Colour Doppler ultrasound of temporal arteries for the diagnosis of giant cell arteritis: a multicentre deep learning study

C. Roncato[1], L. Perez[2], A. Brochet-Guégan[3], C. Allix-Béguec[4], A. Raimbeau[5], G. Gautier[5], C. Agard[5], G. Ploton[5], S. Moisselin[6], F. Lorcerie[1], G. Denis[7], B. Gombert[3], E. Gervais[8], O. Espitia[5]

[1]Vascular Medicine Department, Groupe Hospitalier de la Rochelle Ré Aunis, La Rochelle; [2]Artificial Intelligence Lab, Météo-France, Toulouse; [3]Rheumatology Department, Groupe Hospitalier de la Rochelle Ré Aunis, La Rochelle; [4]Clinical Research Unit, Groupe Hospitalier de la Rochelle Ré Aunis, La Rochelle; [5]Dept. of Internal Medicine, CHU Nantes; [6]Université Paul Sabatier, Toulouse; [7]Medicine and Haematology Department, Centre Hospitalier de Rochefort; [8]Rheumatology Unit, CHU de Poitiers, France.

Christophe Roncato, MD
Lior Perez, Eng
Antoine Brochet-Guégan, MD
Caroline Allix-Béguec, PhD
Alizée Raimbeau, MD
Giovanni Gautier, MD
Christian Agard, MD, PhD
Gaëtan Ploton, MD
Simon Moisselin, Eng
Fanny Lorcerie, MD
Guillaume Denis, MD
Bruno Gombert, MD
Elisabeth Gervais, MD, PhD
Olivier Espitia, MD, PhD

Please address correspondence to:
Christophe Roncato,
Vascular Medicine Department,
Groupe Hospitalier de la Rochelle Ré Aunis,
rue du Dr Schweizer,
17019 La Rochelle, France.
E-mail:
christophe.roncato@ght-atlantique17.fr

Received on February 13, 2020; accepted in revised form on April 16, 2020.

Clin Exp Rheumatol 2020; 38 (Suppl. 124): S120-S125.

Key words: giant cell arteritis, diagnosis, Doppler ultrasound imaging, deep learning

Competing interests: see page S125.

ABSTRACT

Objective. *Giant cell arteritis (GCA) is the most common systemic vasculitis in adults. In recent years, colour Doppler ultrasound of the temporal arteries (CDU) has proven to be a powerful non-invasive diagnostic tool, but its place in the diagnosis of GCA remains to be defined. A limitation of the CDU is the inter-operator reproducibility. Image analysis from a different perspective is now possible with the development of artificial intelligence algorithms. We propose to assess this technology for the detection of the halo sign on CDU images.*

Methods. *Three public hospitals retrospectively collected data from 137 patients suspected of having GCA between January 2017 and April 2019. CDU images (n=1,311) were labelled with the VIA software. Three sets (training, validation and test) were created and analysed with a semantic segmentation technique using a U-Net convolutional neural network.*

Results. *The area under the curve (AUC) was 0.931 and 0.835 on the validation and test set, respectively. An image positivity threshold was determined by focusing on the specificity. With this threshold, a specificity of 95% and a sensitivity of 60% were obtained for the test set. The analysis of the false interpretation showed that the acquisition modalities and the presence of thrombus caused confusion for the algorithm.*

Conclusion. *We propose an automated image analysis tool for GCA diagnosis. The 2018 EULAR guidelines for image acquisition must be respected before generalising this algorithm. After external validation, this tool could be used as an aid for diagnosis, staff training and student education.*

Introduction

The most common systemic vasculitis in adults over 50 years is giant cell arteritis (GCA). It mainly affects large and medium vessels, including the branches of the external carotid artery and the temporal arteries (1, 2). Its incidence rate is about 1 per 10,000 inhabitants (3). Clinical symptoms include unusual headaches, ophthalmological disorders (diplopia, visual acuity loss), and a decrease in temporal pulse (4, 5). To confirm the clinical suspicion of GCA, the reference examination is the temporal artery biopsy (6), but more and more studies support the use of colour Doppler ultrasound (CDU) as a first-line examination (7-13). It is a non-invasive and inexpensive imaging technique that detects specific signs of parietal inflammation, such as the halo sign describe since 1995 (14), and the more recently described compression sign (15). A good reproducibility in these signs interpretation was showed by a group of experts (16). However, this technique is not yet widespread, as the quality of CDU depends on the transducer frequency and images can be misinterpreted, especially when the operator lacks experience (17). Currently, artificial intelligence is one of the most disruptive technologies in the health system. It includes Deep Learning (DL) technology that provides new image reading tools, such as convolutional neural networks (18–23). The power of these networks is based on their ability to extract relevant data from a multitude of information (big data). Images selected and annotated by experienced operators are required as input data. Through successive calculations, convolutional neural network increases its performance in the interpretation of images.

We hypothesised that the use of a convolutional neural network applied to CDU images in the case of suspected GCA could allow for more robust identification of the halo sign.
The objective of this study was to evaluate a semantic segmentation algorithm (positive or negative areas in images) in the automated interpretation of halo signs from CDU images.

Methods

Study design
This is a multicentre retrospective study assessing the accuracy of halo sign identification by a convolutional neural network.

Setting and participants
In three French public hospitals, patients with a GCA suspicion and who underwent a CDU examination between January 2017 and April 2019 were included.
GCA cases were suspected by expert clinicians and/or by detection of an aortitis or an arteritis of one or more arteries by aortic imaging (CT angiography, magnetic resonance angiography or positron emission tomography).
CDU images were acquired by 6 different operators using Siemens S2000 or Toshiba Aplio 400 ultrasound machines, and 18 Mhz transducer. Each participating centre acquired CDU images with their current practice settings. Two operators systematically used 9 cm/sec pulse repetition frequency colour duplex ultrasound, with maximum use of the zoom, and with adjusted Colour Doppler for noise elimination. Four operators either used adjusted Colour Doppler or power Doppler for noise elimination, without systematic use of the zoom.

Ethics
The study was approved by the local ethics committee on May 7, 2019. In accordance with the data protection act (Loi Informatique et Liberté), an information letter was sent to the concerned patients with all the necessary information about the study, and contact details to enable them to express their views in the event of a request to object to the use of their data.

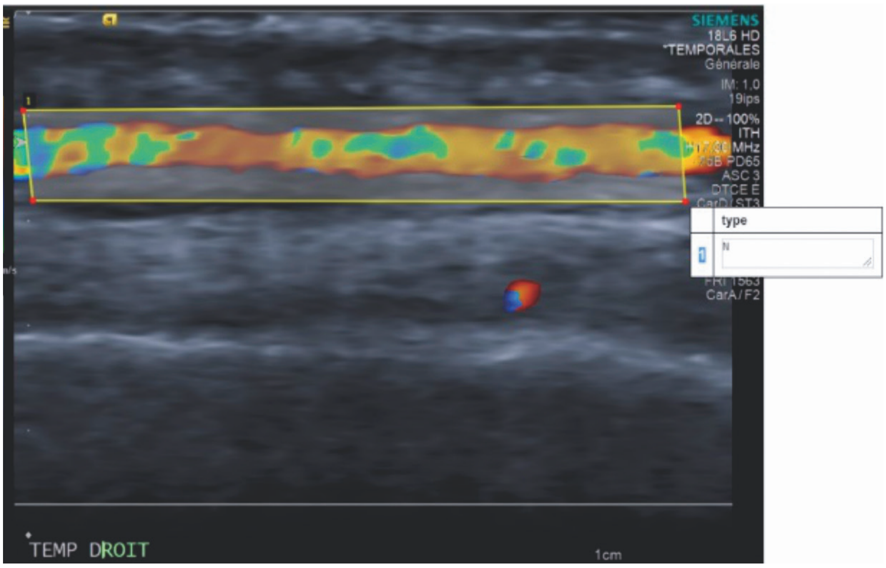

Fig. 1. Example of labelling a CDU image. The vascular area and its immediate vicinity have been selected and labelled as normal artery (N).

Deep learning approach
The semantic segmentation technique, which achieves fine-grained inference by making dense predictions classifying every pixel, was used. Chosen classes were as follows: positive artery (halo sign), negative artery (normal artery), and non-artery. CDU images of temporal arteries were labelled by cutting each segment of arteries, indicating whether it was positive or negative for GCA (Fig. 1). The model used was a convolutional neural network of U-Net architecture (24), with a loss function corresponding to the sum of categorical cross-entropy and Jaccard distance. Data augmentation was performed during model training: rotation, zoom, symmetry, colour shift, and contrast modification. The resulting model predicts, for each pixel of interest, whether it belongs to one of the two classes (positive artery and negative artery). An overall positive image score was then calculated by adding all pixels scores from the positive artery class.

Statistical analysis
Results were analysed globally and then categorised in two groups according to the types of ultrasound machines and image acquisition techniques. In Group 1, the single operator used the same criteria for image acquisition: zooming to obtain a clear, centred image of a longitudinal section of the temporal artery, using colour Doppler ultrasound. In Group 2, these parameters were subjected to variations due to multiple operators, and multiple transducer modes.
ROC curves were calculated on the validation and test sets to determine the performance of the algorithm in identifying a halo sign.

Results

Participants and image datasets
Patients' initial diagnosis included clinical evaluation, CDU, and TAB in doubtful cases. A total of 1,311 CDU images were obtained from 137 patients. Final diagnosis of giant cell arteritis was confirmed in 71 patients by their positive outcome on corticosteroid therapy. Of these, 43 had a halo on the temporal arteries as evidenced by CDU (Table I). Alternative diagnoses included polymyalgia rheumatica, cephalea, primary cerebral vasculitis, lymphocytic vasculitis, postherpetic neuralgia, lymphoma, viral infection, urinary infection, chondrocalcinosis, renal cyst rupture, and spontaneous improvement without diagnosis. Patients' images were randomly separated to form three sets (training, validation and test) (Table II).

Scoring
Based on the training set resulting model, the average scores of positive and negative images were of 2633 pix-

Table I. Baseline characteristics.

	Group 1 N = 83	Group 2 N = 54
Demographics		
Age (years), mean (SD)	73.6 (± 8.9)	72.6 (± 12.1)
Female, n (%)	46 (55.4%)	35 (64.8%)
Diagnosis		
Giant cell arteritis, n (%)	34 (41.0%)	37 (68.5%)
Alternative diagnosis, n (%)	39 (53.4%)	17 (31.5%)
CDU* result		
Thrombus, n(%)	2 (2.4%)	4 (2.2%)
Halo sign, (%)	25 (30.1%)	18 (33.3%)

*Colour duplex ultrasound.

Table II. Content of training, validation and test sets.

sets	Number of patients		Total number of images
	Halo positive	Halo negative	
Training	20	45	627
Validation	13	25	342
Test	10	24	342

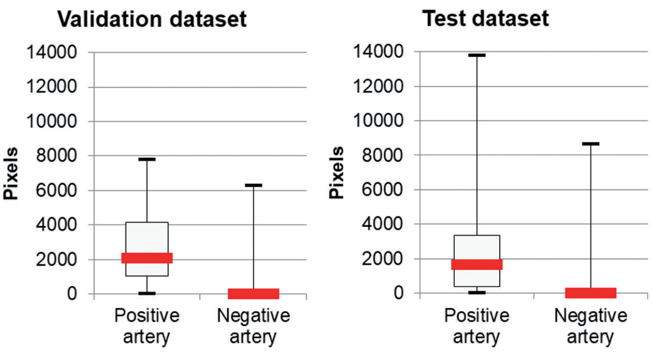

Fig. 2. Box plot of positive and negative image scores for the validation and test sets.

els (± 2137) and 161 (± 618), and 2564 pixels (± 3105) and 190 (± 736) for the validation and test sets, respectively (Fig. 2).

Convolutional neural network classification

To evaluate the model's performance in detecting the halo sign, areas under the curve (AUC) were calculated (Fig. 3). AUC of 0.931 and 0.835 were obtained for the validation and test sets, respectively. From the validation set and by focusing on the specificity, a threshold was determined at 1,200 pixels as the minimum score to define a positive artery. With this threshold, the algorithm had a sensitivity and a specificity of 97.1% and 68.8%; and 95.4% and 59.7% for the validation and test sets, respectively (Table III). For Group 1, the algorithm showed overall better performance. Examples of segmentation are shown in Figure 4, where the experts labelled the longitudinal section of the arteries as positive (example A) and negative (example B). The semantic segmentation algorithm correctly classified the images, and the scores of 4,330 and 3 were over and below the 1,200 pixel threshold, respectively.

Analysis of misclassification

Images from group 2 of small transverse section of arteries were not classified successfully. In example C (Fig. 4), the semantic segmentation algorithm detected a majority of positive pixels but, since the image of the artery was small, a score of 357 below the defined threshold was obtained. In example D, neither the segmentation algorithm nor the score correctly classified the image as positive according to the expert's interpretation. Thrombus (example E), and the use of power ultrasound instead of colour mode (example F) have also been source of misclassification. In the test set, four negative arteries had images that met standardised quality criteria, but the algorithm detected positive pixels with a sum higher than 1,200. Example G illustrated this situation with a score of 3,860 for a negative artery.

Alternative scoring proposal

In order to reduce this bias, the image score could be refined by a score no longer based on an absolute sum but on a ratio $\frac{\text{positive pixel score}}{\text{positive pixel score + negative pixel score}}$. With this method, AUC of 0.943 and 0.880 were obtained for the validation and test sets, respectively. Using a threshold ratio of 0.60, three false positive (6%) and 19 false negative (6%), due to image size bias in group 2 and quality bias, were found.

Discussion

Ultrasound of the temporal arteries is an easily accessible, non-invasive, inexpensive examination. While standardised image acquisition should be performed by a trained and experienced physician (25), the use of deep learning would assist in analysis for an inexperienced operator, and could be used for the continuing education and training of students.

We propose a deep learning model as an aid tool for the detection by colour Doppler ultrasound of halo sign on temporal arteries. This model identified halo signs from pathological arteries in nearly 90% of cases from non-standardised images. On standardised images the accuracy of the algorithm is excellent, and this accuracy remains good on non-standardised images acquired with different ultrasound machines and by different operators. Temporal artery thrombus seems to be associated with GCA (26) and may be a secondary ultrasound sign. Our dataset contained few cases of temporal artery thrombus images. They were included in the test set, but caused confusion for the algorithm that classified them as negative artery (normal artery). In current practice, transversal images are mainly

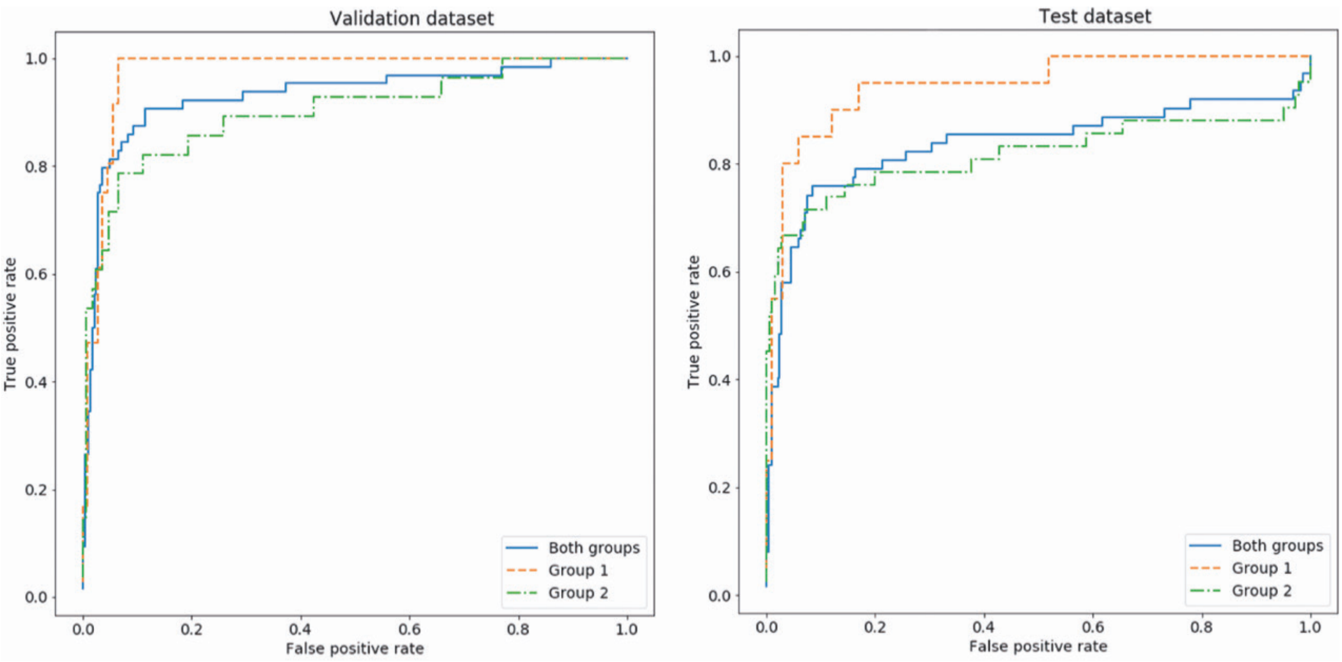

Fig. 3. Receiver operating characteristic analyses for all the images, group 1 and group 2 for the test and validation sets.

Table III. Model performance on the validation and test sets for the whole images, groups 1 and 2.

	Validation			Test		
	Total set	Group 1	Group 2	Total set	Group 1	Group 2
AUC*	0.931	0.974	0.901	0.835	0.946	0.817
Specificity	97.1%	93.5%	99.4%	95.4%	88.0%	99.4%
Sensitivity	68.8%	97.2%	32.1%	59.7%	85.0%	47.6%
Positive likelihood ratio	23.89	15.00	54.64	12.85	7.08	85.71
Negative likelihood ratio	0.32	0.03	0.68	0.42	0.17	0.53

*Area under the curve.

used, as they are easier to acquire than longitudinal images due to the frequent tortuosity of the temporal artery. In this study, the algorithm also correctly classified longitudinal sections.

These encouraging results could be improved by using more data for the training set, which would likely improve the overall performance of our algorithm, and by using normalised CDU images. Townend *et al.* (27), warned about data collection and the importance of harmonisation for deep learning approach. For the diagnosing of rheumatoid arthritis, Andersen *et al.* (19), recommended the use of standardised methods to obtain high-quality ultrasound images of the wrist for the OMERACT-EULAR Synovitis Scoring (OESS) system. The automated analysis of wrist images from different centres did not achieve the same accuracy. Similarly, in our study the accuracy between centres was not the same for the automated analysis of CDU image. By following the EULAR 2018 recommendations on imaging large-vessel vasculitis (25), standardisation would reduce the number of images needed to train an algorithm, and would likely reduce image quality bias.

Another bias was related to the image scoring. The overall positive image score was determined by adding all pixels scores from the positive artery class. The optimisation of the score calculation based on a ratio seemed interesting, but could not be confirmed due to the absence of a second blank test set. A subsequent study on a larger dataset would probably help to refine this point. Other optimisations of the image score could be implemented like the detection of non-normalised images on which the algorithm does not see any arteries, and the detection of situations where the algorithm clearly identifies a part of the artery as pathological. However, this has to be done after the convolutional neural network of semantic segmentation has been properly trained.

A limitation of our study was the choice of algorithm. Since the objective was to make a positive or negative diagnosis on each image, it would have been natural to use a classification algorithm (28). But this approach would have resulted in overfitting, given the small number of images available for training. To train a classifier on such a limited dataset, it would have been necessary to use the transfer learning technique, which consists in specialising a pre-trained model on another set of similar images. However, we did not have such a model already trained on ultrasound images. We have therefore favoured the technique of semantic segmentation.

In conclusion, this study suggests that deep learning technology using convolutional neural network can contribute to the diagnosis of GCA through automated analysis of CDU images. This method could be improved by collecting data on a larger scale, optimising the image score and the positivity threshold. A tool integrated into

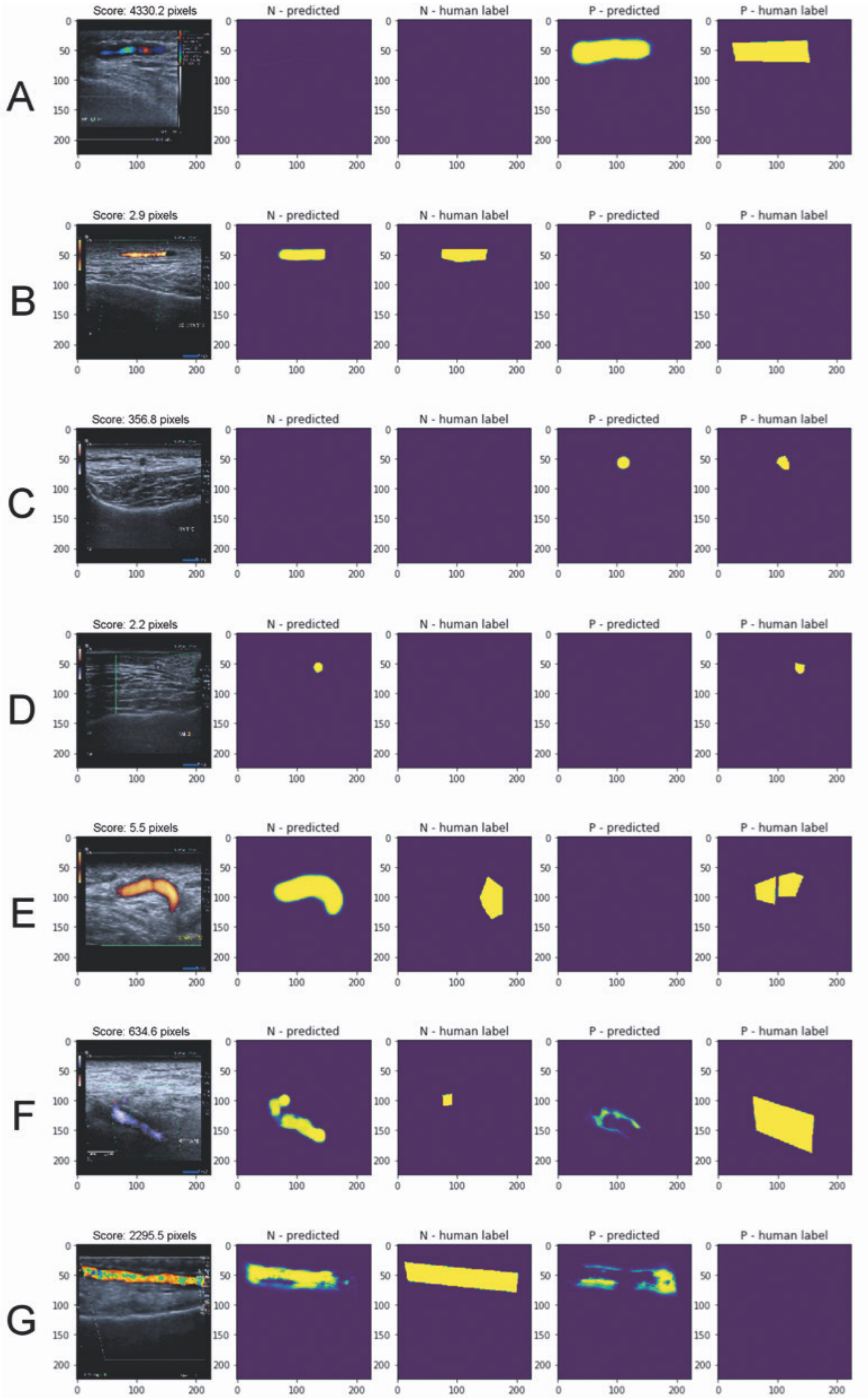

Fig. 4. Examples of classification by the algorithm. **A**. longitudinal section of a positive artery; **B**. longitudinal section of a negative artery; **C**. transverse section of a positive artery correctly classified by the segmentation algorithm but with a score below the threshold ; **D**. transverse section of a positive artery incorrectly classified either by the segmentation algorithm or the score; **E**. longitudinal section of a positive artery with a thrombus; **F**. longitudinal section of a negative artery examined with an power ultrasound; **G**. longitudinal section of a negative artery. Positive image scores and labelling by experts and the model are shown above the images.

ultrasound machines would require the standardisation of large-vessel ultrasound imaging.

Acknowledgements
Part of the ultrasound images were gathered within the ECHORTON trial (NCT02703922) financed as part of the tender "Recherche CH-CHU de la région Poitou-Charentes".

Competing interests
C. Roncato reports grants and personal fees from Bayer HealthCare SAS, Pfizer SAS, LEO Pharma, Aspen France, Bristol-Myers Squibb and Amgen SAS, outside the submitted work; C. Allix-Béguec reports grants from Horizon Atlantique 17 endowment fund, outside the submitted work; B. Gombert reports personal fees from Lilly, BMS, and Pfizer, outside the submitted work; the other authors have declared no competing interests.

References

1. WEYAND CM, GORONZY JJ: Giant-cell arteritis and polymyalgia rheumatica. *New Engl J Med* 2014; 371: 1653.
2. SMITH JH, SWANSON JW: Giant cell arteritis. *Headache* 2014; 54: 1273-89.
3. GONZALEZ-GAY MA, VAZQUEZ-RODRIGUEZ TR, LOPEZ-DIAZ MJ *et al.*: Epidemiology of giant cell arteritis and polymyalgia rheumatica. *Arthritis Care Res* 2009; 61: 1454-61.
4. GONZALEZ-GAY MA, BARROS S, LOPEZ-DIAZ MJ, GARCIA-PORRUA C, SANCHEZ-ANDRADE A, LLORCA J: Giant cell arteritis: Disease patterns of clinical presentation in a series of 240 patients. *Medicine* (Baltimore). 2005; 84: 269-76.
5. GONZÁLEZ-GAY MA, PRIETO-PEÑA D, MARTÍNEZ-RODRÍGUEZ I *et al.*: Early large vessel systemic vasculitis in adults. *Best Pract Res Clin Rheumatol* 2019; 33: 101424.
6. HUNDER GG, BLOCH DA, MICHEL BA *et al.*: The American College of Rheumatology 1990 criteria for the classification of Giant Cell Arteritis. *Arthritis Rheum* 1990; 33: 1122-8.
7. RONCATO C, ALLIX-BÉGUEC C, BROTTIER-MANCINI E, GOMBERT B, DENIS G: Diagnostic performance of colour duplex ultrasonography along with temporal artery biopsy in suspicion of giant cell arteritis. *Clin Exp Rheumatol* 2017; 35 (Suppl. 103): S119-22.
8. RINAGEL M, CHATELUS E, JOUSSE-JOULIN S *et al.*: Diagnostic performance of temporal artery ultrasound for the diagnosis of giant cell arteritis: a systematic review and meta-analysis of the literature. *Autoimmun Rev* 2019; 18: 56-61.
9. MURATORE F, BOIARDI L, RESTUCCIA G *et al.*: Comparison between colour duplex sonography findings and different histological patterns of temporal artery. *Rheumatology* (Oxford) 2013; 52: 2268-74.
10. CONWAY R, O'NEILL L, McCARTHY GM *et al.*: Performance characteristics and predictors of temporal artery ultrasound for the diagnosis of giant cell arteritis in routine clinical practice in a prospective cohort. *Clin Exp Rheumatol* 2019; 37 (Suppl. 117): S72-8.
11. BALL EL, WALSH SR, TANG TY, GOHIL R, CLARKE JMF: Role of ultrasonography in the diagnosis of temporal arteritis. *Br J Surg* 2010; 97: 1765-71.
12. ARIDA A, KYPRIANOU M, KANAKIS M, SFIKAKIS PP: The diagnostic value of ultrasonography-derived edema of the temporal artery wall in giant cell arteritis: A second meta-analysis. *BMC Musculoskelet Disord* 2010; 11: 44.
13. MONTI S, BOND M, FELICETTI M *et al.*: One year in review 2019: vasculitis. *Clin Exp Rheumatol* 2019; 37 (Suppl. 117): S3-19.
14. SCHMIDT W, KRAFT H, VOLKER L, VORPAHL K, GROMNICA-IHLE E: Colour Doppler sonography to diagnose temporal arteritis. *Lancet* 1995; 345: 866.
15. ASCHWANDEN M, DAIKELER T, KESTEN F *et al.*: Temporal artery compression sign - A novel ultrasound finding for the diagnosis of giant cell arteritis. *Ultraschall Med* 2013; 34: 47-50.
16. CHRYSIDIS S, DUFTNER C, DEJACO C *et al.*: Definitions and reliability assessment of elementary ultrasound lesions in giant cell arteritis: A study from the OMERACT large vessel vasculitis ultrasound working group. *RMD Open* 2018; 4: e000598.
17. LUQMANI R, LEE E, SINGH S *et al.*: The Role of Ultrasound Compared to Biopsy of Temporal Arteries in the Diagnosis and Treatment of Giant Cell Arteritis (TABUL): a diagnostic accuracy and cost-effectiveness study. *Health Technol Assess* 2016; 20: 1-238.
18. ABRÀMOFF MD, LAVIN PT, BIRCH M, SHAH N, FOLK JC: Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med* 2018; 1: 39.
19. ANDERSEN JKH, PEDERSEN JS, LAURSEN MS *et al.*: Neural networks for automatic scoring of arthritis disease activity on ultrasound images. *RMD Open* 2019; 5: e000891.
20. BURLINA P, BILLINGS S, JOSHI N, ALBAYDA J: Automated diagnosis of myositis from muscle ultrasound: Exploring the use of machine learning and deep learning methods. *PLoS One* 2017; 12: 1-15.
21. ESTEVA A, KUPREL B, NOVOA RA *et al.*: Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; 542: 115-8.
22. MIELNIK P, FOJCIK M, SEGEN J, KULBACKI M: A Novel Method of Synovitis Stratification in Ultrasound Using Machine Learning Algorithms: Results From Clinical Validation of the MEDUSA Project. *Ultrasound Med Biol* 2018; 44: 489-94.
23. GOSSEC L, GUYARD F, LEROY D *et al.*: Detection of Flares by Decrease in Physical Activity, Collected Using Wearable Activity Trackers in Rheumatoid Arthritis or Axial Spondyloarthritis: An Application of Machine Learning Analyses in Rheumatology. *Arthritis Care Res* (Hoboken) 2019; 71: 1336-43.
24. RONNEBERGER O, FISCHER P, BROX T: U-Net: Convolutional Networks for Biomedical Image Segmentation. *Int Conf Med Image Comput Comput Interv* 2015; 234-5.
25. DEJACO C, RAMIRO S, DUFTNER C *et al.*: EULAR recommendations for the use of imaging in large vessel vasculitis in clinical practice. *Ann Rheum Dis* 2018; 77: 636-43.
26. CAVAZZA A, MURATORE F, BOIARDI L *et al.*: Inflamed temporal artery: histologic findings in 354 biopsies, with clinical correlations. *Am J Surg Pathol* 2014; 38: 1360-70.
27. TOWNEND D: Conclusion: harmonisation in genomic and health data sharing for research: an impossible dream? *Hum Genet* 2018; 137: 657-64.
28. LITJENS G, KOOI T, BEJNORDI BE *et al.*: A survey on deep learning in medical image analysis. *Med Image Anal* 2017; 42: 60-88.