analyse du TLB (29,30). Sa fiabilité et sa reproductibilité n'ont en revanche pas été évaluées à ce jour et semblent même remises en cause en ce qui concerne la partie distale extra-articulaire du tendon, avec de nombreux faux négatifs (19,20). L'évaluation diagnostique arthroscopique est en effet exigeante sur le plan technique, rendant la prise de décision associée fortement conditionnée par l'expérience du chirurgien (21). C'est notamment le cas pour la gestion du TLB dans le cadre des ruptures distales de petite taille de la coiffe des rotateurs. Celle-ci relève davantage du dogme que d'un algorithme décisionnel précis (23).

Les champs d'application de l'intelligence artificielle (IA) dans le domaine médical, et notamment en chirurgie orthopédique, sont en constante évolution (36). Les données cliniques et paracliniques d'imagerie des patients sont principalement utilisées (37,38). En arthroscopie, le signal vidéo produit par la caméra constitue une donnée d'analyse à part entière, pouvant être aussi exploitée par une solution d'IA dans un but diagnostique et d'aide à la décision. Cette mise en relation entre des données visuelles (images ou vidéos), et un apprentissage machine est l'une des applications du computer vision. Le principe est d'extraire et de convertir des données descriptives et analytiques du monde visuel en un format numérique et symbolique,

permettant ensuite grâce à un apprentissage adapté d'interpréter de nouvelles informations visuelles et de guider une décision ou une action en rapport (22). Le *computer vision* fait écho à *l'human vision*, et renvoie de facto à la limite que l'analyse initiale de l'information visuelle est dépendante de l'humain, de ses capacités physiologiques et de sa subjectivité. Un prérequis de qualité concernant la constitution de la base de données initiales est alors indispensable à l'élaboration d'une solution d'IA. Sa performance sera conditionnée par des données d'entrée fiables qui devront correspondre à la « vérité terrain ».

Une solution d'IA pourrait être développée pour aider spécifiquement à la gestion du TLB dans les petites ruptures de coiffe. La création d'une base de données vidéos est l'une des informations d'entrée nécessaire. Sa qualité doit être évaluée pour être validée. L'objectif principal de cette étude était donc d'évaluer la concordance inter-observateur pour l'analyse spécifique du TLB, dont l'exploration reposait sur un protocole vidéo arthroscopique précis. L'objectif secondaire était de constituer une base de données, dite « vérité terrain arthroscopique », destinée à créer et à entraîner une IA pour la reconnaissance et l'analyse du TLB. L'hypothèse était que la concordance d'analyse inter-observateur était suffisamment forte pour utiliser les vidéos à la construction d'une « vérité terrain » en vue d'une solution d'IA de diagnostic arthroscopique du TLB.

Matériel et méthodes

Cette étude s'inscrit dans le cadre d'une étude clinique prospective, multicentrique, randomisée, en simple aveugle, ayant pour objectif d'analyser le résultat d'une réparation arthroscopique d'une rupture du tendon du supra-épineux de stade 1 selon Patte (39), en fonction de l'aspect macroscopique, normal ou pathologique, du TLB. Elle a été menée conformément à la Déclaration d'Helsinki et selon les Bonnes Pratiques Cliniques (40). Cette étude a été approuvée par les comités d'éthique institutionnels des différents établissements participants. Le consentement éclairé a été obtenu de l'ensemble des patients inclus (ID-RCB 2018-A01382-53).

Patients

Entre le 1er février 2019 et le 1er février 2021, 361 patients ont été inclus. Les critères d'inclusion étaient les suivants : âge compris entre 40 et 70 ans, bilan préopératoire complet (examen clinique, score de Constant bilatéral, imagerie en coupe), rupture distale du tendon du supra-épineux de stade 1 selon Patte (39) dans le plan frontal diagnostiquée sur l'imagerie en coupe et candidat à une réparation arthroscopique de la rupture. Ont été tolérées, pour l'inclusion, une délamination débutante du tendon de l'infra-épineux à partir de la rupture du tendon du supra-épineux et une rupture du tendon du sous-scapulaire de stade I selon la classification de la Société Francophone d'Arthroscopie (SFA) (41).

Les critères d'exclusion étaient les suivants : raideur préopératoire, rupture du tendon du supra-épineux à partir du stade II selon Patte, rupture du tendon du supra-épineux étendue au tendon de l'infra-épineux et/ou au tendon du sous-scapulaire, à partir du stade II selon la classification de la SFA, dégénérescence graisseuse du muscle supra-épineux supérieure au

stade 2 selon les classifications de Goutallier et Fuchs (42,43) et pathologie acromio-claviculaire diagnostiquée cliniquement et confirmée sur l'imagerie en coupe.

Vidéos arthroscopiques

Dans une précédente étude prospective (travail du service en cours de publication), il a été montré que l'évaluation arthroscopique du TLB sans protocole d'exploration opératoire précis, participait au faible niveau de concordance diagnostique mesuré entre différents opérateurs. Un protocole arthroscopique détaillé a par conséquent été mis au point pour l'analyse du TLB (Vidéo 1). Il comprenait un premier temps d'exploration intra-articulaire, par une voie d'abord postérieure optique (soft point). Quatre zones bicipitales étaient successivement analysées : 1) insertion à la glène, 2) portion intra-articulaire proximale, 3) portion intra-articulaire distale, 4) gouttière bicipitale. Un testing dynamique au crochet palpeur en élévation antérieure, et de stabilité médiale et latérale du TLB, était systématiquement réalisé. Un second temps d'exploration extra-articulaire, par une voie d'abord latérale, permettait d'évaluer la découverture potentielle du TLB par la rupture distale du tendon du supra-épineux.

En fin d'intervention, chaque opérateur ayant suivi ce protocole d'analyse précis devait renseigner le statut du TLB sur le cahier d'observation (e-case report form, eCRF) de l'étude:

- normal ;
- pathologique avec lésions intrinsèques : rupture partielle, hypertrophie avec sablier en élévation antérieure d'épaule (44), instabilité médiale ou latérale (45), fissuration, désinsertion (SLAP 2, superior labrum anterior-posterior tears) ;

- normal mais avec des lésions dites extrinsèques : cela correspondait à un tendon macroscopiquement normal, mais avec des lésions dites périphériques : chondral print (46,47), poulie pathologique (48,49), découverture du TLB par la rupture.

Des lésions intrinsèques et extrinsèques associées pouvaient être renseignées pour un même TLB analysé.

Chaque vidéo était enfin anonymisée et sauvegardée pour la suite de l'étude.

Analyse vidéo inter-observateur du TLB

Il s'agissait d'une analyse rétrospective, enrôlant 2 observateurs : 1 junior (O1), 1 sénior (O2), indépendants du chirurgien opérateur (C) ayant réalisé la vidéo arthroscopique lors de l'intervention.

Chaque vidéo pour être sélectionnée pour l'analyse inter-observateur devait remplir les critères de qualité suivants :

- l'exploration devait être complète et respecter le protocole vidéo mis en place;
- le format de chaque vidéo, la définition et la qualité des images (netteté, grossissement, angle de vue, présence de bulles, saignement…) devaient être suffisants pour être utilisés secondairement pour la création de la base de données nécessaires à la solution d'IA (50).

Chaque vidéo ainsi sélectionnée faisait ensuite l'objet d'une préparation standardisée, qui consistait à retenir 4 images d'intérêt (screenshots), sur lesquelles le zonage (bounding box) pour la détection du TLB (ROI, region of interest) étaient réalisées (*Figure 1*). Les 4 captures d'images correspondaient aux 4 zones d'analyse obligatoire du TLB selon le protocole d'exploration arthroscopique décrit précédemment. La quantité d'informations contenue dans

un flux vidéo ou une image entière étant trop importante et confondante, un processus de segmentation progressive de l'image est généralement nécessaire pour aboutir à une détection précise de l'objet par la machine dans l'image, basée sur le pixel. Cela permet ainsi à l'algorithme d'apprentissage de se focaliser sur la zone d'intérêt d'analyse, telle que le fait le chirurgien pour sa prise de décision lors de l'intervention. Ces 4 images sélectionnées et segmentées pour être implémentées dans la solution d'IA, étaient également celles soumises aux chirurgiens enrôlés pour l'analyse inter-observateur.

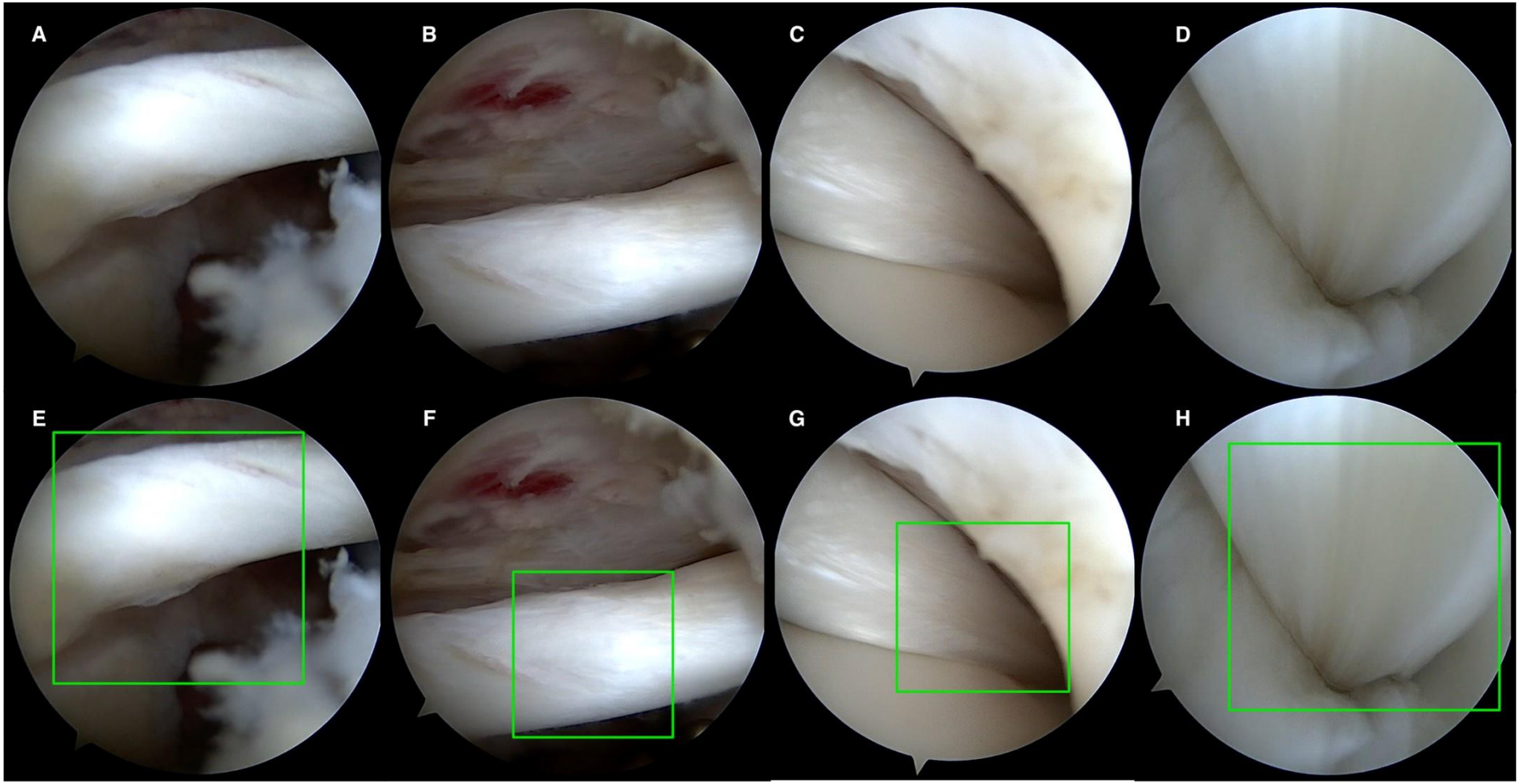


Figure 1 : *Images avec zones d'analyses du long biceps avant (A, B, C et D) et après (E, F, G, H) bounding box.*

Au final, les 2 observateurs indépendants suivaient la même grille d'analyse du TLB que celle renseignée par le chirurgien opérateur sénior, à la différence que ce dernier disposait lors de son intervention de l'ensemble des informations cliniques et para-cliniques du patient opéré, et que son analyse se faisait sur une exploration vidéo complète (non segmentée).

Analyse statistique

Il a été décidé de réaliser 2 analyses de concordance inter-observateurs. Ce choix a été justifié par le fait qu'il a été nécessaire de transformer les vidéos opératoires en 4 images segmentées, en vue de leur implémentation et leur traitement machine dans la solution d'IA prévue dans ce projet d'étude. A l'issue de ces 2 analyses, il était en effet attendu qu'une base de données arthroscopiques dites de « vérité terrain arthroscopique » puisse être établie.

La première analyse avait ainsi pour objectif de valider la qualité des images segmentées, constituant la base de données support au computer vision et à son implémentation dans la solution IA. Une analyse de concordance pour l'évaluation diagnostique du TLB sur les séries d'images segmentées, entre les 2 observateurs indépendants (O1, O2), a donc été conduite.

La seconde analyse avait pour objectif d'identifier quelles données d'évaluation du TLB devaient être retenues et associées aux données images pour compléter la base de données nécessaires à la création de l'IA. Une analyse de concordance pour l'évaluation diagnostique du TLB entre le chirurgien opérateur sénior (C) et les 2 observateurs indépendants (O1, O2) a été effectuée.

L'évaluation de la concordance inter-observateurs a été faite au moyen du coefficient Kappa (K) de Cohen (51). La force de l'évaluation était évaluée par l'échelle de Landis et Koch

(52) (Tableau 1). Cette échelle évalue cependant les scores de concordances « dans l'absolu », alors que plusieurs facteurs de confusion (prévalence et biais) peuvent en modifier l'interprétation (53). En effet, en fonction des cas, la valeur pratique du score n'atteint pas forcément 1, car la prévalence de chacun des codes employés pour une variable donnée ne sont pas égales. Une alternative méthodologique est de calculer des valeurs seuils de l'échelle de comparaison, en fonction de la précision supposée de l'observateur et du nombre de codes possibles pour une variable. Le Kappa Accuracy (54) est alors une information intéressante, nécessitant de renseigner en amont les informations suivantes :

- Nb Cas : le nombre de cas en commun et comparables entre 2 séries de données ;
- κ : (kappa) la valeur du score de Cohen ;
- κ max : la valeur max que κ peut prendre dans la configuration du test ;
- %Ag : pourcentage du nombre de cas où les observateurs sont d'accord ;
- %Ch : pourcentage de cas où les observateurs pourraient tomber d'accord de façon aléatoire, d'après les fréquences d'apparition observées des lésions ;
- Acc : précision estimée ou taux de fiabilité d'annotation pour les observateurs, pour avoir un score de Cohen à la valeur observée ;

La valeur Acc (pour accuracy) et correspondant au Kappa Accuracy s'interprête comme la précision que chaque observateur doit avoir afin de produire un score de Cohen de la magnitude observée.

L'analyse statistique a été réalisée au moyen des implémentations diffusées dans les distributions Scikit-Learn 1.1.2, Scipy.Stats 1.9.1 et Pandas 1.4.1, en Python 3.10.4 sous Jupyter Notebook. Les statistiques concernant le score de Cohen présenté ont été produits par KappaAcc (54).

Résultats

Sur les 361 vidéos des patients inclus dans l'étude, seules 199 remplissaient parfaitement les critères de qualité, en vue d'une exploitation secondaire pour la création d'une solution d'IA. Les vidéos exclues, avec une sélection volontairement drastique, l'ont été principalement sur des arguments de qualité insuffisante. Ce sont donc ces 199 vidéos qui ont été analysées par les 2 observateurs indépendants (O1, O2), en plus du chirurgien opérateur sénior (C) de chaque patient inclus ayant une vidéo exploitable.

Analyse descriptive : occurrence des lésions du TLB selon le diagnostic du « chirurgien opérateur » (C)

Dans l'analyse faite par C sur les vidéos arthroscopiques complètes, le TLB était identifié comme sain dans 119 cas (59,8%) et pathologique dans 80 cas (40,2%). Parmi les lésions du TLB visualisées, on recensait 148 lésions intrinsèques, réparties en : 29 (14,6 %) ruptures partielles, 27 (13,6 %) instabilités, 57 (28,6 %) fissurations, 13 (6,5 %) hypertrophies avec sablier, 22 (11,1%) SLAP 2. Il était retrouvé 58 lésions extrinsèques, réparties en : 15 (7,5 %) chondral print, 20 (10,1 %) cas de poulie anormale isolée et lorsque le TLB était sain, 23 (11,6 %) découvertures. Seuls 96 (48,2 %) TLB étaient finalement parfaitement sains.

Analyses des concordances inter-observateurs

(a) <u>Première analyse inter-observateur : O1 Vs O2 (Tableau 2)</u>

Selon l'échelle d'interprétation de Landis et Koch pour le coefficients Kappa (K) de Cohen (51), le niveau de concordance pour définir le caractère sain ou pathologique du TLB était fort (κ= 0,65).

Le niveau de concordance pour caractériser précisément l'aspect lésionnel du TLB était fort à modéré (κ= 0,53 à 0,7) en cas de rupture partielle et de désinsertion. Il était faible pour les diagnostics d'hypertrophie en sablier et d'instabilité tendineuse (κ= 0,24 à 0,3). Il était mauvais pour l'identification d'une fissuration du TLB (κ= -0,5).

Il n'a pas été possible de mesurer le niveau de concordance pour le diagnostic de lésions extrinsèques associées à un TLB normal en raison d'un défaut de prévalence lésionnelle.

En considérant le Kappa Accuracy, l'analyse de concordance pour définir le caractère sain ou pathologique du TLB s'interprétait comme suit : 191 cas comparables avaient été traités indépendamment par O1 et O2 ; l'accord était observé dans 83,25 % des cas alors qu'en notant aléatoirement il n'aurait été que de 52,11 %. Le score de κ mesuré était de 0,65, pour une valeur maximum de 0,868 (soit 74,9% de la valeur maximum). Pour atteindre ce score au moyen de 2 observateurs indépendants, leur précision devait être de 90 %, c'est à dire se tromper en moyenne 1 fois sur 10 dans leurs annotations. Ainsi, la précision d'évaluation des observateurs pour caractériser les différentes lésions intrinsèques en cas de TLB pathologique était de 86 et 91% dans les conditions de l'étude, à l'exception de l'aspect « fissuration ».

(b) Deuxième analyse inter-observateur : C Vs O1 Vs O2 (Tableau 3)

Le niveau de concordance pour définir le caractère sain ou pathologique du TLB étaient forts entre C et O1 (κ= 0,7), et modéré entre C et O2 (κ= 0,54), avec des précisions minimales selon le Kappa Accuracy respectivement de 92% et 86%.

Le niveau de concordance pour caractériser précisément l'aspect lésionnel du TLB était fort entre C et O1 (κ= 0,7) et modéré entre C et O2 (κ =0,49 à 0,54) en cas de rupture partielle et de désinsertion. Il était mauvais à faible entre C et O1 et O2 pour les diagnostics

d'hypertrophie en sablier et d'instabilité tendineuse (κ= 0,08 à 0,3). Il était fort entre C et O1 (κ= 0,7) et mauvais entre C et O2 pour l'identification d'une fissuration du TLB (κ= -0,05).

Il n'a pas été possible d'évaluer le niveau de concordance pour le diagnostic de lésions extrinsèques associées à un TLB normal, car il n'y avait aucune interprétation commune entre les 3 chirurgiens pour l'ensemble des TLB sains visualisés.

Discussion

L'hypothèse de l'étude est confirmée. Les vidéos arthroscopiques dans le cadre d'un protocole précis, peuvent être utilisées comme support de « vérité terrain arthroscopique » pour la création d'une solution d'IA d'aide au diagnostic lésionnel du TLB dans les petites ruptures de la coiffe des rotateurs. La forte concordance inter-observateur (O1 Vs O2) valide d'une part la fiabilité des images segmentées issues des vidéos arthroscopiques pour l'analyse du caractère sain ou pathologique du TLB. La concordance modérée à forte entre ces mêmes observateurs et le chirurgien opérateur (C) valide d'autre part l'utilisation des données d'analyse du TLB faite par ce dernier lors de son intervention. C'est ainsi l'association des images arthroscopiques segmentées et des données d'interprétation per opératoire, qui peut être utilisée pour « vérité terrain arthroscopique » de la solution d'IA envisagée dans ce projet d'étude.

Cette interprétation des résultats peut être pondérée après démembrement lésionnel du TLB. De forte ou très forte pour les lésions dites statiques (« SLAP », ruptures partielles), la concordance selon les 2 niveaux d'analyse (O1 Vs O2 et O1-O2 Vs C) diminuait pour les lésions à caractère dynamique (hypertrophie en sablier, instabilité), ou de faible prévalence (poulie anormale isolée, chondral print), ou pour le diagnostic d'une fissuration du TLB. Ce résultat pouvait être attendu pour les lésions dynamiques, puisque leur analyse a été faite sur des images statiques, alors qu'elles sont recherchées en pratique par un testing au crochet palpeur, associé à une mobilisation en rotation du membre du patient.

Lorsque la confrontation d'analyse se faisait avec le chirurgien opérateur (C), il a aussi été retrouvé une différence de concordance en faveur de l'observateur junior (O1), pour le diagnostic lésionnel du TLB et son démembrement. Celle-ci peut s'expliquer par le fait que c'est O1 qui a préparé les images segmentées. Leur sélection et le positionnement des bounding

box ont nécessité une visualisation répétée des vidéos, induisant un biais d'entraînement pour l'analyse du TLB. O1 se rapprochait ainsi des conditions de lecture de C, qui en plus de la vidéo arthroscopique complète, disposait de l'ensemble des données cliniques et paracliniques du patient opéré. Par extension, on peut admettre que plus l'observateur possède d'informations (vidéo, données patient…), plus fine et précise est son analyse du TLB. Il s'agit en somme d'autant d'arguments pour retenir les données issues de l'interprétation du chirurgien opérateur comme « vérité terrain » pour la solution IA envisagée.

Certaines lésions, à un stade débutant, sont aussi sujettes à la subjectivité de l'observateur. Une fissuration minime (*Figure 2*) ou un biceps en sablier légèrement hypertrophique (*Figure 3*) peuvent être analysés comme sain ou lésé selon l'observateur. Les conditions d'analyses du TLB dans cette population avec rupture du tendon du supra-épineux de stade 1 sont les moins favorables à un consensus. En effet, en cas de ruptures importantes de la coiffe, voire de rupture massive (55), le TLB est souvent le siège d'une lésion évidente en stade de pré-rupture, et donc de diagnostic consensuel (35,56).

Pour les lésions extrinsèques, le niveau de concordance était beaucoup plus faible, voire proche de l'aléatoire. Cela peut s'expliquer par le fait que la prévalence de ces lésions est faible en pratique, et encore plus dans le cadre de lésions distales ou de petite taille de la coiffe des rotateurs. Le faible nombre de publications pour chacune des lésions extrinsèques, avec parfois des descriptions non consensuelles (46,47), témoignent aussi de leur rareté, et par conséquent de la difficulté pour les chirurgiens à les rechercher et à les diagnostiquer, si tant est qu'ils en connaissent eux-mêmes la description. Les informations cliniques et paracliniques (imagerie) pré-opératoires revêtent alors toute leur importance pour l'analyse du TLB et sa gestion.

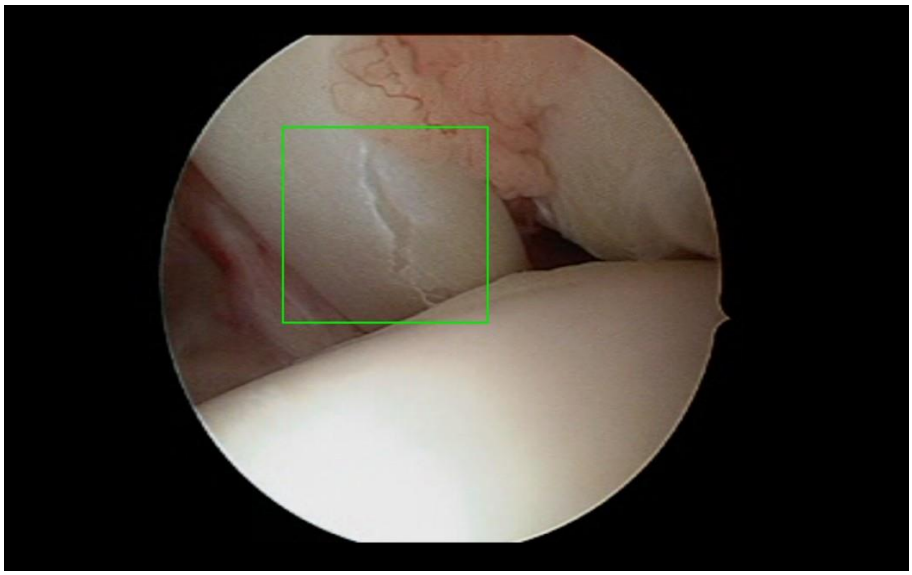


Figure 2 *: Image avec bounding box sur une fissuration, pouvant être considérée comme un TLB sain ou pathologique selon la subjectivité de l'observateur.*

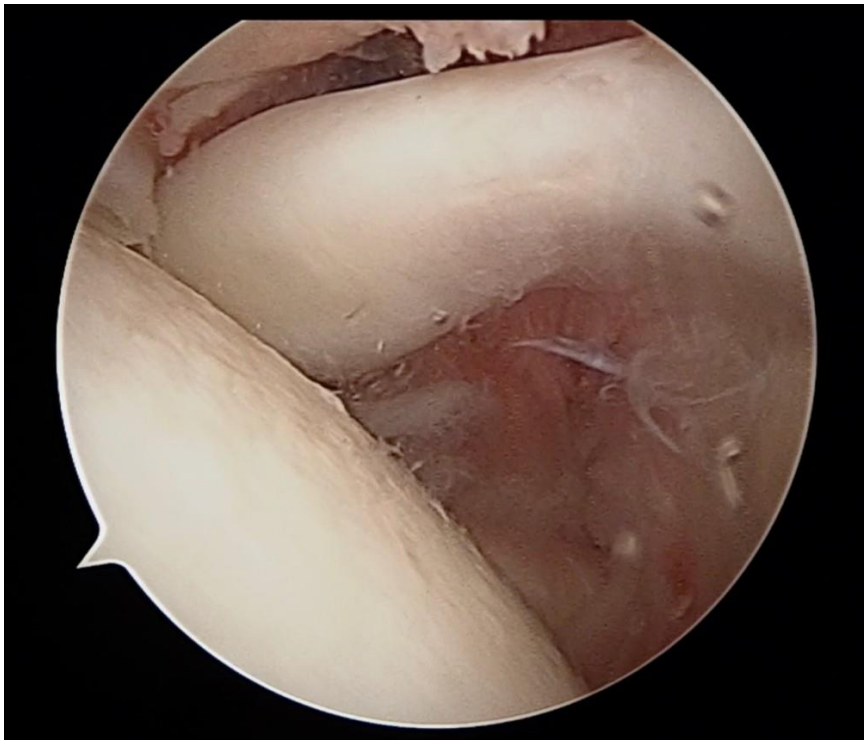


Figure 3 : Image avec hypertrophie en sablier, pouvant être considérée comme un TLB sain ou pathologique selon la subjectivité de l'observateur.

Le traitement des observations contradictoires de différents experts reste une question ouverte (57). L'une des solutions est le raisonnement par incertitude, tel que modélisé par la théorie des fonctions de croyance de Dempster et Shafer (58). Cette théorie permet de combiner des preuves provenant de différentes sources, et d'arriver à un degré de croyance qui prend en compte toutes les observations disponibles. Concernant l'apprentissage d'une IA, il pourrait s'agir d'une solution pour établir une « vérité terrain » fondée sur l'observation de plusieurs experts, limitant ainsi les biais de confirmation ou d'influence auxquels pourrait être sujet un expert synthétique supplémentaire et qui ne serait pas en aveugle. Dans le cadre de cette étude pour l'analyse du TLB, seuls les opérateurs (C) avaient connaissance des données cliniques et d'imagerie préopératoire de chacun des patients qu'ils avaient à charge. C'est sur ce postulat qu'il peut être aussi validé le fait que l'apprentissage de l'IA pour le diagnostic du TLB peut utiliser pour « vérité terrain » les données d'analyse arthroscopique réalisée par les opérateurs (C).

Les points forts de cette étude ont été d'évaluer le TLB sur un grand nombre de patients, selon un protocole d'analyse arthroscopique standardisé, statique et dynamique. La reproductibilité de l'analyse vidéo atteste de la qualité du protocole proposé, alors même que les observateurs O1 et O2 étaient en aveugle de l'opérateur C. Les biais de confirmation et d'influence étant ainsi limités, ce protocole d'analyse per opératoire pourrait être applicable en routine clinique. Ces mêmes observateurs indépendants ne disposaient pas non plus des informations cliniques et paracliniques préopératoires. S'il peut s'agir d'une limite méthodologique de l'étude, il s'agit aussi d'un facteur de consolidation supplémentaire du protocole vidéo mis en place, et du choix qui a été fait de retenir l'analyse de l'opérateur C pour « vérité terrain » dans la création de l'IA en projet. Les points faibles de l'étude sont d'avoir

utilisé un protocole statique, ne permettant pas d'analyser correctement les lésions dynamiques, et l'effet d'entraînement qui est un biais d'analyse pour O1.

Conclusion

L'analyse arthroscopique diagnostique du TLB, considérée comme le gold-standard, est reproductible à condition d'être protocolisée, pour déterminer son caractère sain ou pathologique et caractériser les lésions intrinsèques statiques. Cette analyse est en revanche mise à défaut avec le démembrement lésionnel du TLB en devenant de plus en plus aléatoire à mesure que la lésion est rare. L'analyse dynamique s'est avérée peu pertinente du fait de la segmentation des images. Une IA qui serait entraînée par une analyse humaine serait donc confrontée à ces mêmes difficultés si l'on se basait uniquement sur l'analyse arthroscopique. L'intégration de données cliniques et paracliniques constitue un prérequis obligatoire à l'amélioration du diagnostic arthroscopique des lésions du TLB. L'ensemble de ces informations doit être intégrée pour la constitution d'une base de données, dit de « vérité terrain », en vue de la création d'une IA performante.

Conflits d'intérêts :

En lien avec ce travail : aucun des auteurs.

En dehors de ce travail : aucun des auteurs.

Financement :

Aucun financement n'a été reçu pour cette étude.

Contributions :

Rayane Benhenneda : collecte et analyse des données, rédaction du manuscrit.

Thierry Brouard : analyse des données, statistiques, relecture.

Julien Berhouet : conception de l'étude, analyse des données, rédaction, relecture.

Tableaux

Tableau 1 : Interprétation du score Kappa de Cohen selon Landis et Koch (52)

Kappa	**Interprétation**
< 0	Désaccord
0,00 — 0,20	Accord très faible
0,21 — 0,40	Accord faible
0,41 — 0,60	Accord modéré
0,61 — 0,80	Accord fort
0,81 — 1,00	Accord presque parfait

Tableau 2 : Analyse des concordances inter-observateur O1 versus O2

Critère	Nombre	Cohen κ	κ max	%Ag	%Ch	Acc
Pathologique	191	0,650	0,868	83,25	52,11	90,00
Hypertrophie	58	0,301	0,651	93,1	90,1	91
Rupture partielle	59	0,532	0,666	76,3	49,3	86
Instabilité	57	0,248	0,549	91,2	88,3	88
Fissuration	59	-0,498	0,160	32,2	35,4	<36
SLAP 2	52	0,707	0,902	88,5	60,6	93

Tableau 3 : Analyse des concordances inter-observateur C versus O1 et O2

Critère	Obs	Nombre	Cohen κ	κ max	%Ag	%Ch	Acc
Pathologique	O1-C	198	0,701	0,936	85,86	52,65	92,00
	O2-C	192	0,537	0,925	77,60	51,60	86
Hypertrophie	O1-C	62	0,281	0,461	87,1	82,0	86
	O2-C	57	0,301	0,700	87,7	82,4	87
Rupture partielle	O1-C	62	0,708	0,903	85,5	50,2	92
	O2-C	57	0,494	0,562	73,7	47,9	85
Instabilité	O1-C	62	0,039	0,140	69,3	68,1	<36
	O2-C	55	0,076	0,353	63,6	60,7	65
Fissuration	O1-C	61	0,698	0,871	88,5	62,0	93
	O2-C	56	-0,048	0,109	28,6	31,8	<36
SLAP 2	O1-C	59	0,882	0,961	94,9	56,9	97
	O2-C	48	0,543	0,804	85,4	68,1	90

Légendes des figures

Figure 1 : Images correspondant aux 4 zones d'analyse du TLB : 1) insertion à la glène, 2) portion intra-articulaire proximale, 3) portion intra-articulaire distale, 4) gouttière bicipitale, avant (A, B, C et D) et après (E, F, G, H) l'étape de segmentation (bounding box en vert).

Figure 2 : Image avec bounding box sur une fissuration, pouvant être considérée comme un TLB sain ou pathologique selon la subjectivité de l'observateur.

Figure 3 : Image avec hypertrophie en sablier, pouvant être considérée comme un TLB sain ou pathologique selon la subjectivité de l'observateur.

Article original

L'Intelligence Artificielle Peut Elle Aider à La Décision en Arthroscopie ?

Partie 2 : Le Modèle IA-RTHRO. Aide à la Décision Diagnostique pour le Statut du Tendon du Long Biceps Dans Les Petites Ruptures de la Coiffe des Rotateurs

Titre anglais : The IA-RTRHO Model for Help Decision Making for the Diagnosis of the Long Head of the Biceps Status in Small Rotator Cuff Tears

Rayane Benhenneda [a,*], Thierry Brouard [b], Julien Berhouet [a,b]

a Service de Chirurgie Orthopédique, Hôpital Trousseau, CHRU de Tours, Faculté de Médecine, Université de Tours Centre-Val de Loire, France.

b LIFAT (EA6300), Ecole Polytechnique Universitaire de Tours, 64 avenue Jean Portalis, Tours, 37200, France.

* Auteur correspondant : Rayane Benhenneda, Service de Chirurgie Orthopédique, Hôpital Trousseau, CHRU de Tours, avenue de la République, Chambray-lès-Tours, 37044 Tours cedex 9, France. Faculté de Médecine, université de Tours, France.

Mail : rayane.benhenneda@gmail.com

Résumé

Introduction : Les applications possibles de l'intelligence artificielle (IA) en chirurgie orthopédique sont prometteuses. Le deep learning est utilisable en chirurgie arthroscopique grâce au signal vidéo exploitable par computer vision. La gestion per-opératoire du tendon du long biceps (TLB) est l'objet d'une controverse ancienne. L'objectif principal de cette étude était de modéliser une IA diagnostique capable de statuer sur le caractère sain ou pathologique du TLB sur des images arthroscopiques. L'objectif secondaire était de créer un deuxième modèle IA diagnostique en fonction d'images arthroscopiques et des données médicales de chaque patient, cliniques et imagerie, pour statuer sur le caractère sain ou pathologique du TLB.

Hypothèse : L'hypothèse de l'étude était qu'il était possible de construire un modèle IA à partir d'images opératoires d'arthroscopie d'aide au diagnostic du caractère sain ou pathologique du TLB.

Matériels et Méthodes : Les données prospectives cliniques et d'imagerie de 199 patients ont été collectées et associées aux images issues d'une analyse vidéo arthroscopique protocolisée validée, dite « vérité terrain », faite par le chirurgien opérateur. Un modèle basé sur un réseau neuronal convolutif (RNC) modélisé via un *transfer learning* sur le modèle Inception V3 a été construit pour l'analyse des images arthroscopiques. Ce modèle a ensuite été couplé à un Perceptron MultiCouche (*MultiLayer Perceptron, MLP*) intégrant les données cliniques et imageries. Chaque modèle a été entraîné et testé selon apprentissage supervisé.

Résultats : La précision du RNC sur le diagnostic du statut sain ou pathologique du TLB était de 93,7% en apprentissage et 80,66% en généralisation. Couplé aux données cliniques de chaque patient, la précision du modèle assemblant le RNC et un MLP étaient respectivement de 77% et 58% en apprentissage et en généralisation.

Conclusion : Le modèle IA construit à partir d'un RNC parvient a déterminer le caractère sain ou pathologique du TLB avec un taux de précision de 80,66%. Ce modèle n'est pas suffisamment pertinent pour une application en pratique clinique. Une augmentation des données d'entrée pour limiter l'overfitting, et l'automatisation de la phase de détection par un *Mask-R-CNN* sont des voies d'amélioration du modèle. Cette étude est la première à évaluer la capacité d'une IA à analyser des images arthroscopiques, et ses résultats doivent être confirmés par d'autres études sur le sujet.

Mots-clés : arthroscopie ; intelligence artificielle ; réseau neuronal convolutif ; tendon du long biceps ; apprentissage profond

Type d'étude : Étude diagnostique

Niveau de preuve : III

Abstract

Introduction : Artificial Intelligence (AI) is a promising asset in orthopedic surgery. Deep learning can be used in arthroscopic surgery due to the video signal that can be used by computer vision. The intraoperative management of the long biceps tendon (LTB) is the subject of a long-standing controversy. The main objective of this study was to model a diagnostic AI capable of deciding on the healthy or pathological character of LTB on an arthroscopic image. The secondary objective was to create a second diagnostic AI model based on arthroscopic images and the medical data of each patient, clinical and imaging, to decide on the healthy or pathological nature of the LTB.

Hypothesis : The hypothesis of the study was that it was possible to build an AI model from operative arthroscopic images to help diagnose the healthy or pathological nature of LTB.

Materials & Methods : The clicnal and imaging data of 199 patients were collected prospectively and associated with images from a validated protocolized arthroscopic video analysis, called “ground truth”, made by the operating surgeon. A model based on a convolutional neural network (CNN) modeled via transfer learning on the Inception V3 model was built for the analysis of arthroscopic images. This model was then coupled to a MultiLayer Perceptron (MLP) integrating clinical and imaging data. Each model has been trained and tested using supervised learning.

Results : The accuracy of the CNN on the diagnosis of the healthy or pathological status of the LTB was 93.7% in learning and 80.66% in generalization. Coupled with the clinical data of each patient, the accuracy of the model assembling the CNN and the MLP were respectively 77% and 58% in learning and in generalization.

Conclusion : The AI model built from a CNN manages to determine the healthy or pathological character of the LTB with an accuracy rate of 80.66%. This model is not sufficiently relevant for application in clinical practice. An increase in input data to limit overfitting, and the automation of the detection phase by a Mask-R-CNN are ways of improving the model. This study is the first to assess the ability of an AI to analyze arthroscopic images, and its results need to be confirmed by other studies on the subject.

Keywords: shoulder arthroscopy; artificial intelligence ; convolutional neural network; long tendon biceps; deep learning

Study type : diagnostic study

Level of evidence : III

Introduction

L'intelligence artificielle (IA) est la capacité d'une machine à simuler l'intelligence humaine. Sa capacité à gérer et optimiser de grands ensembles de données en fait un atout prometteur dans le secteur de la santé, et particulièrement en chirurgie orthopédique (59,60). L'apprentissage profond, ou « deep-learning », est une méthode d'apprentissage tentant de modéliser des données avec un haut niveau d'abstraction (61). Les techniques de deep-learning ont atteint des niveaux de performance équivalents, voire supérieurs à la performance humaine (62). L'effort de recherche pour l'application du deep-learning en chirurgie orthopédique porte actuellement sur 3 domaines d'application : la planification pré opératoire (63,64), la navigation et le contrôle per opératoires (65), et la prédiction du résultat clinique post opératoire (66). En chirurgie arthroscopique, la caméra produit un signal visuel utilisé pour le diagnostic et la prise de décision. Ce signal est une donnée exploitable en IA par « computer vision » (22).

L'arthroscopie nécessite une courbe d'apprentissage longue et difficile (13–15), s'expliquant par une visualisation anatomique spécifique et différente du ciel ouvert (16), et par une gestuelle subtile basée sur la triangulation dans un espace réduit (17). Ses spécificités techniques font de l'arthroscopie une discipline expérience-dépendante (18), dont la sensibilité et la spécificité d'analyse diagnostique per opératoire pour une structure anatomique donnée deviennent discutables (19). C'est en effet la fiabilité de la prise de décision per opératoire qui peut être remise en cause, la rendant alors plus difficile et exigeante car potentiellement source d'erreurs et de complications évitables en pratique (20) (21).

La gestion per opératoire du tendon du long biceps (TLB) est l'objet d'une controverse ancienne dans le cadre de la prise en charge arthroscopique des ruptures de la coiffe des rotateurs (23). Si la question d'ordre technique entre ténotomie ou ténodèse semble résolue dans

le cadre de lésions évidentes du TLB lors des lésions larges à massives de la coiffe (67,68), l'indication thérapeutique elle-même reste discutée en cas de ruptures distales ou de petite taille (69–71). Le TLB est en effet fréquemment intact dans ce contexte, ou présente à défaut des lésions minimes, dont l'anticipation diagnostique pré opératoire est également difficile, tant lors de l'examen physique du patient qu'à l'imagerie (échographie, scanner et IRM) (72,73). Le diagnostic du statut du TLB est alors posé au moment même de l'intervention, exposant le chirurgien à une prise de décision parfois difficile quant à la conduite à tenir (68) face à une situation non préparée (74,75). Le fait de pouvoir anticiper le statut du TLB avant l'opération pourrait ainsi constituer une aide intéressante pour l'opérateur quant à sa stratégie chirurgicale.

L'objectif principal de cette étude était de proposer un premier modèle expérimental d'IA d'aide à la décision dans le but de statuer sur le caractère sain ou pathologique du TLB dans le cadre des petites ruptures distales de la coiffe des rotateurs. L'objectif secondaire était de préparer un second modèle d'IA, cette fois-ci d'aide à la décision sur la nature du geste thérapeutique à réaliser en fonction du statut du TLB. L'hypothèse de l'étude était qu'il était possible de construire un modèle d'IA à partir d'images opératoires d'arthroscopie, capable d'aider à définir le caractère sain ou pathologique du TLB.

Matériel et méthodes

Cette étude s'inscrivait dans le cadre d'une étude clinique prospective, multicentrique, randomisée, en simple aveugle, ayant pour objectif d'analyser le résultat d'une réparation arthroscopique d'une rupture du tendon du supra-épineux de stade 1 selon Patte (39), en fonction de l'aspect macroscopique, normal ou pathologique, du TLB. Elle a été menée conformément à la Déclaration d'Helsinki et selon les Bonnes Pratiques Cliniques (40). Cette étude a été approuvée par les comités d'éthique institutionnels des différents établissements participants. Le consentement éclairé a été obtenu de l'ensemble des patients inclus (ID-RCB 2018-A01382-53).

Base de données pour le modèle d'IA expérimental : la « vérité terrain »

L'ensemble des données utilisées pour la construction du modèle d'IA a été recueilli de façon prospective lors de l'inclusion des patients et de leur suivi. Il s'agissait pour chaque patient inclus des données cliniques, d'examens complémentaires et des images d'arthroscopie.

Données cliniques

Les données cliniques recueillies étaient celles recueillies lors de l'anamnèse (sexe, âge, côté dominant, score ASA (76), indice de masse corporelle (IMC), antécédents médicaux, Functional Comorbidity Index (FCI) (77), activité professionnelle et sportive, antécédent traumatique sur l'épaule opérée) et de l'examen clinique (score de Constant (78) pré opératoire, score SSV, Subjective Shoulder Value (79), score ASES, American Shoulder and Elbow surgeons Score (80), Echelle Visuelle Analogique (EVA) douleur, et tests diagnostiques cliniques). L'ensemble des données cliniques sont résumées dans le Tableau 1.

Données d'imagerie

Les données paracliniques étaient celles des imageries pré opératoires. Pour chaque patient, étaient renseignées :

- Radiographies standard de face et de profil : CSA, Critical Shoulder Angle (81) et stade de Bigliani (82) ;
- IRM : taille de la rupture du supra-épineux dans le plan frontal et sagittal, présence ou non d'un clivage de l'infra-épineux ou d'une lésion du sous-scapulaire, espace dans la gouttière, diamètre et position du TLB, présence ou non d'une anomalie de signal en T1 ou en T2 ;
- Échographie par un radiologue : taille, surface, grade de vascularisation, position et échogénicité du TLB évalués à 2 localisations : dans la gouttière et à la partie haute de celle-ci.

Images d'arthroscopie

Le protocole vidéo évalué en première partie de ce travail a permis de valider une base de données vidéos constituant les données d'entrée du programme d'IA. Chaque vidéo a été segmentée en 4 images visualisant le TLB dans 4 zones d'intérêt : 1) insertion à la glène, 2) portion intra-articulaire proximale, 3) portion intra-articulaire distale et 4) gouttière bicipitale. Le zonage des ROI, region of interest, était matérialisé par une bounding box (Figure 1). La première partie de notre étude a permis d'une part de valider l'utilisation d'images pour caractériser le caractère sain ou pathologique du TLB et en faire l'analyse pathologique, et d'autre part de déterminer que l'analyse « vérité terrain » par le chirurgien opérateur était la plus précise. C'est donc celle-ci qui a servi de base d'entraînement.

Construction du modèle d'IA

Le modèle de l'IA diagnostique a été formé par la fusion de deux modèles préalables construits séparément.

Un 1er modèle, basé sur un réseau de neurones convolutifs (*convolutional neural network, CNN*), a été construit pour traiter les données d'imagerie, via un *transfer learning* (83). Notre choix s'est tourné vers le modèle InceptionV3 (84), entraîné sur la base ImageNet et disponible dans TensorFlow™ 2.0 (Google Brain, Santa Clara, États-Unis, tensorflow.org). Le *CNN* est schématisé en *Figure 4*.

Un second modèle a été construit pour traiter les données médicales (données cliniques et paracliniques). Ces données ont été réencodées à partir d'un tableur CSV, puis intégrées à un *MultiLayer Perceptron (MLP)* en entrée sous forme de vecteurs de nombres en une dimension, avec un vecteur pour chaque patient.

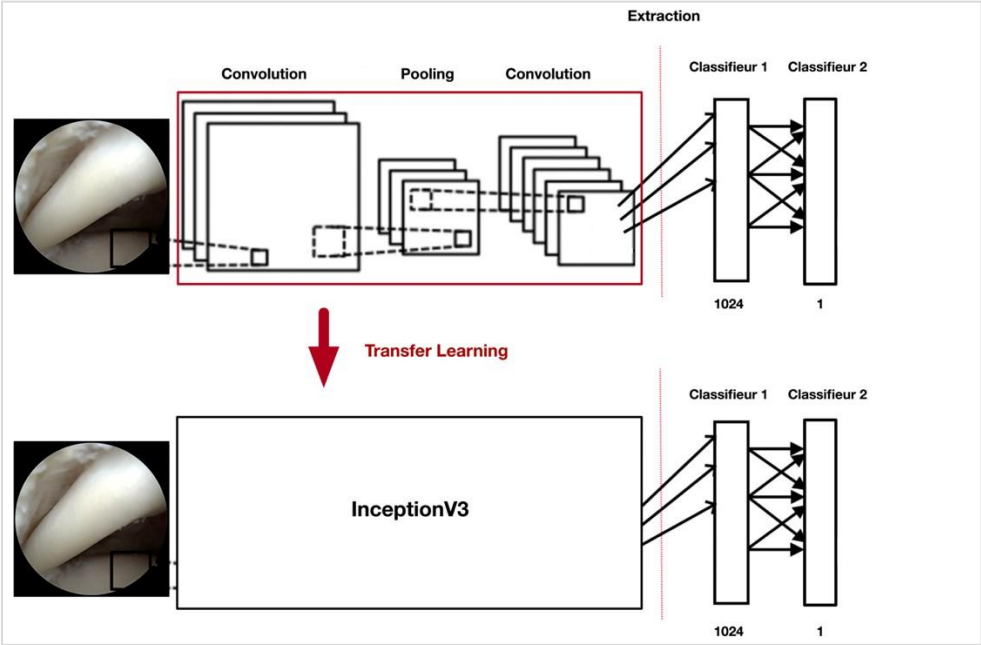


Figure 4 : *Architecture du réseau neuronal convolutif construit*

Chacun de ces deux modèles a été appris séparément, avant d'être associé pour construire le 3ème modèle définitif d'IA d'aide à la décision diagnostique sur le statut sain ou pathologique du TLB. Le but était que ce dernier modèle utilise un maximum d'informations (données médicales et images) pour rendre sa décision (*Figure 5*).

Le codage a été réalisé en langage Python pour chacun des 3 modèles.

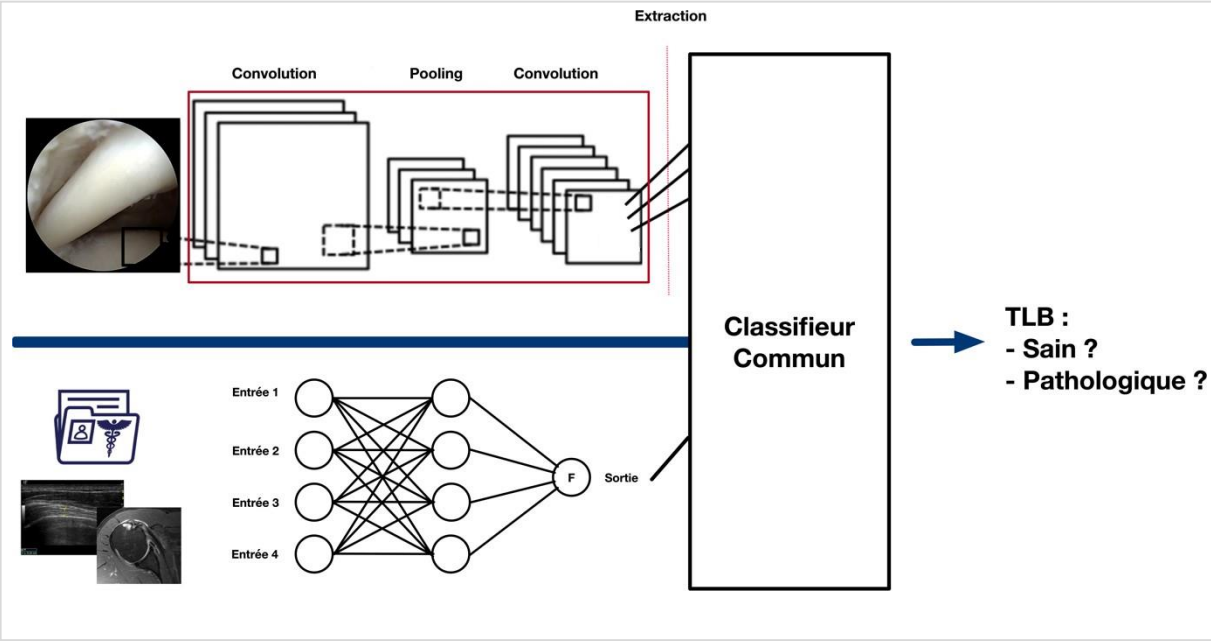


Figure 5 : *Architecture du 3ème modèle construit, basé sur le CNN et le MLP. F = Fonction d'activation ReLU. TLB = Tendon du long biceps.*

Entraînement des modèles d'IA

L'entraînement de chaque modèle se faisait en deux temps :

- *l'apprentissage*, consistant à tester l'apprentissage par cœur de l'algorithme, en lui présentant uniquement des données pour lesquelles il connaît déjà la réponse ;
- *la généralisation*, consistant à tester la capacité de l'algorithme à donner une réponse sur une donnée qu'il n'a pas rencontré par le passé.

Modèle n°1 : Méthode d'apprentissage du CNN

<u>Implémentation</u>

L'apprentissage de la fonction de détection des images a été fait par *transfer learning* (83) en utilisant le modèle pré-entraîné InceptionV3 recueilli sur TensorFlow™ 2.0. La fonction de classification a été entrainée par un apprentissage supervisé, utilisant les données « vérité terrain » préparées en amont.

Hyperparamètres

Le *learning rate* alpha choisi pour la descente de gradient comme étant le plus optimal était 10^{-4}. Une couche de linéarisation a été ajoutée, donnant à la sortie un bloc de 3x3x2048 = 18432 valeurs. Une première couche de classification sur 1024 niveaux, puis une deuxième couche sur 1 niveau permettait d'obtenir un nombre entre 0 et 1 à la sortie. Si ce nombre était inférieur à 0,5, le modèle reconnaissait un tendon normal, sinon il reconnaissait un tendon pathologique.

Preprocessing

La base d'apprentissage était constituée de 580 images (243 TLB « pathologiques » et 337 TLB « normaux »). La série d'apprentissage (*training set*) était constituée de 90% des images, et la série de test *(testing set*) était constituée de 10% des images. Afin de pallier le faible nombre d'images d'apprentissage et éviter un phénomène d'*over-fitting*, une augmentation automatique des données d'apprentissage a été utilisée avec les paramètres suivants :

- Rotation : 10% ;
- Etirement en largeur : 20% ;
- Etirement en hauteur : 20% ;

- Zoom : 20% ;
- Symétrie horizontale.

Le modèle avait donc la possibilité de produire de manière aléatoire des images qui étaient ajoutées lors de la phase d'apprentissage en utilisant une ou plusieurs de ces transformations (*Figure 6*).

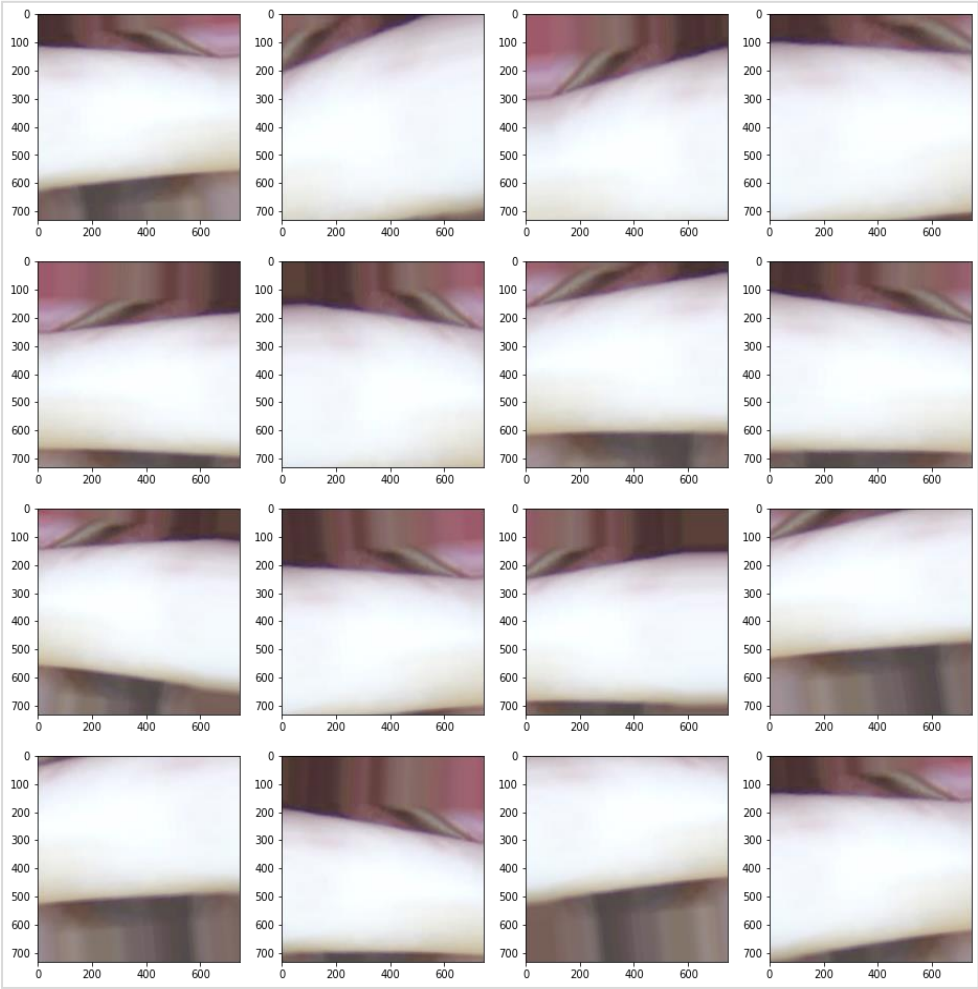


Figure 6 : *Exemple de transformations subies par l'une des images de la base d'apprentissage.*

Entraînement

Le modèle InceptionV3 était composé de 40 679 201 paramètres, dont 18 876 417 étaient modifiables par l'algorithme d'apprentissage. Les 21 802 784 paramètres restants représentaient le « savoir-faire » acquis durant l'apprentissage d'ImageNet. L'apprentissage s'est fait par lots de 20 images sur 200 époques. Durant une époque, le modèle apprenait l'ensemble des images de la base d'apprentissage par paquets de 20 constitués aléatoirement, et cherchait à prédire les classes (sain ou pathologique) des images qu'il venait d'apprendre. Ensuite il constituait des lots aléatoires de 20 images issues de la base de test et évaluait le nombre d'erreur commises sur les images qu'il n'avait pas apprises.

Test

L'exécution a été réalisée en ligne, sur l'infrastructure *Cloud* de Google (Google Inc, Moutain View, Etats-Unis) sur une architecture *GPU* (*Graphics Processing Units*).

Modèle n°2 : Méthode d'apprentissage du MLP

Implémentation

La fonction de classification a été entrainée par un apprentissage supervisé, utilisant les données 1D sous la forme de vecteurs de nombres. Sur les 199 patients, 156 avaient des données médicales complètes et donc exploitables.

Hyperparamètres

Les données d'entrée 1D étaient soumises à un réseau de neurones artificiels à 3 couches avec fonction d'activation ReLU (*Rectified Linear Unit*). Deux couches de 64 et 32 neurones étaient traversées, avant d'arriver sur la dernière couche proposant 2 sorties « tendon sain » ou « tendon pathologique ».

Entraînement

Chaque neurone des différentes couches recevant l'intégralité des informations de la couche précédente, et délivrant sa sortie à tous les neurones de la couche suivante, le modèle MLP était composé de 4194 paramètres à estimer. L'apprentissage s'est fait par lots de 20 patients sur 500 époques. Le ratio apprentissage/généralisation a été de 80/20.

Test

L'exécution a été réalisée hors ligne, sur un ordinateur de bureau (Quad-Core i7, 16 GO de RAM, sans *GPU*).

Modèle n°3 : Association des deux modèles IA

Les deux modèles ont été associés en utilisant un classifieur commun sous forme de *MultiLayer Perceptron*, basé sur les données obtenues par l'analyse des images d'une part et sur celles des données cliniques et d'imagerie d'autre part (*Figure 5*).

Implémentation

Le modèle a été entrainé par un apprentissage supervisé. On disposait pour chaque patient de 2 vecteurs : un vecteur image et un vecteur données médicales qui étaient associés pour n'en former qu'un. En tout, 156 vecteurs correspondant à 156 patients ont été utilisés. Chaque patient était soit dans le set d'apprentissage, soit dans le set de test.

Hyperparamètres

Le *learning rate* alpha choisi pour la descente de gradient comme étant le plus optimal était $9x10^{-8}$. Le *MLP* chargé de traiter les données mixtes comportait 4 couches. Une première était composée de 1056 entrées (1024 issues de l'image, et 32 issues des données médicales) traitées par 256 neurones artificiels. Les deuxième et troisième couches intermédiaires étaient composées respectivement de 128 et 32 neurones. Une dernière couche comportait 2 neurones.

Preprocessing

La série d'apprentissage (*training set*) était constitué de 80% des vecteurs, et la série de test (*testing set*) était constitué de 20% des vecteurs. Comme pour le 1er modèle, une augmentation a été nécessaire pour limiter l'*over-fitting*. Seules les images de la base d'apprentissage étaient augmentées avec les paramètres suivants :

- Rotation : 10% ;
- Étirement en largeur : 10% ;
- Étirement en hauteur : 10% ;
- Zoom : 10% ;

- Perspective : 10% ;
- Symétrie horizontale.

Après augmentation de la base de données d'images, la base d'apprentissage était constituée de 500 vecteurs : 400 pour la base d'apprentissage et 100 pour la base de set.

Entraînement

Les modèles n°1 et n°2 ayant été préalablement entraînés séparément (cf supra), leurs couches de classification finale ont été retirées, de façon à obtenir en sortie du modèle 3 la projection des données dans les espaces de classification qui devaient être utilisés en entrée d'un dernier *MLP* « classifieur commun ». Celui-ci tenait compte des 2 résultats intermédiaires et en faisait une synthèse finale.

Test

L'exécution a été réalisée en ligne, sur l'infrastructure *Cloud* de Google (Google Inc, Moutain View, Etats-Unis) sur une architecture *GPU* (*Graphics Processing Units*).

La *Figure 7* résume les 3 modèles d'IA construits.

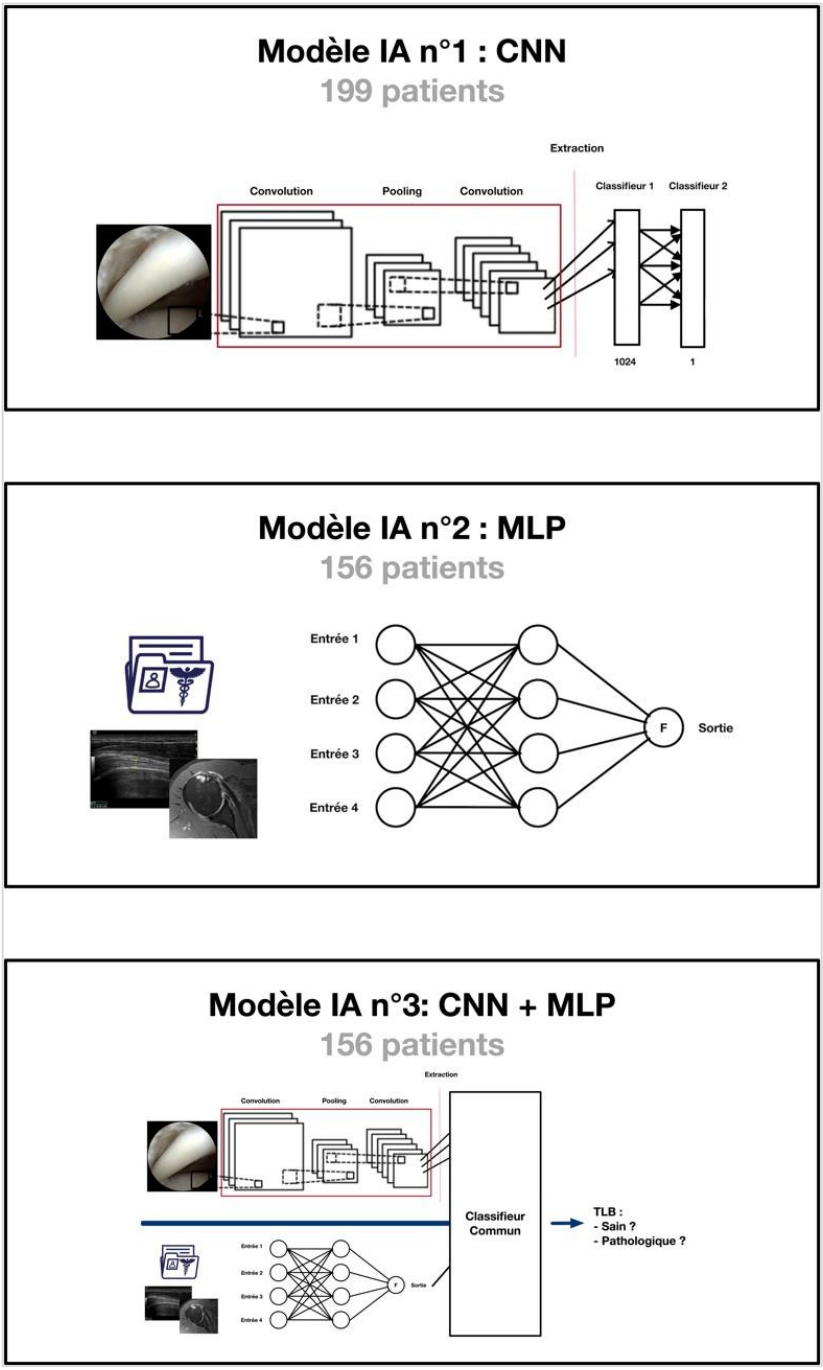


Figure 7 : *Résumé des 3 modèles d'IA conçus. IA = Intelligence Artificielle ; CNN = Convolutional Neural Network ; MLP = MultiLayer Perceptron ; TLB = Tendon du Long Biceps*

Résultats

Précision du modèle n°1 *(Figure 8)*

L'évolution de la précision du modèle durant les étapes d'apprentissage (200 époques) a été croissante, avec une précision initiale de l'ordre de 60% pour atteindre un plateau en fin de processus entre 90 et 95%. La précision moyenne sur les 5 dernières époques était de 93,7% en apprentissage et de 80,66% en généralisation.

Le temps d'exécution d'une époque était en moyenne de 13 secondes, et le temps total d'apprentissage était de 45 minutes.

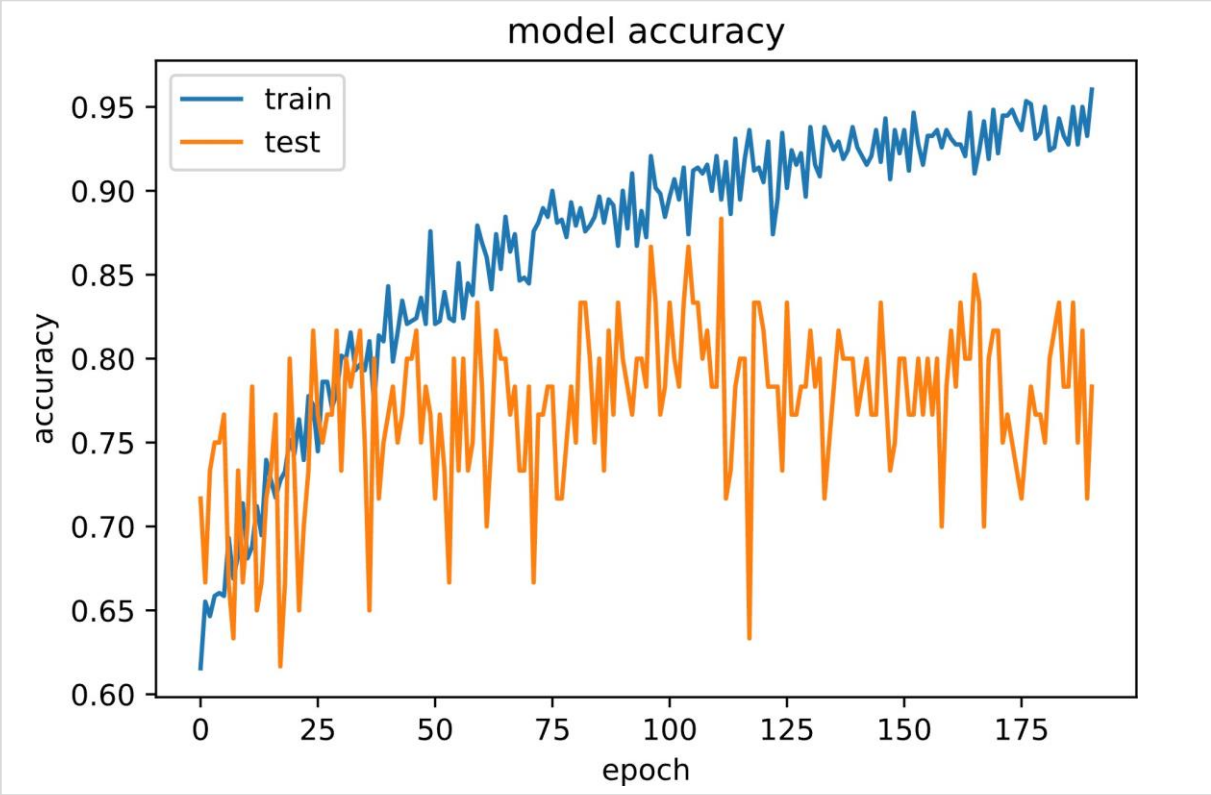


Figure 8 : *Précision du modèle n°1 (CNN). En bleu, courbe d'apprentissage (training set) et en orange, courbe de test (testing set). Model accuracy : précision du modèle, Accuracy : précision, epoch : époque.*

Précision du modèle n°2 *(Figure 9)*

Les résultats en précision de l'apprentissage et de la généralisation en fin de processus (500 époques) étaient respectivement de 85% et 75%. Le temps d'apprentissage total était de 2 minutes.

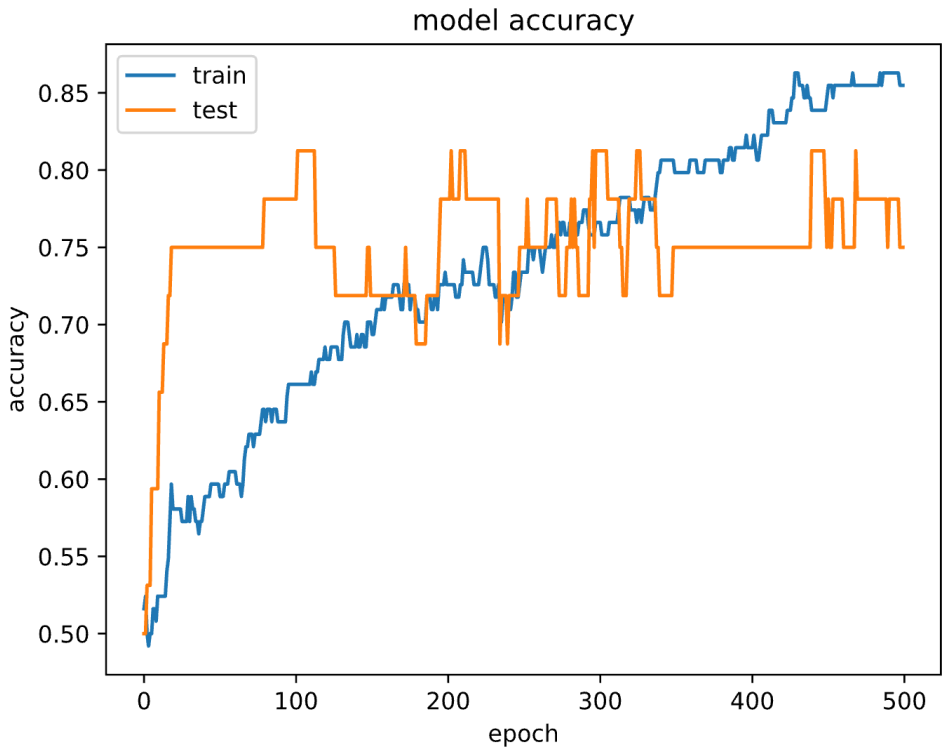


Figure 9 : *Précision du modèle n°2 (MLP). En bleu, courbe d'apprentissage (training set) et en orange, courbe de test (testing set). Model accuracy : précision du modèle, Accuracy : précision, epoch : époque.*

Précision du modèle n°3 *(Figure 10)*

Les résultats en précision de l'apprentissage et de la généralisation en fin de processus (200 époques) étaient respectivement de 77% et 58%.

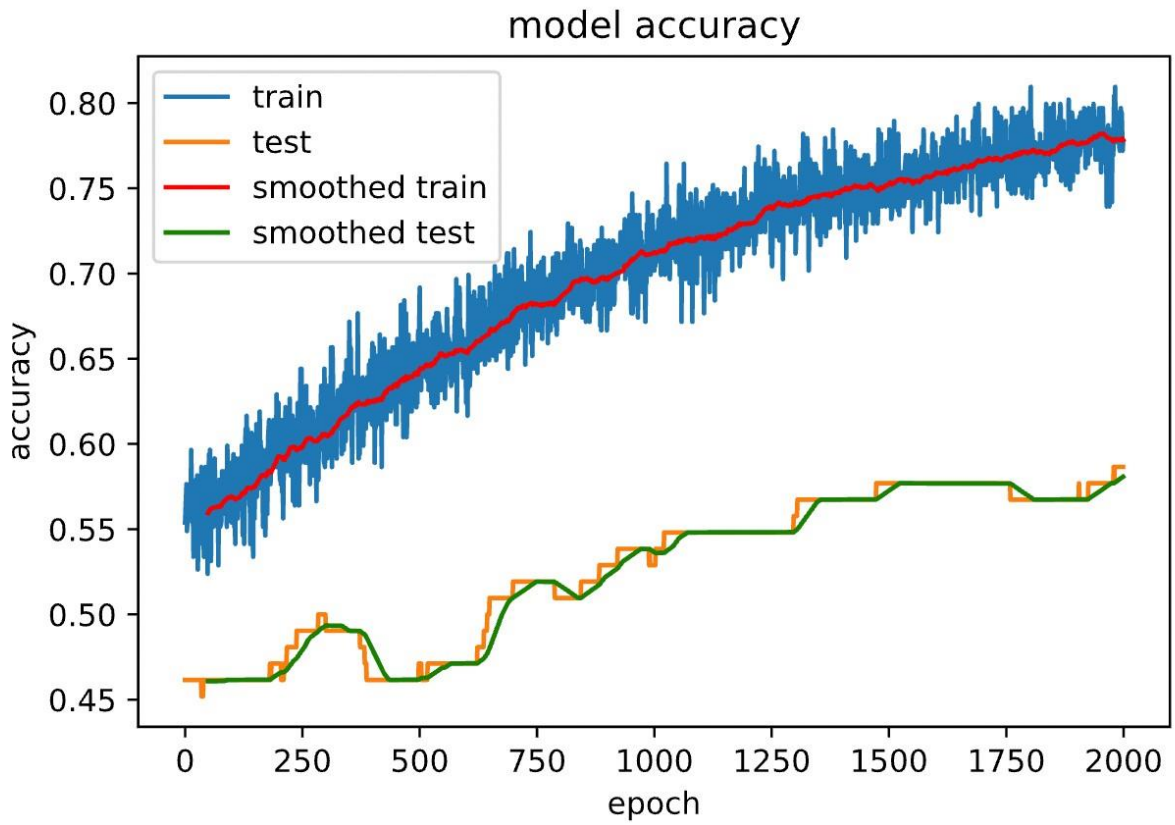


Figure 10 : *Précision du modèle n°3 (assemblage des modèles 1 et 2). En bleu, courbe d'apprentissage (training set) et en orange, courbe de test (testing set). Model accuracy : précision du modèle, Accuracy : précision, epoch : époque.*

Discussion

Nous avons rapporté les résultats d'un modèle d'IA original, destiné à statuer sur le caractère sain ou pathologique du TLB en se basant sur des images issues de vidéos arthroscopiques. Notre hypothèse a été confirmée. Le modèle n°1 est parvenu à déterminer le caractère sain ou pathologique du TLB avec un taux de précision de 80,66%. Le set d'apprentissage était proche de 95% de prédiction correctes. Il s'agissait d'un comportement normal, montrant la capacité de l'algorithme d'apprentissage à agir sur les dernières couches du modèle de façon à discriminer les deux catégories d'images de façon précise. D'après les paramètres d'apprentissage, une réponse au hasard entraînait 243/580 = 41,9% de chances de reconnaître un TLB comme pathologique et donc 58,1% de le qualifier de normal. L'évolution de la précision calculée sur la base de test mettait en évidence trois éléments remarquables. 1/ Sa tendance suivait la précision de la courbe d'apprentissage, ce qui montrait une capacité du modèle à généraliser, c'est-à-dire à prédire de manière correcte les images non apprises en se basant sur ce que l'algorithme avait appris. 2/ Sa performance restait en retrait de celle de l'apprentissage, ce qui était attendu avec ce type de modèle. 3/ Le caractère oscillant de la courbe de test était dû au faible nombre d'images et au fait que celles-ci n'étaient pas augmentées, contrairement à la base d'apprentissage. L'homogénéité des lots envoyés pour l'évaluation de la précision du testing set était faible, ce qui conduisait à des scores par lots plus faibles. En moyenne, la précision du testing set était de 75% sur les 75 dernières époques. Étant donné la grande variabilité des images et leur faible nombre, le modèle utilisé donnait de très bons résultats en apprentissage. En généralisation, les performances étaient aussi très bonnes et l'écart s'expliquait aisément par la quantité d'images disponibles et leur hétérogénéité dans les scènes représentées.

Concernant le modèle n°2, le caractère oscillant de la courbe de test mettait en évidence le faible nombre de données d'entrée, permettant toutefois d'améliorer la capacité de généralisation du modèle jusqu'à 75% en fin de processus. Cela traduisait un cas d'over-fitting qui s'expliquait par une quantité de données d'entrée limitée. Une réponse au hasard entraînait pour ce modèle 50% de chances de qualifier le TLB comme normal et 50% comme pathologique.

Le troisième modèle ne semblait pas bénéficier de l'association des données images et médicales, puisque sa précision était moins bonne que chacun des deux premiers modèles pris séparément. Une réponse au hasard entraînait pour ce modèle 50% de chances de qualifier le TLB comme normal et 50% comme pathologique. Le modèle faisait légèrement mieux que le hasard, avec un taux de précision de 58%. Une première hypothèse à cela était que pour ce modèle, chaque patient était soit dans la base d'apprentissage (*training set*), soit dans la base de test (*testing set*). Dans le premier modèle, la machine avait la liberté, sur 5 images d'un même patient d'en utiliser 3 pour l'apprentissage et 2 pour le test. Les performances étaient donc moindres dans ce cas plus restrictif où les tests ne portaient que sur des patients disjoints. Une seconde hypothèse était que l'ajout de données médicales, pour lequel le modèle n°2 a présenté un over-fitting évident, limitait l'ensemble du modèle n°3 dans ses performances (en le rendant lui-même sujet à de l'over-fitting). Ces données médicales pouvaient ajouter un bruit supplémentaire qui, du fait du faible nombre de données, réduisait les performances de l'ensemble.

Pour être pertinent en pratique clinique, un programme doit avoir une performance supérieure ou égale à celle d'un clinicien expérimenté. Les taux de précision inter-observateurs calculés en première partie de ce travail étaient, sur la base de l'image, de 86 à 92%. Avec une

précision de 80,66% pour le premier modèle basé sur l'image uniquement, qui diminuait à 58% lorsque les données images et médicales (clinique et imagerie) étaient couplées, notre programme n'apparaît au final pas suffisamment efficient. Ces résultats s'expliquent par le faible nombre de données utilisées pour entraîner ces trois modèles, malgré des méthodes d'augmentation d'image.

Malgré la performance limitée des 3 modèles décrits, il s'agit de la première étude rapportant les résultats d'un algorithme de *deep learning* entraîné à identifier des lésions sur images arthroscopiques, avec des résultats encourageants. Guy et al. rapportaient une sensibilité et spécificité respectivement de 61 et 67% d'un algorithme entraîné à diagnostiquer des fractures du col fémoral sur deux incidences radiographiques avec une base de donnée de 1309 images (36). Avec une base de données de 256000 images, Olczak et al. obtenaient une précision de 83% de détection de fractures sur différentes radiographies (poignet, main et cheville) (85). Dans ces deux cas, les CNN utilisés n'avaient pas été entraînés au préalable sur des images non-médicales. En utilisant le *transfer learning*, Adams et al. obtenaient une précision de 90,5% pour détecter une fracture du col fémoral sur une incidence radiographique avec une base de donnée de 640 images (86). De même, Kim et al. avec 11112 images analysées par un CNN modélisé via un *transfer learning* sur le modèle Inception V3, obtenaient une sensibilité de 0,9 et une spécificité de 0,88 sur la détection de fractures en imagerie radiographique. Cela montre une première limite que l'on a pour l'utilisation de l'IA en chirurgie orthopédique : le manque de base de données massives (*big data*), qui n'est pas spécifique à notre spécialité (87,88). Ces *big data* peuvent être issues pour le moment de registres cliniques ou de données administratives (89) pour réduire les problèmes d'over-fitting en attendant des solutions nouvelles. Pour le moment, le *transfer learning* semble être une

solution partielle à ce manque de données mais ne suffit pas à lui seul pour atteindre des taux de précision équivalent au clinicien humain.

Les technologies utilisées par les algorithmes sont également une piste d'amélioration. En effet, notre modèle utilisait un *CNN* simple. Il a été montré qu'avec des R-CNN plus complexes de développement récents, les *Mask-R-CNN* (90), les niveaux de performances des algorithmes s'améliorent (91). La nécessité d'une segmentation manuelle des images par un expert, pour permettre au modèle un apprentissage supervisé, est également une limite dans la constitution de larges bases de données. Les *R-CNN* et *Mask-R-CNN* s'affranchissent de cela en automatisant la segmentation d'une image (*Figure 11*).

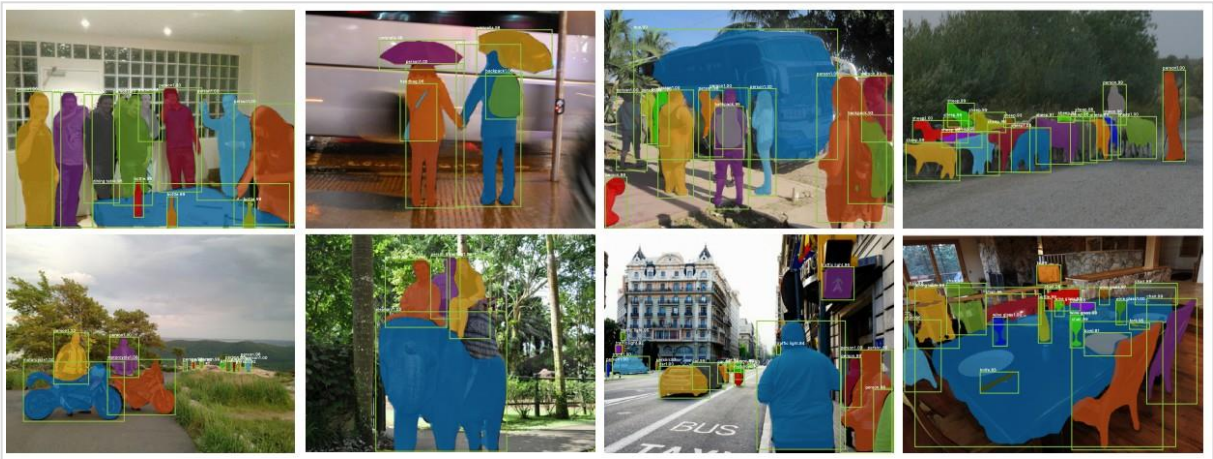


Figure 11 : *Segmentation automatique d'images par un Mask-R-CNN (90).*

Notre étude présente plusieurs limites :

- Un faible nombre de données, à la fois images arthroscopiques (199 patients) et cliniques ou paracliniques (156 patients) expliquant le niveau de performance limité des 3 modèles étudiés ;

- les images pour l'entraînement supervisé ont été labellisées par un seul clinicien, chirurgien junior ;
- il n'a pas été possible de réaliser une analyse de pondération des différents facteurs médicaux pour anticiper le caractère sain ou pathologique du TLB.

Notre étude présente néanmoins des points positifs. Il a été utilisé une méthode de *transfer learning* pour améliorer les performances des modèles IA, ce qui semble être un pré-requis indispensable par comparaison aux modèles sans *transfer learning* rapportés dans la littérature. Il a été réalisé une validation qualitative du *data-set* exploité pour la construction du modèle IA. L'analyse inter-observateur des images d'arthroscopie faite en première partie de ce travail était en effet un pré-requis indispensable à la création d'une IA, dont la performance est étroitement liée à la qualité de ses données d'entrée.

Conclusion

L'analyse du caractère sain ou pathologique du TLB par une IA basée sur des images arthroscopiques est possible avec une précision diagnostique allant jusqu'à 80%. Ce taux n'est cependant pas suffisamment élevé pour avoir une pertinence clinique. L'utilisation de base de données plus importantes et l'utilisation de R-CNN plus complexes de type *Mask-R-CNN* sont des voies d'amélioration qui pourraient permettre d'augmenter ces résultats.

Conflits d'intérêts :

En lien avec ce travail : aucun des auteurs.

En dehors de ce travail : aucun des auteurs.

Financement :

Aucun financement n'a été reçu pour cette étude.

Contributions :

Rayane Benhenneda : collecte et analyse des données, rédaction du manuscrit.

Thierry Brouard : analyse des données/statistiques, relecture.

Julien Berhouet : conception de l'étude, rédaction, relecture.

Tableaux

Tableau 4 : Données démographiques de la population étudiée

Catégorie	Nombre (%)
Patients	199
Épaule opérée	
Côté gauche	64 (32,2)
Côté droit	135 (67,8)
Côté opéré = épaule dominante	137 (68,8)
Sexe	
Masculin	95 (47,7)
Féminin	104 (52,3)
Âge moyen, années	55,3 +/- 8,9 (33-78)
Indice de Masse Corporelle (IMC)	28,7 +/- 3,7 (18-42,8)
Score ASA (American Society of Anesthesiologists) (76)	
ASA 1	131 (65,8)
ASA 2	57 (28,6)
ASA 3	11 (5,5)
ASA 4	0
Comorbidités	
Tabagisme actif	28 (14,1)
Hypercholestérolémie	35 (17,6)
Diabète	12 (6,0)
Hyperuricémie	3 (1,5)
Functional Comorbidity Index (FCI) (77)	0,35 (0-4)
Activité	
Patient retraité au moment de l'intervention	57 (28,6)
AT/MP	35 (17,6)
Pratique d'un sport avant chirurgie (loisir)	70 (35,2)
Antécédent traumatique < 12 mois sur l'épaule opérée	60 (30,2)

Tableau 5 : Scores cliniques pré-opératoires de la population étudiée

Catégorie	Moyenne (Intervalle)
Score de Constant (78)	
Douleur	6,79 +/- 2,72 (0-15)
Activité	10,45 +/- 3,57 (1-20)
Mobilités	32,42 +/- 6,99 (8-40)
Force	6,07 +/- 4,45 (0-22)
Total	55,71 +/- 13,69 (21-86)
Score SSV (79)	48,61 +/- 16,11 (10-90)
Score ASES	43,55 +/- 20,79 (6-92)
EVA douleur (0-10)	5,45 +/- 1,92 (0-10)

Légendes

Figure 4 : Architecture du réseau neuronal convolutif construit.

Figure 5 : Architecture du modèle construit, basé sur le CNN et le MLP. F = Fonction d'activation ReLU. TLB = Tendon du long biceps.

Figure 6 : Exemple de transformations subies par l'une des images de la base d'apprentissage.

Figure 7 : Résumé des 3 modèles d'IA conçus. IA = Intelligence Artificielle ; *CNN = Convolutional Neural Network ; MLP = MultiLayer Perceptron* ; TLB = Tendon du Long Biceps

Figure 8 : Précision du modèle n°1 (*CNN*). En bleu, courbe d'apprentissage (training set) et en orange, courbe de test (testing set). *Model accuracy* : précision du modèle, *Accuracy* : précision, *epoch* : époque.

Figure 9 : Précision du modèle n°2 (*MLP*). En bleu, courbe d'apprentissage (training set) et en orange, courbe de test (testing set). *Model accuracy* : précision du modèle, *Accuracy* : précision, *epoch* : époque.

Figure 10 : Précision du modèle n°3 (assemblage des modèles 1 et 2). En bleu, courbe d'apprentissage (training set) et en orange, courbe de test (testing set). *Model accuracy* : précision du modèle, *Accuracy* : précision, *epoch* : époque.

Figure 11 : Segmentation automatique d'images par un Mask-R-CNN (90).

Conclusion générale

Ce travail a consisté en une première initiation à l'IA. Il a permis d'identifier les difficultés à construire un modèle fiable. La qualité des données d'entrées est déterminante. Le choix de la méthode d'apprentissage elle-même, adaptée à la question d'aide à la décision posée, est aussi décisif.

La première question posée dans ce travail était une aide à la décision diagnostique. A partir de la même base de données, complétée de résultats cliniques à un délai post opératoire suffisant, nous avons pour projet d'investiguer la question d'une aide à la décision thérapeutique quant au geste à réaliser dans la situation particulière d'un TLB normal. Il s'agirait alors d'une IA pronostique, qui pourrait servir de base à la construction d'un score fonctionnel pronostique pour la gestion du TLB dans les petites ruptures de la coiffe des rotateurs.

Bibliographie

1. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys. 1 déc 1943;5(4):115-33.

2. Russell SJ, Norvig P, Davis E. Artificial intelligence: a modern approach. 3rd ed. Upper Saddle River: Prentice Hall; 2010. 1132 p. (Prentice Hall series in artificial intelligence).

3. Jiang Y, Edwards AV, Newstead GM. Artificial Intelligence Applied to Breast MRI for Improved Diagnosis. Radiology. janv 2021;298(1):38-46.

4. Shen YT, Chen L, Yue WW, Xu HX. Artificial intelligence in ultrasound. Eur J Radiol. juin 2021;139:109717.

5. Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann Oncol Off J Eur Soc Med Oncol. 1 août 2018;29(8):1836-42.

6. Anderson O, Davis R, Hanna GB, Vincent CA. Surgical adverse events: a systematic review. Am J Surg. août 2013;206(2):253-62.

7. Sirleo L, Innocenti M, Innocenti M, Civinini R, Carulli C, Matassi F. Post-operative 3D CT feedback improves accuracy and precision in the learning curve of anatomic ACL femoral tunnel placement. Knee Surg Sports Traumatol Arthrosc Off J ESSKA. févr 2018;26(2):468-77.

8. Bouthors C, Dagneaux L, Boisgard S, Garreau de Loubresse C, Benhamou D, Court C. Procedural simulation training in orthopaedics and traumatology: Nationwide survey among surgeon educators and residents in France. Orthop Traumatol Surg Res OTSR. 7 juin 2022;103347.

9. Marcheix PS, Vergnenegre G, Dalmay F, Mabit C, Charissoux JL. Learning the skills needed to perform shoulder arthroscopy by simulation. Orthop Traumatol Surg Res OTSR. juin 2017;103(4):483-8.

10. Tronchot A, Berthelemy J, Thomazeau H, Huaulmé A, Walbron P, Sirveaux F, et al. Validation of virtual reality arthroscopy simulator relevance in characterising experienced surgeons. Orthop Traumatol Surg Res OTSR. déc 2021;107(8):103079.

11. Lacroix PM, Commeil P, Chauveaux D, Fabre T. Learning and optimizing arthroscopic knot-tying by surgery residents using procedural simulation. Orthop Traumatol Surg Res OTSR. déc 2021;107(8):102944.

12. Walbron P, Common H, Thomazeau H, Hosseini K, Peduzzi L, Bulaid Y, et al. Virtual reality simulator improves the acquisition of basic arthroscopy skills in first-year orthopedic surgery residents. Orthop Traumatol Surg Res OTSR. juin 2020;106(4):717-24.

13. Hoppe DJ, de Sa D, Simunovic N, Bhandari M, Safran MR, Larson CM, et al. The learning curve for hip arthroscopy: a systematic review. Arthrosc J Arthrosc Relat Surg Off Publ Arthrosc Assoc N Am Int Arthrosc Assoc. mars 2014;30(3):389-97.

14. Hodgins JL, Veillette C, Biau D, Sonnadara R. The knee arthroscopy learning curve: quantitative assessment of surgical skills. Arthrosc J Arthrosc Relat Surg Off Publ Arthrosc Assoc N Am Int Arthrosc Assoc. mai 2014;30(5):613-21.

15. Choe JS, Bin SI, Lee BS, Kim JM, Song JH, Cho HK. Learning Curve For Lateral Meniscal Allograft Transplantation: Preventing Meniscal Extrusion. Arthrosc J Arthrosc Relat Surg Off Publ Arthrosc Assoc N Am Int Arthrosc Assoc. nov 2021;37(11):3326-34.

16. Paxton ES, Backus J, Keener J, Brophy RH. Shoulder arthroscopy: basic principles of positioning, anesthesia, and portal anatomy. J Am Acad Orthop Surg. juin 2013;21(6):332-42.

17. Anetzberger H, Becker R, Eickhoff H, Seibert FJ, Döring B, Haasters F, et al. The Diagnostic Arthroscopy Skill Score (DASS): a reliable and suitable assessment tool for arthroscopic skill training. Knee Surg Sports Traumatol Arthrosc Off J ESSKA. janv 2022;30(1):349-60.

18. Randelli P, Cucchi D, Butt U. History of shoulder instability surgery. Knee Surg Sports Traumatol Arthrosc Off J ESSKA. févr 2016;24(2):305-29.

19. Jordan RW, Saithna A. Physical examination tests and imaging studies based on arthroscopic assessment of the long head of biceps tendon are invalid. Knee Surg Sports Traumatol Arthrosc Off J ESSKA. oct 2017;25(10):3229-36.

20. Sasyniuk TM, Mohtadi NGH, Hollinshead RM, Russell ML, Fick GH. The inter-rater reliability of shoulder arthroscopy. Arthrosc J Arthrosc Relat Surg Off Publ Arthrosc Assoc N Am Int Arthrosc Assoc. sept 2007;23(9):971-7.

21. Loftus TJ, Tighe PJ, Filiberto AC, Efron PA, Brakenridge SC, Mohr AM, et al. Artificial Intelligence and Surgical Decision-making. JAMA Surg. 1 févr 2020;155(2):148-58.

22. Kitaguchi D, Takeshita N, Hasegawa H, Ito M. Artificial intelligence-based computer vision in surgery: Recent advances and future perspectives. Ann Gastroenterol Surg. janv 2022;6(1):29-36.

23. Walch G, Edwards TB, Boulahia A, Nové-Josserand L, Neyton L, Szabo I. Arthroscopic tenotomy of the long head of the biceps in the treatment of rotator cuff tears: clinical and radiographic results of 307 cases. J Shoulder Elbow Surg. juin 2005;14(3):238-46.

24. Plomb-Holmes C, Clavert P, Kolo F, Tay E, Lädermann A, French Arthroscopic Society. An orthopaedic surgeon's guide to ultrasound imaging of the healthy, pathological and postoperative shoulder. Orthop Traumatol Surg Res OTSR. déc 2018;104(8S):S219-32.

25. Armstrong A, Teefey SA, Wu T, Clark AM, Middleton WD, Yamaguchi K, et al. The efficacy of ultrasound in the diagnosis of long head of the biceps tendon pathology. J Shoulder Elbow Surg. févr 2006;15(1):7-11.

26. Taylor SA, Newman AM, Nguyen J, Fabricant PD, Baret NJ, Shorey M, et al. Magnetic Resonance Imaging Currently Fails to Fully Evaluate the Biceps-Labrum Complex and Bicipital Tunnel. Arthrosc J Arthrosc Relat Surg Off Publ Arthrosc Assoc N Am Int Arthrosc Assoc. févr 2016;32(2):238-44.

27. Nourissat G, Tribot-Laspiere Q, Aim F, Radier C. Contribution of MRI and CT arthrography to the diagnosis of intra-articular tendinopathy of the long head of the biceps. Orthop Traumatol Surg Res OTSR. déc 2014;100(8 Suppl):S391-394.

28. Rol M, Favard L, Berhouet J. Diagnosis of long head of biceps tendinopathy in rotator cuff tear patients: correlation of imaging and arthroscopy data. Int Orthop. juin 2018;42(6):1347-55.

29. Bhatnagar A, Bhonsle S, Mehta S. Correlation between MRI and Arthroscopy in Diagnosis of Shoulder Pathology. J Clin Diagn Res JCDR. févr 2016;10(2):RC18-21.

30. Thiagarajan A, Nagaraj R, Marathe K. Correlation Between Clinical Diagnosis, MRI, and Arthroscopy in Diagnosing Shoulder Pathology. Cureus. déc 2021;13(12):e20654.

31. Ahrens PM, Boileau P. The long head of biceps and associated tendinopathy. J Bone Joint Surg Br. août 2007;89(8):1001-9.

32. Neviaser RJ. Lesions of the biceps and tendinitis of the shoulder. Orthop Clin North Am. avr 1980;11(2):343-8.

33. Murthi AM, Vosburgh CL, Neviaser TJ. The incidence of pathologic changes of the long head of the biceps tendon. J Shoulder Elbow Surg. oct 2000;9(5):382-5.

34. Beall DP, Williamson EE, Ly JQ, Adkins MC, Emery RL, Jones TP, et al. Association of biceps tendon tears with rotator cuff abnormalities: degree of correlation with tears of the anterior and superior portions of the rotator cuff. AJR Am J Roentgenol. mars 2003;180(3):633-9.

35. Virk MS, Cole BJ. Proximal Biceps Tendon and Rotator Cuff Tears. Clin Sports Med. janv 2016;35(1):153-61.

36. Guy S, Jacquet C, Tsenkoff D, Argenson JN, Ollivier M. Deep learning for the radiographic diagnosis of proximal femur fractures: Limitations and programming issues. Orthop Traumatol Surg Res OTSR. avr 2021;107(2):102837.

37. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial Intelligence in Surgery: Promises and Perils. Ann Surg. juill 2018;268(1):70-6.

38. Chen JH, Asch SM. Machine Learning and Prediction in Medicine - Beyond the Peak of Inflated Expectations. N Engl J Med. 29 juin 2017;376(26):2507-9.

39. Patte D. Classification of rotator cuff lesions. Clin Orthop. mai 1990;(254):81-6.

40. World Medical Association declaration of Helsinki. Recommendations guiding physicians in biomedical research involving human subjects. JAMA J Am Med Assoc. 19 mars 1997;277(11):925-6.

41. Toussaint B, Barth J, Charousset C, Godeneche A, Joudet T, Lefebvre Y, et al. Nouvelle classification endoscopique des lésions du sub-scapulaire. Rev Chir Orthopédique Traumatol. déc 2012;98(8):S419-26.

42. Goutallier D, Postel JM, Bernageau J, Lavau L, Voisin MC. Fatty muscle degeneration in cuff ruptures. Pre- and postoperative evaluation by CT scan. Clin Orthop. juill 1994;(304):78-83.

43. Fuchs B, Weishaupt D, Zanetti M, Hodler J, Gerber C. Fatty degeneration of the muscles of

the rotator cuff: Assessment by computed tomography versus magnetic resonance imaging. J Shoulder Elbow Surg. nov 1999;8(6):599-605.

44. Boileau P, Ahrens PM, Trojani C, Coste JS, Cordéro B, Rousseau P. [Entrapment of the long head of the biceps: the « hourglass biceps ». Another cause of pain and locking of the shoulder]. Rev Chir Orthop Reparatrice Appar Mot. déc 2003;89(8):672-82.

45. Lafosse L, Reiland Y, Baier GP, Toussaint B, Jost B. Anterior and posterior instability of the long head of the biceps tendon in rotator cuff tears: a new classification based on arthroscopic observations. Arthrosc J Arthrosc Relat Surg Off Publ Arthrosc Assoc N Am Int Arthrosc Assoc. janv 2007;23(1):73-80.

46. Castagna A, Mouhsine E, Conti M, Vinci E, Borroni M, Giardella A, et al. Chondral print on humeral head: an indirect sign of long head biceps tendon instability. Knee Surg Sports Traumatol Arthrosc Off J ESSKA. mai 2007;15(5):645-8.

47. Domos P, Neogi DS, Longo UG, Ahrens PM. The chondral print sign: what does it really mean? J Shoulder Elbow Surg. juin 2017;26(6):e188-92.

48. Braun S, Horan MP, Elser F, Millett PJ. Lesions of the biceps pulley. Am J Sports Med. avr 2011;39(4):790-5.

49. Martetschläger F, Tauber M, Habermeyer P. Injuries to the Biceps Pulley. Clin Sports Med. janv 2016;35(1):19-27.

50. Sabottke CF, Spieler BM. The Effect of Image Resolution on Deep Learning in Radiography. Radiol Artif Intell. janv 2020;2(1):e190015.

51. Cohen J. A Coefficient of Agreement for Nominal Scales. Educ Psychol Meas. 1 avr 1960;20(1):37-46.

52. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics. 1977;33(1):159-74.

53. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther. mars 2005;85(3):257-68.

54. Bakeman R. KappaAcc: Deciding Whether Kappa is Big Enough by Estimating Observer Accuracy. Tech Rep. :9.

55. Godenèche A, Freychet B, Lanzetti RM, Clechet J, Carrillon Y, Saffarini M. Should massive rotator cuff tears be reconstructed even when only partially repairable? Knee Surg Sports Traumatol Arthrosc Off J ESSKA. juill 2017;25(7):2164-73.

56. Desai SS, Mata HK. Long Head of Biceps Tendon Pathology and Results of Tenotomy in Full-Thickness Reparable Rotator Cuff Tear. Arthrosc J Arthrosc Relat Surg Off Publ Arthrosc Assoc N Am Int Arthrosc Assoc. nov 2017;33(11):1971-6.

57. Yuan Y, Tang Y. Fusion of expert uncertain assessment in FMEA based on the negation of basic probability assignment and evidence distance. Sci Rep. 19 mai 2022;12(1):8424.

58. Dempster AP. Upper and Lower Probabilities Induced by a Multivalued Mapping. Ann Math Stat. avr 1967;38(2):325-39.

59. Kurmis AP, Ianunzio JR. Artificial intelligence in orthopedic surgery: evolution, current state and future directions. Arthroplasty Lond Engl. 2 mars 2022;4(1):9.

60. Hui AT, Alvandi LM, Eleswarapu AS, Fornari ED. Artificial Intelligence in Modern Orthopaedics: Current and Future Applications. JBJS Rev. 1 oct 2022;10(10).

61. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 28 mai 2015;521(7553):436-44.

62. Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, et al. Mastering the game of Go without human knowledge. Nature. 18 oct 2017;550(7676):354-9.

63. Lambrechts A, Wirix-Speetjens R, Maes F, Van Huffel S. Artificial Intelligence Based Patient-Specific Preoperative Planning Algorithm for Total Knee Arthroplasty. Front Robot AI. 2022;9:840282.

64. Jayakumar P, Moore MG, Furlough KA, Uhler LM, Andrawis JP, Koenig KM, et al. Comparison of an Artificial Intelligence-Enabled Patient Decision Aid vs Educational Material on Decision Quality, Shared Decision-Making, Patient Experience, and Functional Outcomes in Adults With Knee Osteoarthritis: A Randomized Clinical Trial. JAMA Netw Open. 1 févr 2021;4(2):e2037107.

65. von Eisenhart-Rothe R, Hinterwimmer F, Graichen H, Hirschmann MT. Artificial intelligence and robotics in TKA surgery: promising options for improved outcomes? Knee Surg Sports Traumatol Arthrosc Off J ESSKA. août 2022;30(8):2535-7.

66. Kumar V, Roche C, Overman S, Simovitch R, Flurin PH, Wright T, et al. Using machine learning to predict clinical outcomes after shoulder arthroplasty with a minimal feature set. J Shoulder Elbow Surg. mai 2021;30(5):e225-36.

67. Belk JW, Kraeutler MJ, Houck DA, Chrisman AN, Scillia AJ, McCarty EC. Biceps tenodesis versus tenotomy: a systematic review and meta-analysis of level I randomized controlled trials. J Shoulder Elbow Surg. mai 2021;30(5):951-60.

68. MacDonald P, Verhulst F, McRae S, Old J, Stranges G, Dubberley J, et al. Biceps Tenodesis Versus Tenotomy in the Treatment of Lesions of the Long Head of the Biceps Tendon in Patients Undergoing Arthroscopic Shoulder Surgery: A Prospective Double-Blinded Randomized Controlled Trial. Am J Sports Med. mai 2020;48(6):1439-49.

69. Godenèche A, Kempf JF, Nové-Josserand L, Michelet A, Saffarini M, Hannink G, et al. Tenodesis renders better results than tenotomy in repairs of isolated supraspinatus tears with pathologic biceps. J Shoulder Elbow Surg. nov 2018;27(11):1939-45.

70. Castricini R, Familiari F, De Gori M, Riccelli DA, De Benedetto M, Orlando N, et al. Tenodesis is not superior to tenotomy in the treatment of the long head of biceps tendon lesions. Knee Surg Sports Traumatol Arthrosc Off J ESSKA. janv 2018;26(1):169-75.

71. van Deurzen DFP, Auw Yang KG, Onstenk R, Raven EEJ, van den Borne MPJ, Hoelen MA, et al. Long Head of Biceps Tenotomy Is Not Inferior to Suprapectoral Tenodesis in Arthroscopic Repair of Nontraumatic Rotator Cuff Tears: A Multicenter, Non-inferiority, Randomized, Controlled Clinical Trial. Arthrosc J Arthrosc Relat Surg Off Publ Arthrosc Assoc N Am Int Arthrosc Assoc. juin 2021;37(6):1767-1776.e1.

72. Al-Ani Z, Lauder J. Ultrasound assessment in distal biceps tendon injuries: Techniques, pearls and pitfalls. Clin Imaging. juill 2021;75:46-54.

73. Kim JY, Rhee SM, Rhee YG. Accuracy of MRI in diagnosing intra-articular pathology of the long head of the biceps tendon: results with a large cohort of patients. BMC Musculoskelet Disord. 1 juin 2019;20(1):270.

74. Whelehan DF, Conlon KC, Ridgway PF. Medicine and heuristics: cognitive biases and medical decision-making. Ir J Med Sci. nov 2020;189(4):1477-84.

75. Loftus TJ, Filiberto AC, Li Y, Balch J, Cook AC, Tighe PJ, et al. Decision analysis and

reinforcement learning in surgical decision-making. Surgery. août 2020;168(2):253-66.

76. Doyle DJ, Goyal A, Bansal P, Garmon EH. American Society of Anesthesiologists Classification. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021 [cité 19 déc 2021]. Disponible sur: http://www.ncbi.nlm.nih.gov/books/NBK441940/

77. Groll DL, To T, Bombardier C, Wright JG. The development of a comorbidity index with physical function as the outcome. J Clin Epidemiol. juin 2005;58(6):595-602.

78. Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. Clin Orthop. janv 1987;(214):160-4.

79. Gilbart MK, Gerber C. Comparison of the subjective shoulder value and the Constant score. J Shoulder Elbow Surg. déc 2007;16(6):717-21.

80. Baumgarten KM, Chang PS. The American Shoulder and Elbow Surgeons score highly correlates with the Simple Shoulder Test. J Shoulder Elbow Surg. avr 2021;30(4):707-11.

81. Moor BK, Bouaicha S, Rothenfluh DA, Sukthankar A, Gerber C. Is there an association between the individual anatomy of the scapula and the development of rotator cuff tears or osteoarthritis of the glenohumeral joint?: A radiological study of the critical shoulder angle. Bone Jt J. juill 2013;95-B(7):935-41.

82. McLean A, Taylor F. Classifications in Brief: Bigliani Classification of Acromial Morphology. Clin Orthop. août 2019;477(8):1958-61.

83. Morid MA, Borjali A, Del Fiol G. A scoping review of transfer learning research on medical image analysis using ImageNet. Comput Biol Med. janv 2021;128:104115.

84. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision [Internet]. arXiv; 2015 [cité 18 oct 2022]. Disponible sur: http://arxiv.org/abs/1512.00567

85. Olczak J, Fahlberg N, Maki A, Razavian AS, Jilert A, Stark A, et al. Artificial intelligence for analyzing orthopedic trauma radiographs. Acta Orthop. déc 2017;88(6):581-6.

86. Adams M, Chen W, Holcdorf D, McCusker MW, Howe PD, Gaillard F. Computer vs human: Deep learning versus perceptual training for the detection of neck of femur fractures. J Med Imaging Radiat Oncol. févr 2019;63(1):27-32.

87. Baro E, Degoul S, Beuscart R, Chazard E. Toward a Literature-Driven Definition of Big Data in Healthcare. BioMed Res Int. 2015;2015:639021.

88. Jee K, Kim GH. Potentiality of big data in the medical sector: focus on how to reshape the healthcare system. Healthc Inform Res. juin 2013;19(2):79-85.

89. Lawson EH, Louie R, Zingmond DS, Sacks GD, Brook RH, Hall BL, et al. Using Both Clinical Registry and Administrative Claims Data to Measure Risk-adjusted Surgical Outcomes. Ann Surg. janv 2016;263(1):50-7.

90. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN [Internet]. arXiv; 2018 [cité 9 oct 2022]. Disponible sur: http://arxiv.org/abs/1703.06870

91. Opfer R, Krüger J, Spies L, Ostwaldt AC, Kitzler HH, Schippling S, et al. Automatic segmentation of the thalamus using a massively trained 3D convolutional neural network: higher sensitivity for the detection of reduced thalamus volume by improved inter-scanner stability. Eur Radiol. 20 oct 2022;

Vu, le Directeur de Thèse

Vu, le Doyen
De la Faculté de Médecine de Tours
Tours, le

université de TOURS | Faculté de médecine

Thèse 2021/2022

DOCTORAT en MEDECINE

Diplôme d'Etat

D.E.S. de chirurgie orthopédique et traumatologique
Présentée et Soutenue le 27 octobre 2022
Dépôt de sujet de thèse, proposition de jury,

NOM : BENHENNEDA

Prénom : Rayane

Date de naissance : 29 décembre 1993 **Nationalité : Français**

Lieu de naissance : Colombes, 92

Domicile : 20 rue Victor Hugo 37000 Tours **Téléphone : 06 58 39 82 52**

Directeur de Thèse : Professeur Julien BERHOUET

Titre de la Thèse :
L'intelligence artificielle peut-elle aider à la décision en arthroscopie ?
Partie 1 : Évaluation d'une analyse inter-observateur d'un protocole vidéo diagnostique du tendon du long biceps.
Partie 2 : Réflexion à partir d'un modèle appliqué à la gestion du long biceps dans les ruptures distales du tendon supra-épineux.

JURY

Président du Jury :
Professeur Luc FAVARD, Chirurgie orthopédique, Faculté de médecine - Tours

Membres du Jury :
Professeur Julien BERHOUET, Chirurgie orthopédique, Faculté de médecine - Tours
Monsieur Thierry BROUARD, Département d'informatique, École Polytechnique - Tours
Docteur Louis-Romée LE NAIL, Chirurgie orthopédique, PH, CHRU Tours
Professeur Philippe ROSSET, Chirurgie orthopédique, Faculté de médecine – Tours

Avis du Directeur de Thèse
À Tours, le 22/09/2022

Signature

Pr. J.BERHOUET
Chirurgie Orthopédique / Traumatologie
C.H.R.U TOURS
37044 TOURS CEDEX 9
RPPS 10100182714

Avis du Directeur de l'U.F.R. Tours
à Tours, le 4 10 22

Signature

BENHENNEDA Rayane

112 pages – 5 tableaux – 11 figures

L'intelligence artificielle peut-elle aider à la décision en arthroscopie?

Introduction : L'intelligence artificielle est un atout prometteur en chirurgie orthopédique. Le deep learning est utilisable en chirurgie arthroscopique grâce au signal vidéo exploitable par computer vision. La gestion per-opératoire du tendon du long biceps (TLB) est l'objet d'une controverse ancienne. L'objectif principal de cette étude était de modéliser une IA diagnostique capable de statuer sur le caractère sain ou pathologique du TLB sur des images arthroscopiques. L'objectif secondaire était de créer un deuxième modèle IA diagnostique en fonction d'images arthroscopiques et des données médicales de chaque patient, cliniques et imagerie, pour statuer sur le caractère sain ou pathologique du TLB. **Hypothèse :** L'hypothèse de l'étude était qu'il était possible de construire un modèle IA à partir d'images opératoires d'arthroscopie d'aide au diagnostic du caractère sain ou pathologique du TLB. **Matériels et Méthodes :** Les données cliniques et imagerie de 199 patients ont été collectées de manière prospectives, et associées aux images issues d'une analyse vidéo arthroscopique protocolisée validée, dite « vérité terrain », faite par le chirurgien opérateur. Un modèle basé sur un réseau neuronal convolutif (RNC) modélisé via un *transfer learning* sur le modèle Inception V3 a été construit pour l'analyse des images arthroscopiques. Ce modèle a ensuite été couplé à un Perceptron MultiCouche (*MultiLayer Perceptron, MLP*) intégrant les données cliniques et imageries. Chaque modèle a été entraîné et testé selon apprentissage supervisé. **Résultats :** La précision du RNC sur le diagnostic du statut sain ou pathologique du TLB était de 93,7% en apprentissage et 80,66% en généralisation. Couplé aux données cliniques de chaque patient, la précision du modèle assemblant le RNC et un MLP étaient respectivement de 77% et 58% en apprentissage et en généralisation. **Conclusion :** Le modèle IA construit à partir d'un RNC parvient a déterminer le caractère sain ou pathologique du TLB avec un taux de précision de 80,66%. Ce modèle n'est pas suffisamment pertinent pour une application en pratique clinique. Une augmentation des données d'entrée pour limiter l'overfitting, et l'automatisation de la phase de détection par un *Mask-R-CNN* sont des voies d'amélioration du modèle. Cette étude est la première à évaluer la capacité d'une IA à analyser des images arthroscopiques, et ses résultats doivent être confirmés par d'autres études sur le sujet.

Mots-clés : arthroscopie ; intelligence artificielle ; réseau neuronal convolutif ; tendon du long biceps ; apprentissage profond

Jury :

Président du Jury : Professeur Luc Favard

Directeur de thèse : Professeur Julien Berhouet

Membres du Jury : Professeur Philippe Rosset, Dr Louis-Romée Le Nail, Mr Thierry Brouard

Date de soutenance : le jeudi 27 octobre 2022