Orthopaedics & Traumatology: Surgery & Research 109 (2023) 103652

Contents lists available at ScienceDirect

Orthopaedics & Traumatology: Surgery & Research

journal homepage: www.elsevier.com

Original article

Can artificial intelligence help decision-making in arthroscopy? Part 2: The IA-RTRHO model – a decision-making aid for long head of the biceps diagnoses in small rotator cuff tears

Rayane Benhenneda[a,*], Thierry Brouard[b], Christophe Charousset[c], Julien Berhouet[a,b], The Francophone Arthroscopy Society (SFA)

[a] Service de chirurgie orthopédique, hôpital Trousseau, CHRU de Tours, faculté de médecine, université de Tours, Centre-Val-de-Loire, France
[b] LIFAT (EA6300), école polytechnique universitaire de Tours, 64, avenue Jean-Portalis, 37200 Tours, France
[c] Clinique Turin, 9, rue de Turin, 75008 Paris, France

ARTICLE INFO

Article history:
Received 17 March 2023
Accepted 17 May 2023

Keywords:
Arthroscopy
Artificial intelligence
Convolutional neural network
Long head of biceps
Deep learning

ABSTRACT

Introduction: The possible applications of artificial intelligence (AI) in orthopedic surgery are promising. Deep learning can be utilized in arthroscopic surgery due to the video signal used by computer vision. The intraoperative management of the long head of biceps (LHB) tendon is the subject of a long-standing controversy. The main objective of this study was to model a diagnostic AI capable of determining the healthy or pathological state of the LHB on arthroscopic images. The secondary objective was to create a second diagnostic AI model based on arthroscopic images and the medical, clinical and imaging data of each patient, to determine the healthy or pathological state of the LHB.
Hypothesis: The hypothesis of this study was that it was possible to construct an AI model from operative arthroscopic images to aid in the diagnosis of the healthy or pathological state of the LHB, and its analysis would be superior to a human analysis.
Materials and methods: Prospective clinical and imaging data from 199 patients were collected and associated with images from a validated protocoled 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 MultiLayer Perceptron (MLP), integrating clinical and imaging data. Each model was trained and tested using supervised learning.
Results: The accuracy of the CNN in diagnosing the healthy or pathological state of the LHB 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 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 state of the LHB with an accuracy rate of 80.66%. 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 further studies on this subject.
Level of evidence: III Diagnostic study.

1. Introduction

Artificial intelligence (AI) is the ability of a machine to simulate human intelligence. Its ability to manage and optimize large data sets makes it a promising asset in the health sector, particularly in orthopedic surgery [1]. Deep learning is a learning method which aims to model data with a high level of abstraction. Deep learning techniques have reached levels of performance equivalent to, or even superior to, human performance [2]. Research endeavors

* Corresponding author at: Service de chirurgie orthopédique, hôpital Trousseau, CHRU de Tours, faculté de médecine, université de Tours, avenue de la République, Chambray-lès-Tours, 37044 Tours cedex 9, France.
E-mail address: rayane.benhenneda@gmail.com (R. Benhenneda).

https://doi.org/10.1016/j.otsr.2023.103652
1877-0568/

focused on the application of AI in orthopedic surgery are currently focused on 3 areas: preoperative planning [3], intraoperative navigation and guidance [4], and prediction of postoperative clinical outcomes [5]. In arthroscopic surgery, the camera produces a visual signal used for diagnosis and decision-making that can be used in AI by computer vision [6].

Arthroscopy requires a long and difficult learning curve [7], due to its specific view, which is anatomically different to open surgery, and its precise gestures based on triangulation in a reduced space [8]. Its technical specificities make arthroscopy an experience-dependent discipline [9], whose sensitivity and specificity of intraoperative diagnostic analysis for a given anatomical structure remains debatable [10]. It is indeed the reliability of intraoperative decision-making that can be called into question, making it more difficult and demanding because it is potentially a source of errors and avoidable complications in practice.

The intraoperative management of the long head of biceps (LHB) tendon has long been the subject of controversy in the context of arthroscopic management of rotator cuff tears [11]. While the technical considerations between tenotomy or tenodesis appear agreed upon in the context of obvious LHB lesions in large to massive rotator cuff lesions [12,13], the therapeutic indication remains debatable in the event of distal or small tears [14,15]. The LHB is indeed frequently intact in this context, or otherwise presents with minor lesions, making the preoperative diagnostic evaluation difficult, both during the physical examination of the patient and on imaging.

The main objective of this study was to propose a novel experimental AI model to aid decision-making with regards to the healthy or pathological state of the LHB in the context of small distal rotator cuff tears from arthroscopic images. The secondary objective was to prepare a second AI model determining the healthy or pathological state of the LHB, but in this case from arthroscopic images and preoperative clinical and imaging data. The hypothesis of the study was that it was possible to build an efficient AI model from operative arthroscopic images, capable of helping define the healthy or pathological state of the LHB.

2. Material and methods

This is a prospective, multicenter, randomized, single-blind study. Informed consent was obtained from all the patients included (ID-RCB 2018-A01382-53).

2.1. Database for the experimental AI model: "ground truth"

All the data used for the construction of the AI model was collected prospectively during the inclusion of patients and their follow-up. For each patient included, this involved clinical data, additional examinations and arthroscopic images.

2.1.1. Clinical data

The clinical data collected pertained to the information gathered during history taking (sex, age, dominant side, ASA score [16], Body Mass Index [BMI], medical history, Functional Comorbidity Index [FCI] [17], profession, participation in sports, history of trauma on the operated shoulder) and clinical examination (preoperative Constant score [18], Subjective Shoulder Value [SSV] score [19], American Shoulder and Elbow surgeons Score [ASES] score [20], Visual Analogue Scale [VAS] for pain, and clinical diagnostic tests). All the clinical data are summarized in Table 1 and the scores in Table 2.

2.1.2. Imaging data

Paraclinical data included preoperative imaging. Each patient underwent:

Table 1
Demographics of the study population.

Category	Number (%)
Patients	199
Operated shoulder	
Left side	64 (32.2)
Right side	135 (67.8)
Operated side=dominant shoulder	137 (68.8)
Sex	
Male	95 (47.7)
Female	104 (52.3)
Mean age, years	55.3 ± 8.9 (33–78)
Body Mass Index (BMI)	28.7 ± 3.7 (18–42.8)
American Society of Anesthesiologists (ASA) score [16]	
ASA 1	131 (65.8)
ASA 2	57 (28.6)
ASA 3	11 (5.5)
ASA 4	0
Comorbidities	
Active smoker	28 (14.1)
Hypercholesterolemia	35 (17.6)
Diabetes	12 (6.0)
Hyperuricemia	3 (1.5)
Functional Comorbidity Index (FCI) [17]	0.35 (0–4)
Social	
Patient retired at the time of the intervention	57 (28.6)
Work accident or occupational disease	35 (17.6)
Practiced sport before surgery (leisure)	70 (35.2)
History of trauma to the operated shoulder (< 12 months)	60 (30.2)

Table 2
Preoperative clinical scores of the study population.

Category	Average (range)
Constant score [18]	
Pain	6.79 ± 2.72 (0–15)
Activity	10.45 ± 3.57 (1–20)
Range of motion	32.42 ± 6.99 (8–40)
Strength	6.07 ± 4.45 (0–22)
Total	55.71 ± 13.69 (21–86)
Strength	
SSV Score [19]	48.61 ± 16.11 (10–90)
ASES Score	43.55 ± 20.79 (6–92)
VAS Pain (0–10)	5.45 ± 1.92 (0–10)

- standard AP and lateral radiographs: Critical Shoulder Angle (CSA) [21] and Bigliani classification [22];
- MRI: size of the supraspinatus tear in the frontal and sagittal plane, presence or absence of an infraspinatus or a subscapularis lesion/tear, space in the bicipital groove, diameter and position of the LHB, presence or absence of signal abnormality in T1 or T2;
- ultrasound by a radiologist: size, surface, grade of vascularization, position and echogenicity of the LHB evaluated at 2 locations: in the bicipital groove and at the upper part of it.

2.1.3. Arthroscopy images

The first part of this study made it possible to validate a database of images constituting the input data of the AI program. Each video was segmented into 4 images visualizing the LHB at 4 areas of interest: (1) glenoid insertion, (2) proximal intra-articular portion, (3) distal intra-articular portion, and (4) bicipital groove. The segmenting of the regions of interest (ROI), was done by bounding boxes. Furthermore, the initial part of this study made it possible, to validate the use of images to characterize the healthy or pathological state of the LHB and to carry out the pathological analysis, and subsequently determine that the "ground truth" analysis by the operating surgeon was the most accurate. Thus, this served as a training base.

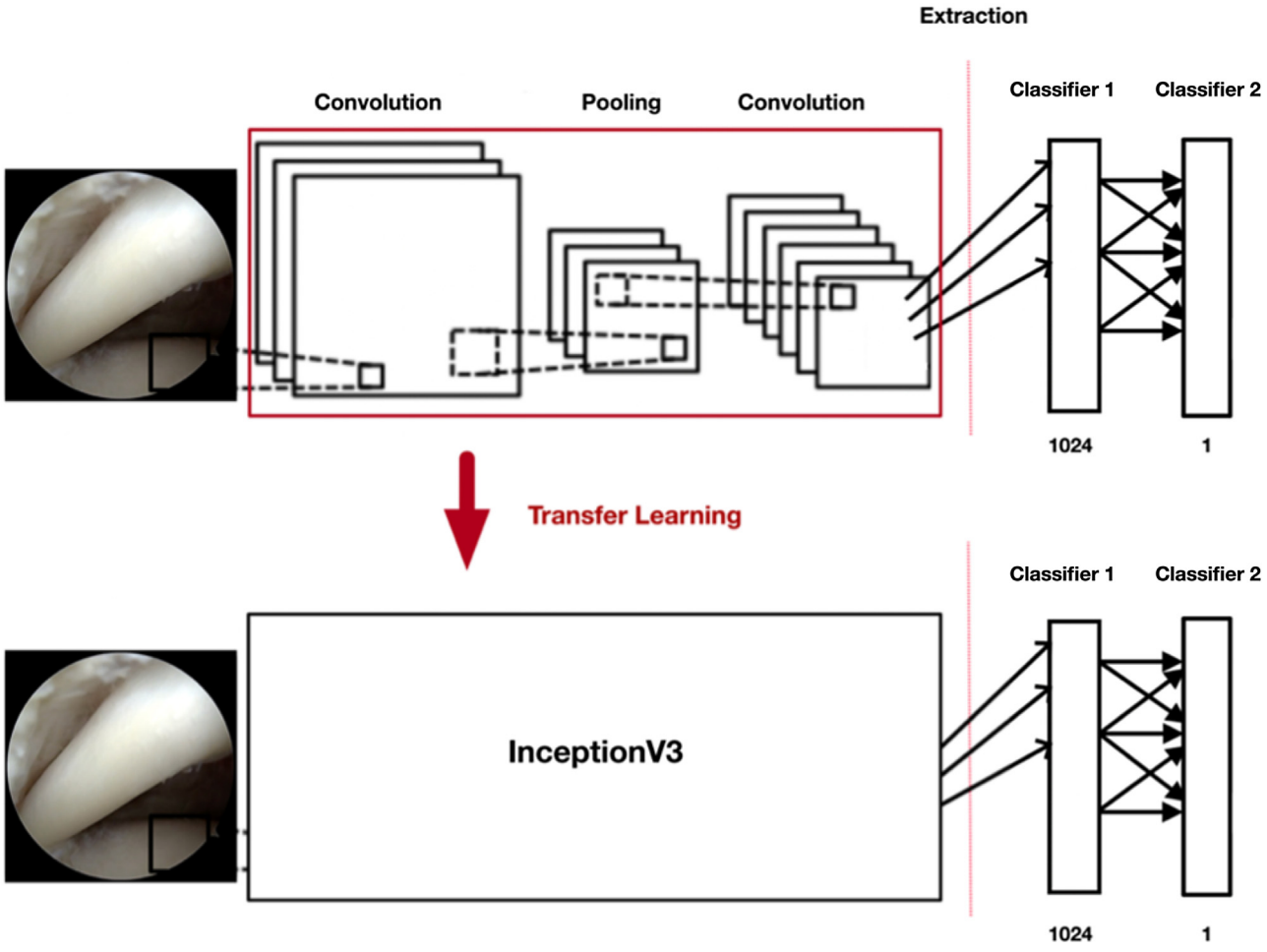

Fig. 1. Architecture of the constructed convolutional neural network.

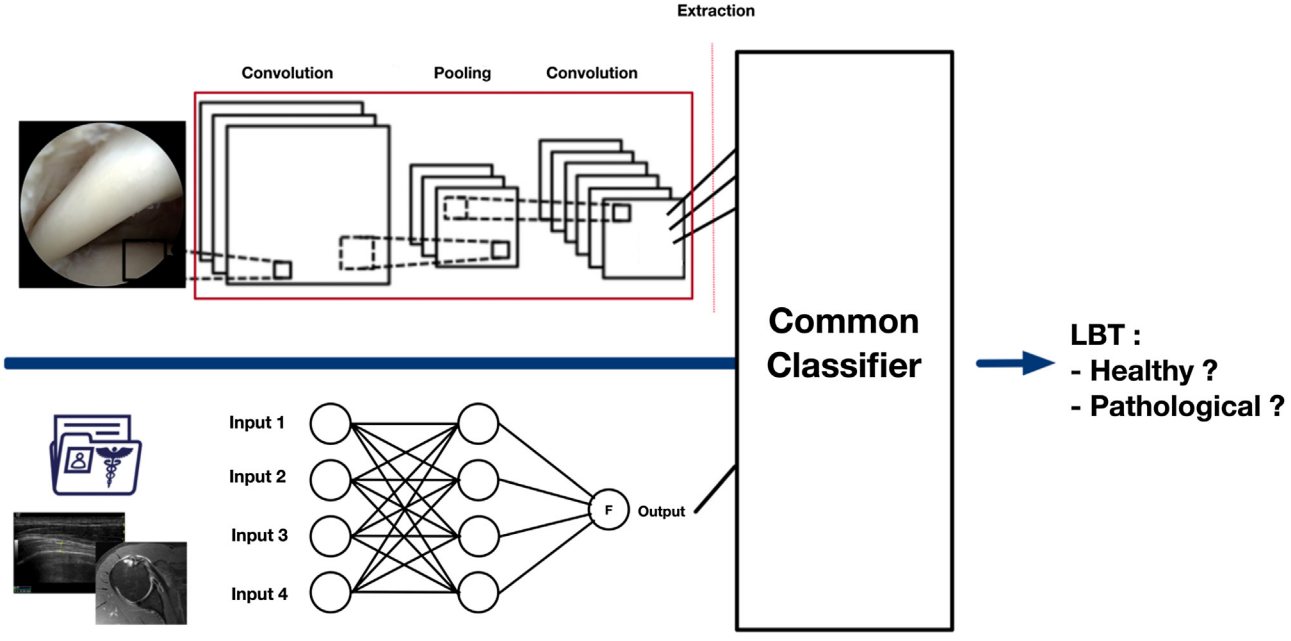

Fig. 2. Architecture of the constructed model, based on CNN and MLP. F: ReLU activation function. LHB: long head of the biceps tendon.

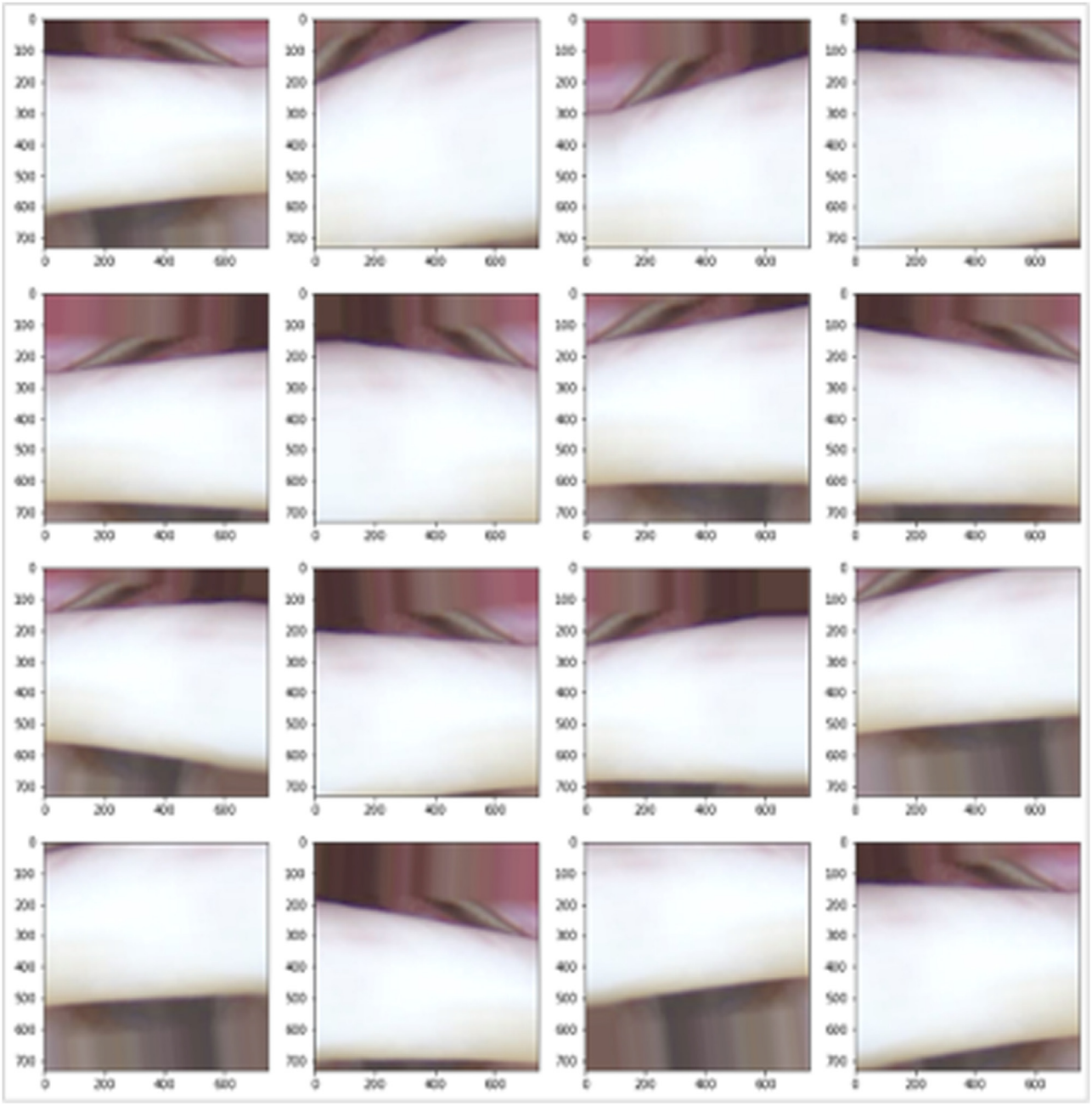

Fig. 3. Example of transformations undergone by one of the images of the learning base.

2.2. Construction of the AI model

The diagnostic AI model was formed by merging two prior, separately built, models. The initial model, based on a convolutional neural network (CCN), was built to process arthroscopic images, via transfer learning [23]. We chose the InceptionV3 model [24], trained on the ImageNet database and available in TensorFlow™ 2.0 (Google Brain, Santa Clara, USA, tensorflow.org). The CNN is schematized in Fig. 1. A second model based on a Multilayer Perceptron (MLP) was built to process medical data (clinical and paraclinical data). A third model associating the CNN of the first model with the MLP was created, using as much information as possible (medical data and arthroscopic images) to make a decision (Fig. 2). The coding was done in Python language for each of the models.

2.3. Training AI models

The training of each model was done in two stages:

- learning, consisting of testing the knowledge of the algorithm verbatim, by presenting it only with data for which it already knows the answer;
- generalization, consisting of testing the ability of the algorithm to give an answer on data that it has not encountered in the past.

2.3.1. Implementation

The learning of the image detection function was done by transfer learning [23] using the pre-trained InceptionV3 model collected on TensorFlow™ 2.0. The classification function was trained by supervised learning, using the “ground truth” data.

2.3.2. Hyperparameters

The CNN learning rate alpha chosen for the descent of the gradient as the most optimal was 10−4. The 1D input data of the MLP was subjected to a 3-layer artificial neural network with a ReLU (Rectified Linear Unit) activation function.

2.3.3. Preprocessing

The training base consisted of 580 images (243 "pathological" LHB and 337 "normal" LHB). The training set consisted of 90% of the images, and the testing set consisted of 10% of the images. In order to overcome the low number of training images and to avoid the over-fitting phenomenon, an automatic augmentation of the training data was used with the following parameters:

- rotate: 10%;
- stretch in width: 20%;
- stretch in height: 20%;
- zoom: 20%;
- horizontal symmetry.

The model therefore had the possibility of randomly producing images that were added during the training phase using one or more of these transformations (Fig. 3).

2.3.4. Training

Learning was done in batches of 20 images over 200 periods. During a period, the model learned all the images of the learning base in groups of 20 randomly formed images, and sought to predict the state (healthy or pathological) of the images it had just learned. Then it composed random batches of 20 images from the test database and evaluated the number of errors made on the images that it had not learned.

2.3.5. Test

The implementation was carried out online, on the Google Cloud software (Google Inc, Mountain View, United States) on GPU (Graphics Processing Units) architecture.

Fig. 4 schematizes each of the 3 models constructed.

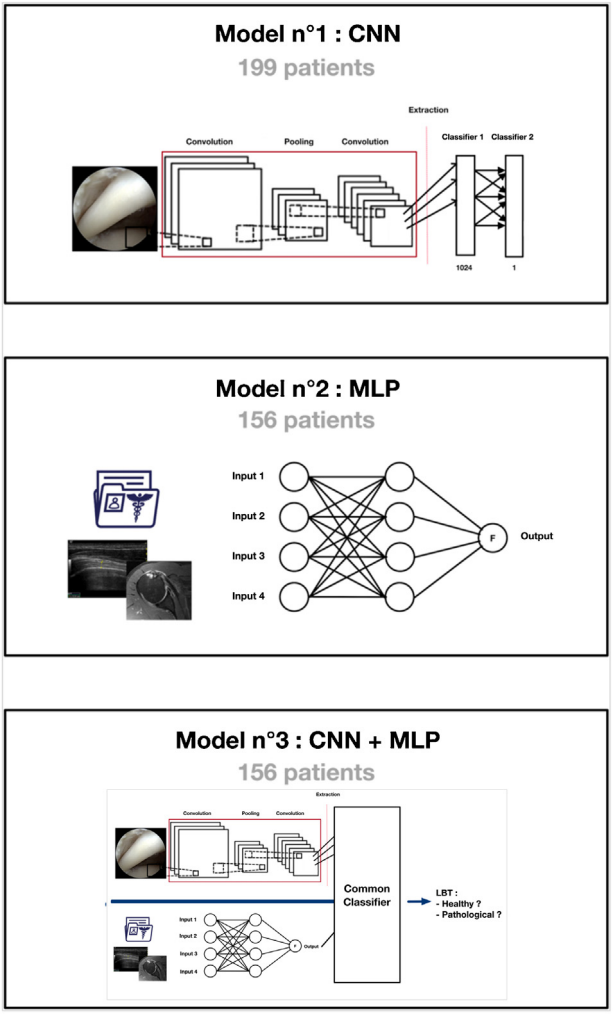

Fig. 4. Summary of the 3 AI models designed. AI: Artificial Intelligence; CNN: Convolutional Neural Network; MLP: MultiLayer Perceptron; LHB: long head of the biceps tendon.

3. Results

3.1. CNN Accuracy

The evolution of the model's accuracy during the learning stages (200 periods) increased, with an initial precision in the order of 60% to reach a plateau at the end of the process to between 90 and 95%. The average accuracy over the last 5 periods was 93.7% for learning and 80.66% for generalization (Fig. 5).

The duration of a period averaged 13 seconds, and the total training time was 45 minutes.

3.2. Model #2 accuracy

The accuracy of learning and generalization at the end of the process (500 periods) were respectively 85% and 75%. The total learning time was 2 minutes (Fig. 6).

3.3. Model #3 accuracy

The accuracy of learning and generalization at the end of the process (2000 periods) were respectively 77% and 58% (Fig. 7).

4. Discussion

We reported the results of an original AI model, intended to determine the healthy or pathological state of the LHB based on images from arthroscopic videos. Our hypothesis was confirmed. The first model managed to determine the healthy or pathological state of the LHB with an accuracy rate of 80.66%. The learning set had correct predictions at a rate of close to 95%.

This demonstrates the ability of the learning algorithm to discriminate the two categories of images accurately. According to the learning parameters, a random response resulted in a 243/580 = 41.9% chance of recognizing a LHB as pathological and therefore 58.1% of qualifying it as normal. The evolution of the accuracy, calculated on the test basis, highlighted three notable elements: (1) Its tendency followed the accuracy of the learning curve, which showed an ability of the model to generalize, i.e. to correctly predict unlearned images based on what the algorithm had learned; (2) Its performance remained behind that of learning, which was expected with this type of model; (3) The oscillating nature of the test curve was due to the low number of images and the fact that these were not increased, unlike the learning base. The homogeneity of batches sent for testing set accuracy assessment was low, leading to lower batch scores. On average, the accuracy of the testing set was 75% over the last 75 periods. Given the high variability of the images and their low number, the model used conferred very good learning results. In generalization, the performances were also very good and the difference was easily explained by the quantity of images available and their heterogeneity in the scenes represented.

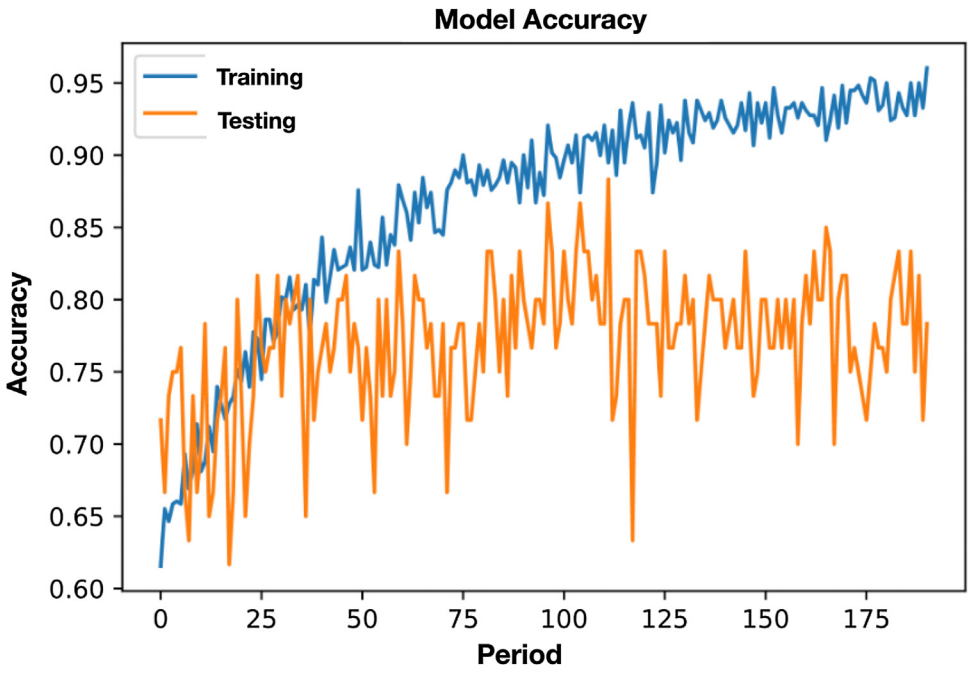

Fig. 5. Accuracy of model #1 (CNN). In blue, learning curve (training set) and in orange, test curve (testing set).

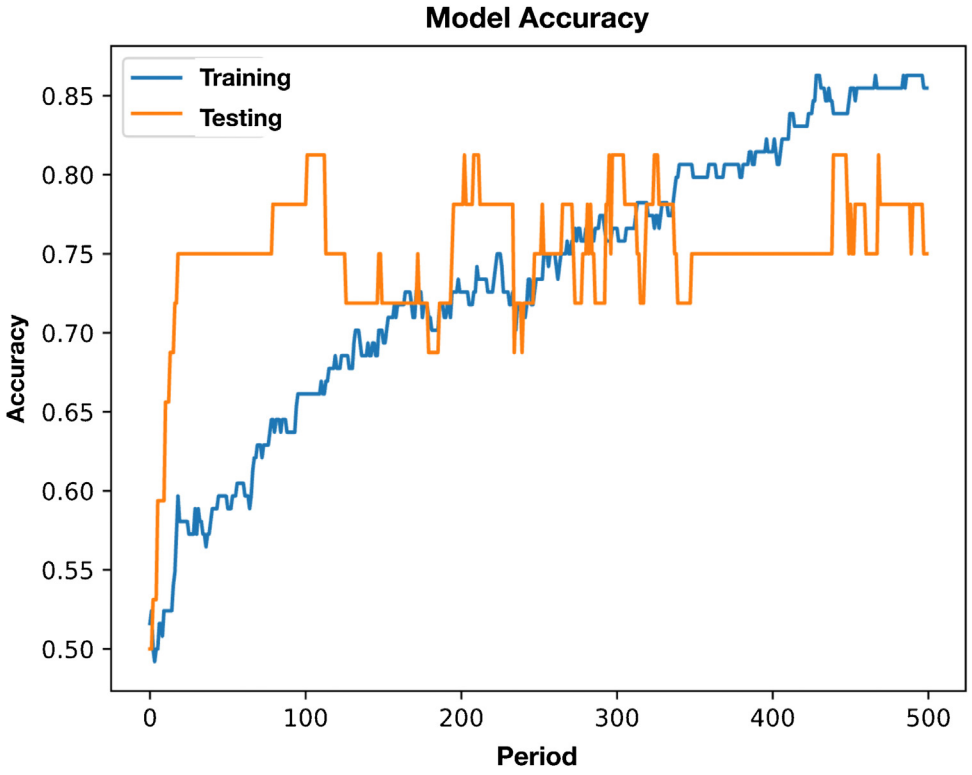

Fig. 6. Accuracy of Model #2 (MLP). In blue, learning curve (training set) and in orange, test curve (testing set).

Concerning second model, the oscillating character of the test curve highlighted the low number of input data, allowing however to improve the generalization capacity of the model up to 75% at the end of the process. This reflected a case of over-fitting which was explained by a limited amount of input data. A random response resulted for this model in a 50% chance of judging the LHB to be normal and 50% as pathological.

The third model did not seem to benefit from the association of image and medical data, since its accuracy was not as good as each of the first two models considered separately. A random response resulted for this model in a 50% chance of considering the LHB as normal and 50% as pathological. The accuracy rate of this model was 58%. An initial hypothesis for this was that for this model, each patient was either in the learning base (training set) or in the test

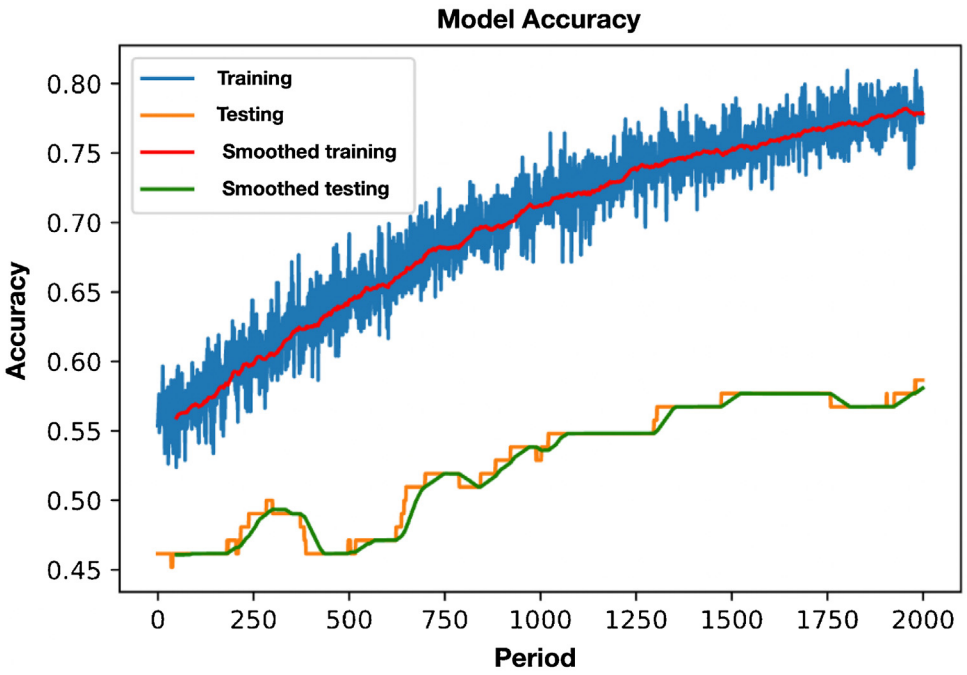

Fig. 7. Accuracy of model #3 (assembly of models 1 and 2). In blue, learning curve (training set) and in orange, test curve (testing set).

base (testing set). In the first model, the machine had the choice, on 5 images of the same patient, to use 3 for learning and 2 for the test. Performance was therefore lower in this more restrictive case where the tests only concerned disjoint patients. A second hypothesis was that the addition of medical data, for which the second model presented an obvious over-fitting, limited the whole of the third model in its performance (by making it itself subject to over-fitting). This medical data could add additional noise which, due to the small number of data, reduced the performance of the whole.

To be relevant in clinical practice, a program must have a performance equal to or greater than that of an experienced clinician. The inter-observer accuracy rates calculated in the first part of this work were, from 86 to 92%. With an accuracy of 80.66% for the first image-based model alone, the AI manages to perform at a level close to human analysis. This shows that it adapts very well to a purely technical analysis. This accuracy decreased to 58% when the image and medical data (clinical and imaging) were coupled. The performance of AI seems to decrease when a subjective part, through the clinical examination, enters the decision-making process. These results are also explained by the low number of data used to train these three models, despite image augmentation methods.

Despite the limited performance of the 3 models described, this is the first study reporting the results of a deep learning algorithm trained to identify lesions on arthroscopic images, with encouraging results. Guy et al. reported a sensitivity and specificity of 61 and 67% respectively for an algorithm trained to diagnose femoral neck fractures on two radiographic views with a database of 1309 images [25]. With a database of 256,000 images, Olczak et al. obtained an accuracy of 83% fracture detection on different radiographs (wrist, hand and ankle) [26]. In these two cases, the CNN used had not been previously trained on non-medical images. Using transfer learning, Adams et al. obtained an accuracy of 90.5% in detecting a femoral neck fracture on a radiographic view with a database of 640 images [27]. Similarly, Kim et al. with 11,112 images analyzed by a CNN modeled via transfer learning on the Inception V3 model, obtained a sensitivity of 0.9 and a specificity of 0.88 for the detection of fractures in radiographic imaging. This highlights an initial limitation in the use of AI in orthopedic surgery: the lack of massive databases (big data), which is not specific to our specialty [28,29]. This big data can for the moment come from clinical registers or administrative data [30] to reduce overfitting problems while waiting for new solutions. For the moment, transfer learning seems to be a partial solution to this lack of data but it is not enough on its own to achieve accuracy rates equivalent to the human clinician.

The technologies used by the algorithms are also an area for improvement. Indeed, our model used a simple CNN. It has been shown that with more complex R-CNN of recent development, the Mask-R-CNN [31], the performance levels of the algorithms improve [32]. The need for manual segmentation of the images by an expert, to ensure the model has supervised learning, is also a limit factor in the constitution of large databases. The R-CNN and Mask-R-CNN overcome this by automating the segmentation of an image (Fig. 8).

Our study has several limitations:

- a small number of data, both arthroscopic images (199 patients) and clinical or paraclinical images (156 patients) explaining the limited level of performance of the 3 models studied;
- images for supervised training were labeled by a single clinician, junior surgeon;
- it was not possible to carry out a weighting analysis of the various medical factors to anticipate the healthy or pathological state of the LHB.

However, our study has positive points. A transfer learning method was used to improve the performance of AI models, which seems to be an essential prerequisite compared to models without transfer learning reported in the literature. A qualitative validation of the data-set used for the construction of the AI model was carried out. The inter-observer analysis of arthroscopy images made in the first part of this work was indeed an essential prerequisite for the creation of an AI, whose performance is closely linked to the quality of its input data.

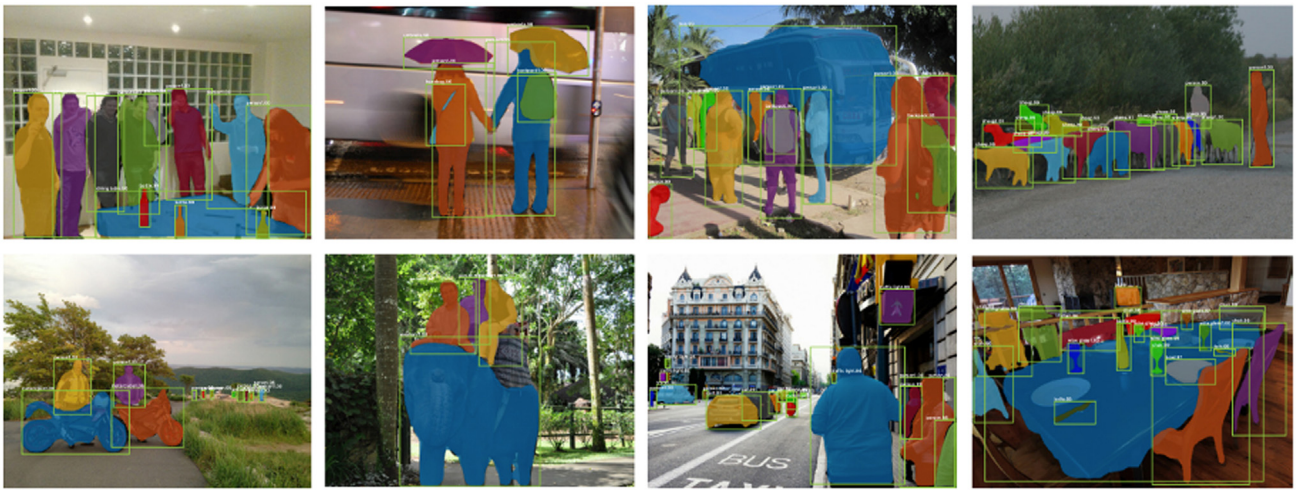

Fig. 8. Automatic segmentation of images by a Mask-R-CNN [31].

5. Conclusion

The analysis of the healthy or pathological state of the LHB by an AI based on arthroscopic images is possible with a diagnostic accuracy of up to 80%. This rate is close to human performance. On the other hand, the performance of AI decreases when a subjective component enters the decision-making process. The use of larger databases and the use of more complex R-CNN of the Mask-R-CNN type are ways to improve and potentially increase these results.

Disclosure of interests

Related to this study: none of the authors.
Outside of this study:
Rayane Benhenneda: none.
Thierry Brouard: none.
Christophe Charousset: consultant and royalties for/by Move-up.
Julien Berhouet: consultant for Stryker.

Funding

No funding was received for this study.

Contribution

Rayane Benhenneda: data collection and analysis, manuscript writing.
Thierry Brouard: data analysis, statistical analysis, proofreading.
Julien Berhouet: study design, data analysis, writing, proofreading.
Christophe Charousset: data analysis.

Acknowledgement

The authors thank all the participating centers for their collaboration in this study (Dr Antoni, Dr Barth, Dr Bonnevialle, Dr Dordain, Dr Gadéa, Dr Galinet, Dr Gasse, Dr Guéry, Dr Joudet, Dr Neyton, Dr Ohl), and all patients who gave consent to participate in this study.

References

[1] Hui AT, Alvandi LM, Eleswarapu AS, Fornari ED. Artificial intelligence in modern orthopaedics: current and future applications. JBJS Rev 2022:10, http://dx.doi.org/10.2106/JBJS.RVW.22.00086.
[2] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44.
[3] 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.
[4] 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 2022;30:2535–7.
[5] Kumar V, Roche C, Overman S, Simovitch R, Flurin P-H, Wright T, et al. Using machine learning to predict clinical outcomes after shoulder arthroplasty with a minimal feature set. J Shoulder Elbow Surg 2021;30:e225–336.
[6] Kitaguchi D, Takeshita N, Hasegawa H, Ito M. Artificial intelligence-based computer vision in surgery: recent advances and future perspectives. Ann Gastroenterol Surg 2022;6:29–36.
[7] Hodgins JL, Veillette C, Biau D, Sonnadara R. The knee arthroscopy learning curve: quantitative assessment of surgical skills. Arthroscopy 2014;30:613–21.
[8] 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 2022;30:349–60.
[9] Randelli P, Cucchi D, Butt U. History of shoulder instability surgery. Knee Surg Sports Traumatol Arthrosc 2016;24:305–29.
[10] 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 2017;25:3229–36.
[11] 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 2005;14:238–46.
[12] 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 2021;30:951–60.
[13] 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 2020;48:1439–49.
[14] Godenèche A, Kempf J-F, 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 2018;27:1939–45.
[15] 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 2018;26:169–75.
[16] Doyle DJ, Goyal A, Bansal P, Garmon EH, American Society of Anesthesiologists Classification. StatPearls. Treasure Island (FL): StatPearls Publishing; 2021.
[17] Groll DL, To T, Bombardier C, Wright JG. The development of a comorbidity index with physical function as the outcome. J Clin Epidemiol 2005;58:595–602.
[18] Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. Clin Orthop 1987:160–4.
[19] Gilbart MK, Gerber C. Comparison of the subjective shoulder value and the Constant score. J Shoulder Elbow Surg 2007;16:717–21.
[20] Baumgarten KM, Chang PS. The American shoulder and elbow surgeons score highly correlates with the simple shoulder test. J Shoulder Elbow Surg 2021;30:707–11.
[21] 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 J 2013;95-B:935–41.

[22] McLean A, Taylor F. Classifications in brief: bigliani classification of acromial morphology. Clin Orthop 2019;477:1958–61.
[23] Morid MA, Borjali A, Del Fiol G. A scoping review of transfer learning research on medical image analysis using ImageNet. Comput Biol Med 2021;128:104115.
[24] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision 2015. https://doi.org/10.48550/arXiv.1512.00567.
[25] Guy S, Jacquet C, Tsenkoff D, Argenson J-N, Ollivier M. Deep learning for the radiographic diagnosis of proximal femur fractures: Limitations and programming issues. Orthop Traumatol Surg Res 2021;107:102837.
[26] Olczak J, Fahlberg N, Maki A, Razavian AS, Jilert A, Stark A, et al. Artificial intelligence for analyzing orthopedic trauma radiographs. Acta Orthop 2017;88:581–6.
[27] 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 2019;63:27–32.
[28] Baro E, Degoul S, Beuscart R, Chazard E. Toward a literature-driven definition of big data in healthcare. BioMed Res Int 2015;2015:639021.
[29] Jee K, Kim GH. Potentiality of big data in the medical sector: focus on how to reshape the healthcare system. Healthc Inform Res 2013;19:79–85.
[30] 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 2016;263:50–7.
[31] He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN 2018. https://doi.org/10.48550/arXiv.1703.06870.
[32] Opfer R, Krüger J, Spies L, Ostwaldt A-C, 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 2022, http://dx.doi.org/10.1007/s00330-022-09170-y.