




## Clinical science

# Vision transformer assisting rheumatologists in screening for capillaroscopy changes in systemic sclerosis: an artificial intelligence model

Alexandru Garaiman<sup>1,†</sup>, Farhad Nooralahzadeh<sup>2,†</sup>, Carina Mihai<sup>1</sup>, Nicolas Perez Gonzalez<sup>2</sup>, Nikitas Gkikopoulos<sup>1</sup>, Mike Oliver Becker <sup>1</sup>, Oliver Distler <sup>1</sup>, Michael Krauthammer<sup>3,‡</sup>, Britta Maurer <sup>1,4,\*</sup>

<sup>1</sup>Department of Rheumatology, University Hospital Zurich, University of Zurich, Zurich, Switzerland

<sup>2</sup>Department of Biomedical Informatics, University of Zurich, Zurich, Switzerland

<sup>3</sup>Department of Quantitative Biomedicine, University of Zurich, Zurich, Switzerland

<sup>4</sup>Department of Rheumatology and Immunology, University Hospital Bern, University of Bern, Bern, Switzerland

\*Correspondence to: Britta Maurer, Department of Rheumatology and Immunology, University Hospital Bern, University of Bern, Freiburgstrasse 16p, 3010 Bern, Switzerland. E-mail: britta.maurer@insel.ch

<sup>†</sup>Alexandru Garaiman and Farhad Nooralahzadeh contributed equally to this study.

<sup>‡</sup>Michael Krauthammer and Britta Maurer contributed equally to this study.

## Abstract

**Objectives:** The first objective of this study was to implement and assess the performance and reliability of a vision transformer (ViT)-based deep-learning model, an ‘off-the-shelf’ artificial intelligence solution, for identifying distinct signs of microangiopathy in nailfold capillaroscopy (NFC) images of patients with SSc. The second objective was to compare the ViT’s analysis performance with that of practising rheumatologists.

**Methods:** NFC images of patients prospectively enrolled in our European Scleroderma Trials and Research group (EUSTAR) and Very Early Diagnosis of Systemic Sclerosis (VEDOSS) local registries were used. The primary outcome investigated was the ViT’s classification performance for identifying disease-associated changes (enlarged capillaries, giant capillaries, capillary loss, microhaemorrhages) and the presence of the scleroderma pattern in these images using a cross-fold validation setting. The secondary outcome involved a comparison of the ViT’s performance vs that of rheumatologists on a reliability set, consisting of a subset of 464 NFC images with majority vote-derived ground-truth labels.

**Results:** We analysed 17 126 NFC images derived from 234 EUSTAR and 55 VEDOSS patients. The ViT had good performance in identifying the various microangiopathic changes in capillaries by NFC [area under the curve (AUC) from 81.8% to 84.5%]. In the reliability set, the rheumatologists reached a higher average accuracy, as well as a better trade-off between sensitivity and specificity compared with the ViT. However, the annotators’ performance was variable, and one out of four rheumatologists showed equal or lower classification measures compared with the ViT.

**Conclusions:** The ViT is a modern, well-performing and readily available tool for assessing patterns of microangiopathy on NFC images, and it may assist rheumatologists in generating consistent and high-quality NFC reports; however, the final diagnosis of a scleroderma pattern in any individual case needs the judgement of an experienced observer.

**Keywords:** SSc, scleroderma, nailfold capillaroscopy, artificial intelligence, automated, image analysis

### Rheumatology key messages

- The Vision Transformer (ViT), an ‘off-the-shelf’ artificial intelligence solution, can be used for the classification of nailfold capillaroscopy (NFC) images.
- Without any prior feature engineering, the ViT’s ability to classify NFC images is on a par with that of practising rheumatologists
- The ViT promises to be able to assist rheumatologists in screening for NFC changes and generating reliable reports.

## Introduction

SSc is a rare but potentially lethal autoimmune disease characterized by three hallmarks: small-vessel damage (microangiopathy), production of disease-specific autoantibodies,

and deposition of extracellular matrix proteins, resulting in tissue fibrosis [1, 2].

Nailfold capillaroscopy (NFC) is a well-established diagnostic tool in SSc, allowing early diagnosis by identifying

Received: 12 April 2022. Accepted: 14 September 2022

© The Author(s) 2022. Published by Oxford University Press on behalf of the British Society for Rheumatology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, [br]distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

disease-specific microangiopathic changes, which are frequently referred to as the capillaroscopic scleroderma pattern [3]. Morphological and architectural capillary changes captured by NFC (i.e. enlarged capillaries, giant capillaries, loss of capillaries, microhaemorrhages) also reflect dynamic SSc microangiopathy changes, classified into the ‘early’, ‘active’ and ‘late’ patterns [4]. Moreover, NFC changes correlate with disease activity and severity and are predictive for disease worsening [5, 6]. Therefore, given the well-established role of NFC in the diagnosis and prognosis of SSc, the adequate assessment of NFC images is of great importance. However, this often comes with some challenges: need of assessor training, operator/observer bias, and time-consuming assessment, especially if all eight fingers are examined [7].

To overcome these challenges, there is increasing interest in automating the evaluation of NFC images, with recent work showing considerable progress towards this goal. Semi-automated systems, such as *CapiAna*, have reduced the analysis time to 25–35 min per patient [8], while the system addressed by Murray *et al.*, with an analysis time of 2 min per image, helps the clinician to save approximately 64 min per patient (eight fingers) [9]. With their system, Cheng *et al.* achieved an analysis time of 30 s per image [10]. However, all these solutions required image processing before analysis. The fully automated solution introduced by Cutolo *et al.*, the AUTOCAPi, is able to correctly assess an NFC image in only 20 s [11].

All of these approaches opened up promising application areas, from improving NFC assessment quality in clinical routine to standardizing outcome measures in clinical trials [7, 12]. Computer vision algorithms based on recent advances in machine learning represent a new and promising avenue for automating the analysis of NFC images. The ‘Manchester system’ includes a fully automated system operating on a *random forest* algorithm to extract quantitative biomarkers from capillaroscopy images [13]. The system is able to process images and to measure capillary density, mean capillary width, maximum capillary width, shape and derangement of capillaries in 1–2 min per finger [7, 13–15]. A further contribution was made by Gracia Tello *et al.* [16], who applied deep learning, i.e. *Retinanet*, to NFC image analysis, albeit with a smaller dataset than the current study.

At the moment, one of the most competitive and widely used deep learning architectures in computer vision is the Vision Transformer (ViT) [17, 18]. With this in mind, our first objective was to implement ViT’s deep learning architecture (see [17]) for identifying patterns of microangiopathy in NFC images from patients with SSc. Our second objective was a comparison of the performance of ViT with that of practising rheumatologists.

## Patients and methods

### Patients and source of data

For this study, we used prospectively collected patient data from two sources, the local European Scleroderma Trials and Research group (EUSTAR) cohort, which includes patients with established SSc [1], and the local Very Early Diagnosis of Systemic Sclerosis (VEDOSS) cohort, which contains patients with early/mild SSc [2]. We included routine NFC images of patients  $\geq 18$  years with visits between 2012 and 2021, who fulfilled either the 2013 ACR/EULAR classification criteria

for SSc or the preliminary criteria for VEDOSS [1, 2]. At baseline and at each annual follow-up visit, a total of 16 images (digits II to V from both hands, two images per digit—at 11 and 1 o’clock) were acquired and analysed. Prior to the image acquisition, the patients were prepared according to Ingegnoli *et al.* [19]. For image acquisition, we used an Optilia Digital Capillaroscopy System, with magnification set to  $\times 200$ . We used Synedra for long-term image storage. The examination of NFC images follows the requirements of the international Delphi consensus proposed by Ingegnoli *et al.* [19].

Ethical approval for this data collection and analysis was issued by the Ethics Committee of Zurich, Switzerland (BASEC no. 2016-01515 and BASEC no. 2018-02165). Written informed consent was obtained from the participants (or their legally authorized representatives).

### Outcome

The primary outcome under investigation was the performance of the ViT in terms of area under the (receiver operator characteristic) curve (AUC), sensitivity and specificity for identifying the following distinct microangiopathic features in NFC images: enlarged capillaries (apex diameter of the capillary  $>20 \mu\text{m}$  and  $<50 \mu\text{m}$ ), giant capillaries ( $>50 \mu\text{m}$ ), loss of capillaries ( $<7$  capillaries/mm), and microhaemorrhages [20]. From now on, these capillary alterations will be referred to simply as NFC changes.

In addition to the NFC changes, we also assessed the performance of the ViT with regard to identifying the presence of a scleroderma pattern. To this end, we used the definition by Smith *et al.* (2019) and labelled an NFC image as scleroderma pattern–positive if it showed very low capillary density ( $\leq 3$  capillaries/mm) OR the presence of giant capillaries [21]. Furthermore, we tested the performance of the ViT to identify the SSc pattern subtypes described by Cutolo [i.e. early, active and late, standardized according to Smith *et al.* 2020 [12], see [Supplementary Table S1](#) (available at *Rheumatology* online) for the respective definitions], which assess the severity of the SSc-associated microangiopathy. The secondary outcome under investigation was the comparison of the ViT’s *vs* rheumatologists’ performance in terms of sensitivity, specificity and accuracy on a reliability set, consisting of a subset of NFC images with majority vote–derived ground-truth labels.

### Statistical analysis

The image analysis was conducted in Python using the Pytorch libraries. The statistical computations were done in R 4.0 language.

### Vision transformer implementation and performance evaluation

A detailed architecture of the ViT model architecture is reported in the [Supplementary Data](#) and displayed in [Supplementary Fig. S1](#), available at *Rheumatology* online. To assess the ViT’s performance, we implemented a *k*-fold ( $k = 5$ ) cross-validation technique: the original dataset was randomly partitioned into five equal-sized subsamples [22]. To ensure a balanced distribution of class labels, we used a stratified random sampling technique. Of the five subsamples, three sets were used for training, one set for validation and the remaining one for testing. This process was repeated five times (i.e. five folds) until each subsample had been given the chance to be in the train, test and validation set, respectively. The AUC was reported for each fold, as a measure of the model performance. The AUC performance

was classified as follows: 90–100% excellent, 80–89.9% good, 70–79.9% fair, 60–69.9% poor, 50–59.9% fail [23]. To express the ViT's performance in terms of a single test sensitivity and specificity, we determined, for each NFC change and fold, the optimal threshold along the receiver operator characteristic (ROC) curve. For this performance evaluation, we used the assessment contained in the clinical report attached to each image as the ground truth. This assessment was provided by the attending physician, who performed the NFC image analysis as part of the routine clinical visit at which the images were generated.

### Rheumatologists vs vision transformer

From the first validation subsample, we randomly selected NFC images ( $n = 464$ , reliability set), which were annotated in a blinded fashion by three invited rheumatologists with different levels of expertise in capillaroscopy: two residents (A.G. and N.G., with 2 months and 2 years of experience assessing NFCs, respectively) and one attending physician (C.M., with 12 years of experience assessing NFCs). The resident physicians completed a 4-week internal course on capillaroscopy offered by C.M. before they started annotating the NFC images. We also had access to the assessment of the same 464 images by the attending physicians who performed NFC image analysis as part of the routine clinical visits at which the images were generated (the human annotators together with the attending physician will be referred to as 'rheumatologists'). To generate ground-truth labels for the images in the reliability set, we derived a majority vote-based label for each of the 464 images (at least three out of the four rheumatologists had to agree on a particular finding).

The rheumatologists and the ViT were then tasked with the identification of NFC image changes in the reliability set. We assessed the performance with regard to the ground-truth labels of the reliability set using AUC (ViT), as well as sensitivity, specificity and accuracy (for both the ViT and the rheumatologists). The ViT's sensitivity, specificity and accuracy were derived by finding an optimal probability threshold along the ROC curve.

We assessed the agreement among the rheumatologists as follows. First, for every image in the reliability set, we measured the disagreement by counting how often two out of three rheumatologists disagreed on the label in a particular NFC image. Second, for the same set of images, we also assessed the agreement between the attending physician and each of the remaining rheumatologists. Agreement cut-offs were considered as follows: 82–100% almost perfect, 64–81% strong, 35–63% moderate, 15–35% weak, 4–15% minimal, 0–4% none [24].

### Sample size considerations

A minimum sample size required a priori was not computed. We assumed an adequate sample size if we obtained similar model performance across all cross-validation folds.

### Missing data

While we could use all capillaroscopy images for model training and validation, not all images had well-documented corresponding clinical and laboratory variables in the local VEDOSS/EUSTAR registries. Thus, the patients' characteristics in Table 1 [25, 26] reflect the baseline visits in our dataset for which we have >50% clinical data and laboratory parameters, and for which a capillaroscopy report was available.

**Table 1.** Clinical and laboratory characteristics of the patients at baseline

|   | EUSTAR         | VEDOSS          |
|---|----------------|-----------------|
| Patients, $n$   | 234            | 55              |
| Age, median (IQR)                                     | 57 (48, 66)    | 44 (33, 57)     |
| Gender: female, $n/N$ (%)                             | 180/234 (76.9) | 51/55 (92.7)    |
| SSc subset: lcSSc, $n/N$ (%)                          | 144/199 (74.2) | NA              |
| Disease duration: time since RP, median (IQR)         | 4.9 (2, 10.7)  | 4.53 (2.2, 8.4) |
| Cigarette smoking ever: yes, $n/N$ (%)                | 63/196 (32)    | NA              |
| RP present: yes, $n/N$ (%)                            | 199/211 (94.3) | 53/54 (98.1)    |
| Digital ulcers: current, $n/N$ (%)                    | 18/202 (8.9)   | 2/55 (3.6)      |
| Digital ulcers: previously, $n/N$ (%)                 | 49/201 (24.3)  | 0/55 0          |
| Gangrene: current, $n/N$ (%)                          | 1/234 (0.4)    | 0/55 0          |
| Pitting scars on fingertips: current, $n/N$ (%)       | 68/226 (30)    | 0/55 0          |
| mRSS, median (IQR)                                    | 2 (0, 8)       | 0 (0, 0)        |
| PH: yes, $n/N$ (%)                                    | 22/204 (10.8)  | 1/55 (1.8)      |
| Arterial hypertension: yes, $n/N$ (%)                 | 49/234 (20.9)  | 7/53 (13.2)     |
| Renal crisis  | 1/77 (1.3)     | 0/55 0          |
| CRP elevation: yes, $n/N$ (%)                         | 38/208 (18.3)  | 4/55 (7.3)      |
| Any vasoactive medication: yes                        | 178/234 (76)   | 55/55 (100)     |
| Scleroderma pattern: present, $n/N$ (%)               | 8/37 (21.6)    | NA              |
| Giant capillaries <sup>a</sup> : extensive, $n/N$ (%) | 9/43 (20.9)    | 1/55 (1.8)      |
| Haemorrhages <sup>a</sup> : extensive, $n/N$ (%)      | 3/33 (9.1)     | 1/55 (1.8)      |
| Capillary loss <sup>a</sup> : extensive, $n/N$ (%)    | 29/69 (42)     | 1/55 (1.8)      |

<sup>a</sup> Scoring system after Smith *et al.* [25]: 'None' is defined as no capillary changes, 'Rare' as  $\leq 33\%$ , 'Moderate' as  $> 33$  but  $< 66\%$ , and 'Extensive' as  $\geq 66\%$  capillary changes per linear millimetre, respectively. mRSS: modified Rodnan skin score;  $n$ : number; NA: not available; EUSTAR: European Scleroderma Trial and Research; IQR: Interquartile range; VEDOSS: Very Early Diagnosis of Systemic Sclerosis; PH: pulmonary hypertension detected by echocardiography [26].

## Results

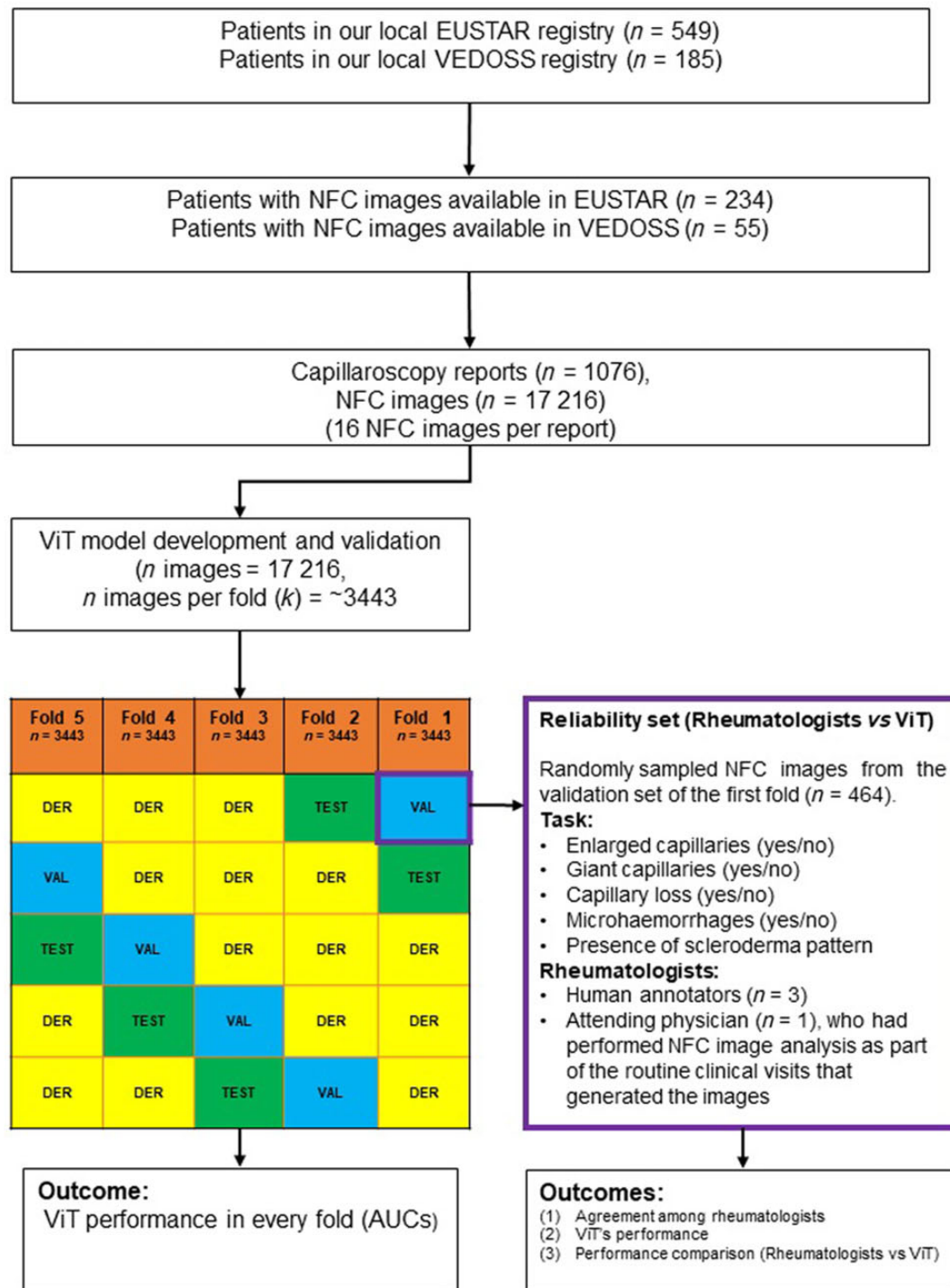
### Patients

We included 234 patients from our local EUSTAR (76.9% female sex, 74.2% lcSSc, median age 57 years, median disease duration 4.9 years) and 55 from our local VEDOSS (92.7% female sex, median age 44 years, median time elapsed since first RP 4.5 years) cohorts that had corresponding NFC images and reports (from a total of 549 EUSTAR and 185 VEDOSS patients). Clinical characteristics of the patients are summarized in Table 1.

### Vision transformer implementation and performance evaluation

From the 17 126 NFC images across the patients in the two registries, five subsets of 3443 NFC images were created to assess the performance of the ViT in a cross-validation setting (see Fig. 1).

The ViT had good mean performance in identifying the NFC changes in all five folds (Fig. 2). The AUC ranged from 81.8% to 84.5% for identifying the different microangiopathic changes. We also tested the ViT's performance to detect scleroderma patterns. Particularly, the accuracy of the ViT to delineate the presence of early, active or late scleroderma pattern ranged from 85.8% to 93.5% (Supplementary Fig. S2, available at *Rheumatology* online), corresponding to a good to excellent accuracy among all cross-validation sets. In terms of



**Figure 1.** Flow chart depicting patient selection across the study. AUC: Area under the curve; EUSTAR: European Scleroderma Trial and Research;  $k$  = folds;  $n$ , number; DER: derivation; NFC: nail-fold capillaroscopy; VAL: validation; VEDOSS: Very Early Diagnosis of Systemic Sclerosis; ViT: Vision Transformer

inference time, the ViT was able to label a NFC image in 0.19 s, which means that for 16 NFC (two images per digit and a total of eight digits) a report can be generated in ~3 s.

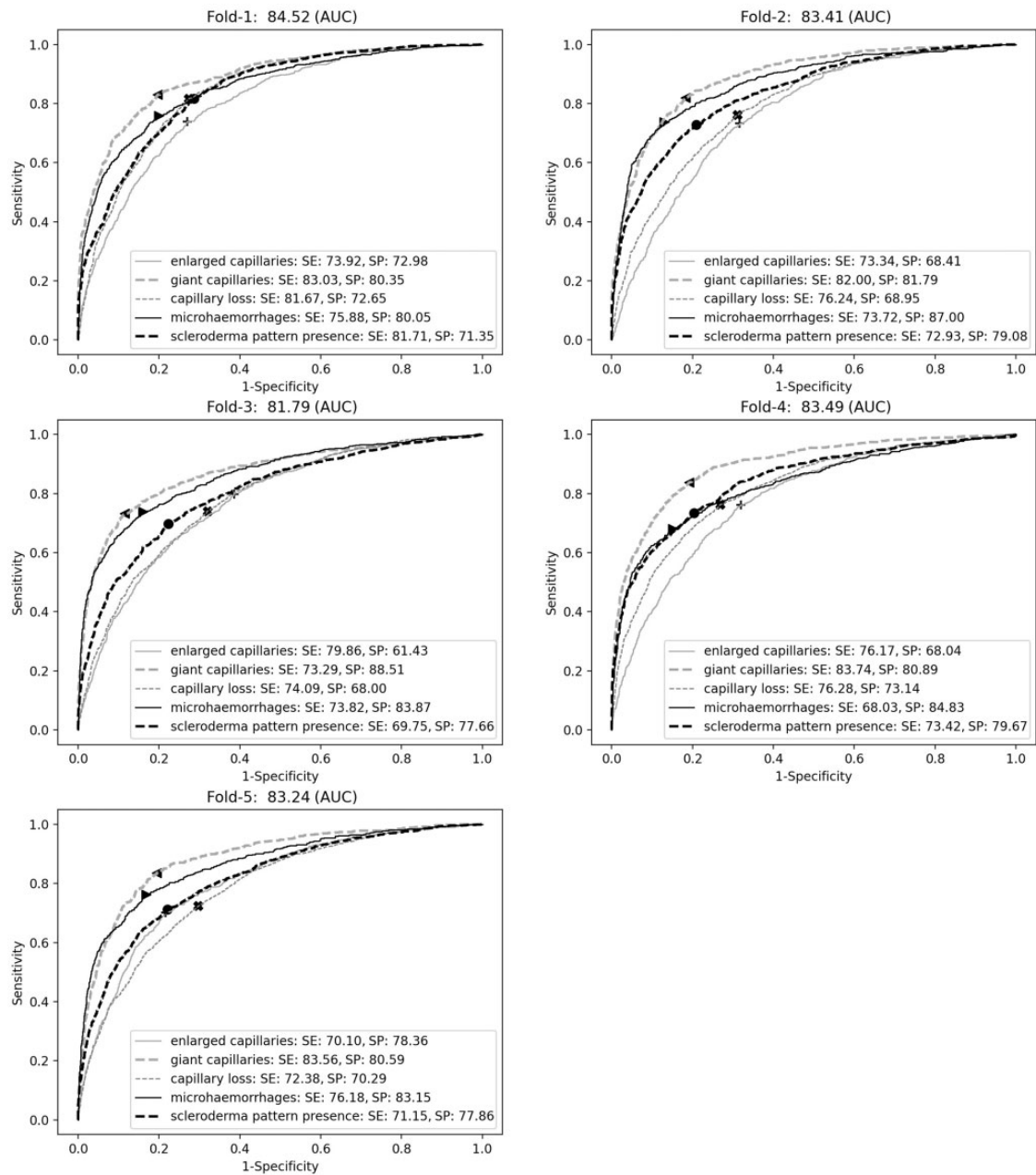
### Rheumatologists vs vision transformer

The reliability subset contained 464 NFC images sampled at random from the validation subset of the first fold. We first assessed the agreement among four rheumatologists

with regard to the findings in these images. We then derived a majority vote-derived ground truth for the images in the reliability set, based on the findings from each rheumatologist.

### Agreement among rheumatologists

When comparing the labels of the four rheumatologists, we observed (Fig. 3A) almost perfect agreement for giant



**Figure 2.** ViT performance in each cross-validation fold (evaluation set,  $n = 17216$  images). The panels show the average AUC across the NFC changes (panel title), as well as the ‘optimal’ sensitivity and specificity per NFC change (box). AUC: area under the (receiver operating characteristic) curve; NFC: nail-fold capillaroscopy; SE: sensitivity; SP: specificity

capillaries (91%) and microhaemorrhages (85%) as well as strong agreement for enlarged capillaries and capillary loss (72% for both changes). Regarding the classification of NFC into presence and absence of scleroderma pattern, the human annotators reached a strong agreement of 81%.

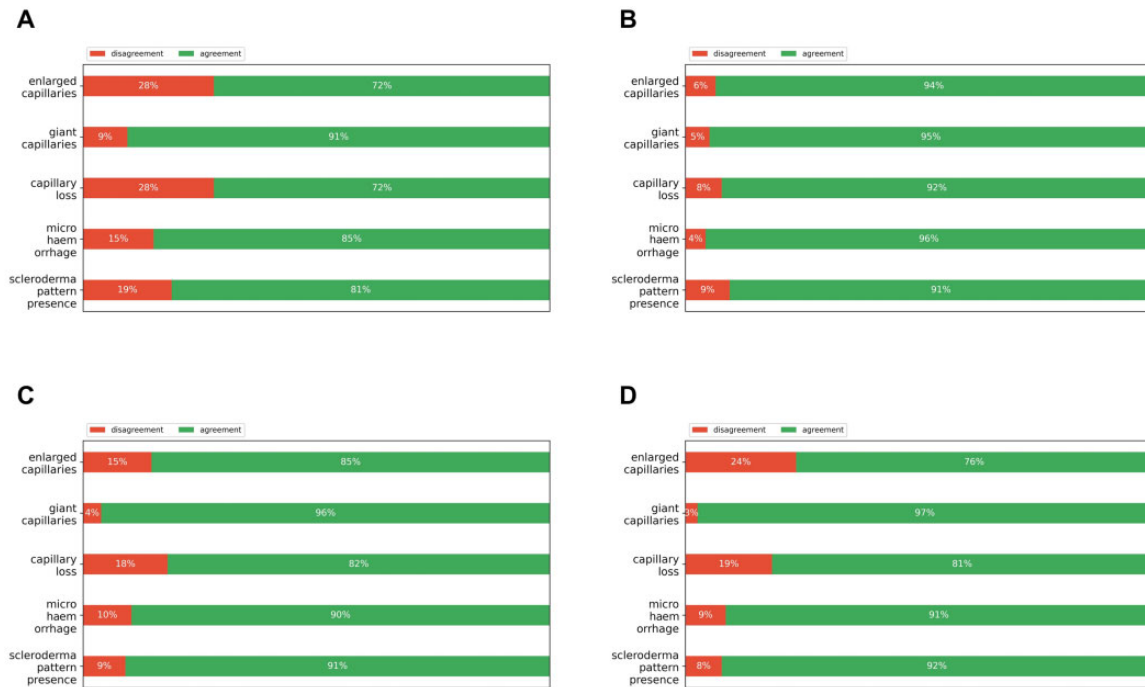
**Agreement with the attending physician**

The agreement reached between the attending physicians and the remaining three rheumatologists was strong to almost perfect (Fig. 3B–D). The highest agreement was reached in assessing of the giant capillaries by annotators C and B (agreement = 97% and 96%, respectively) and for microhaemorrhages by annotator A (agreement = 96%). Lowest agreement was seen in the assessment of capillary loss (annotators

A and B, agreement = 92% and 82%, respectively) and enlargement (annotator C, agreement = 76%). By wrapping these features into presence/absence of scleroderma pattern, we observed an agreement of 91% for annotators A and B and 92% for annotator C, respectively.

**Vision transformer’s performance on the reliability set**

We assessed the ViT’s performance with regard to the ground-truth labels in the reliability set (see Fig. 4). We observed highest performance for diagnosing giant capillaries (AUC = 92.6%), followed by identification of enlarged capillaries (AUC = 90.2%). Good AUCs were seen in depicting capillary loss (AUC = 86.7%), microhaemorrhages (AUC = 85.1%), and the scleroderma pattern (AUC = 88.6%).



**Figure 3.** Agreement among human annotators in assessing the NFC changes in the reliability set ( $n = 464$  images). Panel A: agreement among four annotators (three rheumatologists and an attending physician who assessed the NFC image in the first place at the respective follow-up visit). Panels B, C and D: agreement between the attending physician and each of the three rheumatologists (invited annotator A, B and C, respectively)

### Vision transformer's performance compared with the rheumatologists' performance

The rheumatologists' performance was measured by comparing their individual assessments with the majority vote-derived labels of the ground-truth dataset. The point performance of individual rheumatologists (their specific performance in terms of sensitivity and specificity) is shown in Fig. 4. Point performances that lie within the ViT's ROC curves indicate higher performance of the ViT compared with the human assessors. Accordingly, while the rheumatologists generally demonstrated better performance, for the prediction of capillary loss (assessor B), the ViT classified the NFC images at least as well as an experienced rheumatologist (Fig. 4).

We also measured the rheumatologists' and the ViT's accuracy in correctly identifying the ground-truth labels (Supplementary Fig. S3A–F, available at *Rheumatology* online). The mean accuracy for the human assessors ranged from 90% to 98% (Supplementary Fig. S3A, available at *Rheumatology* online). By deriving an optimal threshold along the ROC curve, a similar accuracy measure was derived for the ViT (Supplementary Fig. S3B, available at *Rheumatology* online). Overall, the rheumatologists' accuracy was higher than the ViT's, ranging between 80% and 100%. The accuracy of the ViT did not drop under 81%, and it showed equal or higher accuracy than one rheumatologist in identifying enlarged capillaries (Supplementary Fig. S3E, available at *Rheumatology* online).

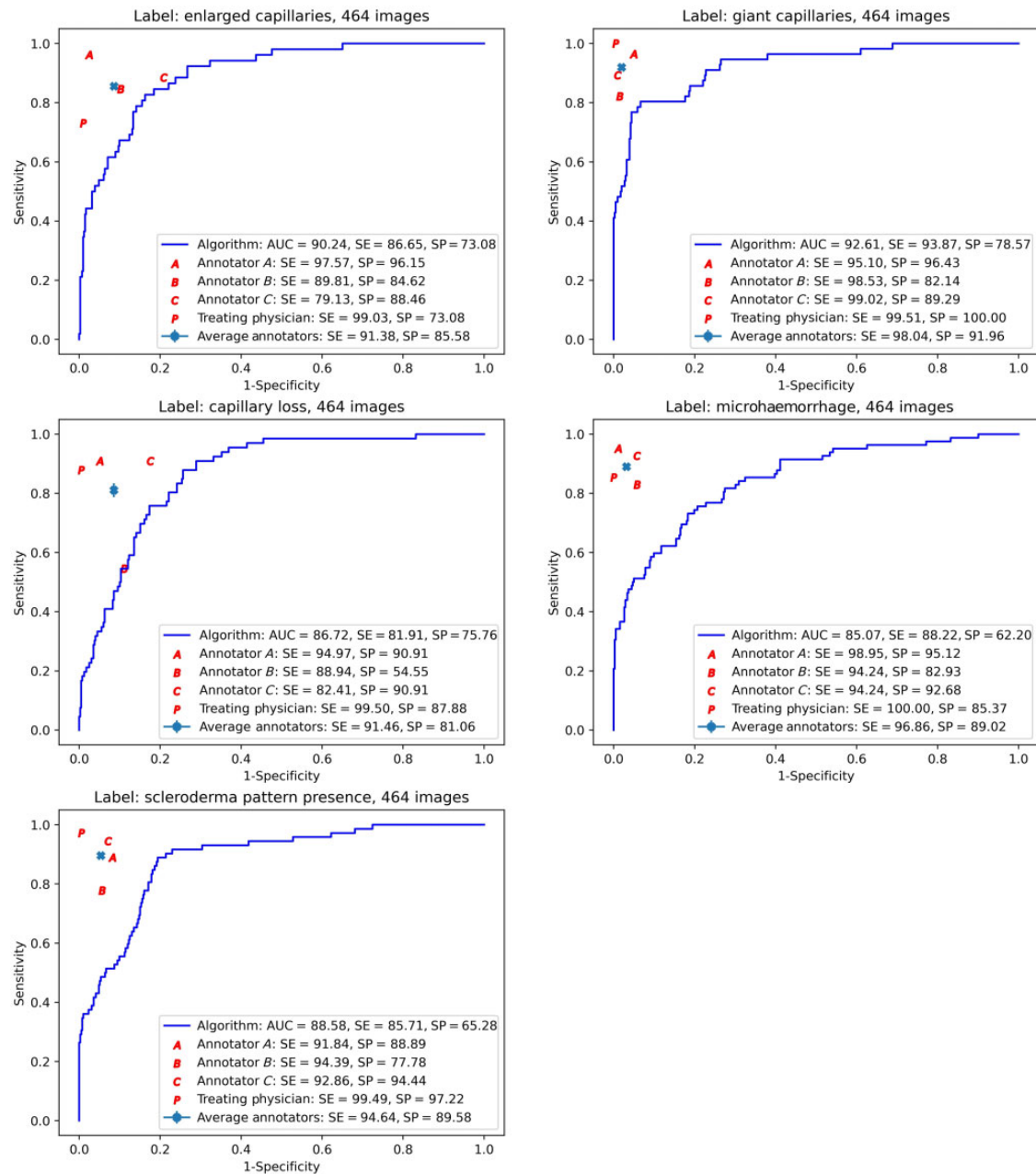
### Discussion

In this study, we addressed the implementation of a powerful 'off-the shelf' deep learning algorithm based on the ViT architecture to identify characteristic features of microangiopathy in NFC images. The ViT system achieved a fair to excellent

performance in our evaluation set-up, with best performance and agreement observed in detecting giant capillaries and microhaemorrhages, followed by capillary loss and enlarged capillaries. The key finding of our study, based on a reliability set with ground-truth labels, is the interindividual variance of the human annotators and the comparable and robust ViT classification performance. Particularly, while the rheumatologists on average had good to excellent classification performance, the ViT showed similar performance, matching one human assessor for specific NFC changes. Given this robust performance, our data warrant further exploration of the use of automated NFC assessment in clinical practice. Particularly, a deep learning-based NFC tool could serve as an assistive device supporting rheumatologists, independent of expertise, in assessing NFC images.

In addition to its good performance, the ViT provides all the advantages of a computational solution: it performs consistently faster than a human assessor does (0.19 s per NFC image). In fact, according to our knowledge, the ViT reached the fastest inference time compared with any other previous automated solutions for NFC assessment [7–11, 13–15]. Furthermore, the software can be hosted on a digital platform and could thus be readily and easily available for use by any clinician. Furthermore, the application of the ViT to a wider spectrum of rheumatic and non-rheumatic diseases affecting the small vessels is straightforward. Naturally, validation of the ViT for these diseases should be addressed in separate studies.

The performance of the ViT was in line with the performance of previously published solutions for automated NFC image classification. Similar AUCs were reported in a study using a random forest classifier for identifying capillary parameters such as density, width or shape [13, 14]. A recent study using a convolutional neural network/RetinaNet



**Figure 4.** Sensitivity/specificity trade-off across the images in the reliability set. The continuous lines show the ROC curves corresponding to the ViT predictions, with the ViT’s AUC, ‘optimal’ sensitivity and specificity shown in the box. Also shown are the point performances of each of the four human annotators, with performances below the continuous line indicating lower performance than the ViT. A: annotator A; B: annotator B; C: annotator C; AUC: area under the ROC curve; P: attending physician; ROC: receiver operating characteristic; SE: sensitivity; SP: specificity; ViT: Vision Transformer

solution revealed similar sensibility and specificity in predicting enlarged capillaries, giant capillaries and microhaemorrhages [16]. Our below-perfect agreement numbers between human annotators are also reflected in previous studies. In Smith *et al.*, the agreement for experts, attendees and novices ranged between 63% and 67% [12, 25]. In our study, different levels of expertise in assessing the NFC images and freedom to select the region of interest on the NFC images might have played a role in the observed agreement differences.

Our study has several limitations. First, it was conducted at a single institution and we have yet to demonstrate the generalizability of our tool in different environments. Nevertheless, despite being a monocentric study, two different cohorts of

SSc patients at different stages of the disease were included. Second, the images were collected with a single type of capillaroscope. Thus, we do not know whether our ViT model can deal with images collected with a different type of optical system. Third, the consistency of our NFC image labels might be sub-optimal. For example, different attending physicians in the clinical routine of our centre might have assessed NFC images of the same patient acquired at different time points. Another limitation may arise from not having tested the ViT on healthy control subjects or on patients with primary RP. The cohort could not be enriched with healthy controls or with patients with primary RP, as the current ethical approval did not allow the inclusion of NFC images from these

subjects. Finally, even though we included baseline and all available follow-up images from all patients, the results do not reflect on the ViT's abilities to track patients longitudinally. Therefore, future studies are needed to evaluate more rigorously whether the ViT can also be used for the assessment of severity over time.

Despite these limitations, our study has several strengths. To our knowledge, this study reports on a deep learning architecture trained on the largest NFC image set to date. The image set included NFC images from SSC patients with established disease from the local EUSTAR registry as well as patients with mild/early disease from the local VEDOSS registry. The innovation brought by the ViT in the capillaroscopy field is that the model represents the state-of-art in computer vision, outperforming convolutional neural network or other previously developed AI solutions [17]. Particularly, no feature engineering was performed, which was undertaken in earlier studies by Berks *et al.* [13]. Moreover, we included all NFC images available, irrespective of any image artefacts. Given the good performance on these unfiltered images, we conclude that the ViT is a capable tool for successfully dealing with image artefacts.

In conclusion, the ViT is an easy-to-use and reliable instrument for assessing NFC images and could be of assistance in detecting NFC changes. Further prospective studies are encouraged to assess the performance of rheumatologists in detecting NFC changes when assisted by automated NFC image classification. However, the final diagnosis of a scleroderma pattern in any individual case needs the judgement of an experienced observer.

## Supplementary data

Supplementary data are available at *Rheumatology* online.

## Data availability statement

The underlying ethics proposal prevents us sharing the data beyond the study team. The software code can be accessed at [https://github.com/uzh-dqbm-cmi/NFC\\_VIT](https://github.com/uzh-dqbm-cmi/NFC_VIT).

## Funding

No specific funding was received from any bodies in the public, commercial or not-for-profit sectors to carry out the work described in this article.

**Disclosure statement:** C.M.—reports personal fees from Boehringer Ingelheim, Eli-Lilly, Mepha, and MEDTalks Switzerland, and congress support from Actelion and Roche, outside the submitted work. M.O.B.—has received personal fees from Amgen and Bayer, outside the submitted work, received speaker fees from Mepha, Novartis, and Vifor, and congress support from GSK, and has an unrestricted research grant from the Novartis Foundation for Medical-Biological Research. O.D.—has received grants or contracts from Kymera and Mitsubishi Tanabe, consulting fees from Abbvie, Acceleron, Alcedim, Amgen, AnaMar, Arxx, AstraZeneca, Baecon, Blade, Bayer, Boehringer Ingelheim, ChemomAb, Corbus, CSL Behring, Galapagos, Glenmark, GSK, Horizon, Inventiva, iQvia, Kymera, Lupin, Medac, Medscape, Miltenyi Biotec, Mitsubishi Tanabe, MSD, Prometheus Biosences, Roche, Roivant, Topadur and UBC, payment or honoraria

for lectures, presentations, speakers bureaus, manuscript writing or educational events from Bayer, Boehringer Ingelheim, Medscape, Novartis, Roche, Pfizer, and Roche and Sanofi, has been issued a patent ('mir-29 for the treatment of systemic sclerosis'), has had leadership or a fiduciary role on other boards or committees or in societies or advocacy groups, paid or unpaid with the ERS/EULAR Guideline on CTD-ILDs, EUSTAR, Pfizer, Swiss Clinical Quality Management in Rheumatic Diseases (SCQM), SAMW, and the Hartmann Müller Foundation. M.K.—has received personal fees from Oncobit, outside of the submitted work. B.M.—has consultancies with Novartis, Boehringer Ingelheim, and Janssen-Cilag, had grant/research support from AbbVie, Protagen, and Novartis Biomedical Research, received speaker fees from Boehringer-Ingelheim as well as congress support from Medtalk, Pfizer, Roche, Actelion, Mepha, and MSD. In addition, B.M. has been issued a patent 'mir-29 for the treatment of systemic sclerosis' (US8247389, EP2331143). The other authors have declared no conflicts of interest.

## Acknowledgements

All authors have made substantial contributions to the paper. A.G. and F.N. contributed equally to this paper. All authors were involved in the review of the paper and approved the final version of the manuscript. A.G. contributed to acquisition of data, NFC images assessment, design of the study, analysis and interpretation of data, drafting and revision of the article. N.F. contributed to design of the study, analysis and interpretation of data, and drafting and revision of the article. C.M. contributed to acquisition of data, NFC images assessment, design of the study, interpretation of data, and revision of the article. N.P.G. contributed to interpretation of data, and revision of the article. N.G. contributed to acquisition of data, NFC images assessment, and revision of the article. M.O.B. contributed to acquisition of data, design of the study, analysis and interpretation of data, and revision of the article. O.D. contributed to acquisition of data, design of the study, analysis and interpretation of data, and revision of the article. M.K. contributed to idealization, design of the study, analysis and interpretation of data, and drafting and revision of the article. B.M. contributed to idealization, acquisition of data, design of the study, analysis and interpretation of data, and drafting and revision of the article. M.K. and B.M. contributed equally to this paper.

## References

1. van den Hoogen F, Khanna D, Fransen J *et al.* 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/ European League Against Rheumatism collaborative initiative. *Ann Rheum Dis* 2013;72:1747–55.
2. Avouac J, Fransen J, Walker UA *et al.*; EUSTAR Group. Preliminary criteria for the very early diagnosis of systemic sclerosis: results of a Delphi Consensus Study from EULAR Scleroderma Trials and Research Group. *Ann Rheum Dis* 2011;70:476–81.
3. Avouac J, Lepri G, Smith V *et al.* Sequential nailfold videocapillaroscopy examinations have responsiveness to detect organ progression in systemic sclerosis. *Semin Arthritis Rheum* 2017;47:86–94.
4. Cutolo M, Sulli A, Pizzorni C, Accardo S. Nailfold videocapillaroscopy assessment of microvascular damage in systemic sclerosis. *J Rheumatol* 2000;27:155–60.



5. Emrani Z, Karbalaie A, Fatemi A, Etehadtavakol M, Erlandsson BE. Capillary density: an important parameter in nailfold capillaroscopy. *Microvasc Res* 2017;109:7–18.
6. Mihai C, Smith V, Dobrota R *et al.* The emerging application of semi-quantitative and quantitative capillaroscopy in systemic sclerosis. *Microvasc Res* 2018;118:113–20.
7. Herrick AL, Berks M, Taylor CJ. Quantitative nailfold capillaroscopy—update and possible next steps. *Rheumatology* 2021;60:2054–65.
8. Gronenschild EHB, Muris DMJ, Schram MT *et al.* Semi-automatic assessment of skin capillary density: proof of principle and validation. *Microvasc Res* 2013;90:192–8.
9. Murray AK, Feng K, Moore TL *et al.* Preliminary clinical evaluation of semi-automated nailfold capillaroscopy in the assessment of patients with Raynaud's phenomenon: semi-automated nailfold capillaroscopy. *Microcirculation* 2011;18:440–7.
10. Cheng C, Lee CW, Daskalakis C. A reproducible computerized method for quantitation of capillary density using nailfold capillaroscopy. *J Vis Exp* 2015;105:e53088.
11. Cutolo M, Trombetta AC, Melsens K *et al.* Automated assessment of absolute nailfold capillary number on videocapillaroscopic images: proof of principle and validation in systemic sclerosis. *Microcirculation* 2018;25:e12447.
12. Smith V, Herrick AL, Ingegnoli F *et al.* Standardisation of nailfold capillaroscopy for the assessment of patients with Raynaud's phenomenon and systemic sclerosis. *Autoimmun Rev* 2020;19:102458.
13. Berks M, Tresadern P, Dinsdale G *et al.* An automated system for detecting and measuring nailfold capillaries. *Med Image Comput Assist Interv* 2014;17:658–65.
14. Berks M, Dinsdale G, Murray A *et al.* Automated structure and flow measurement – a promising tool in nailfold capillaroscopy. *Microvasc Res* 2018;118:173–7.
15. Berks M, Dinsdale G, Murray A *et al.* Improved diagnosis of systemic sclerosis using nailfold capillary flow. In: Ourselin S, Joskowicz L, Sabuncu MR, Unal G, Wells W, eds. *Medical image computing and Computer-Assisted Intervention – MICCAI 2016*. Cham: Springer International Publishing, 2016: 344–52.
16. Gracia Tello B, Ramos Ibañez E, Fanlo Mateo P *et al.* The challenge of comprehensive nailfold videocapillaroscopy practice: a further contribution. *Clin Exp Rheumatol* 2021; Advance Access published 16 December 2021, doi: 10.55563/clinexprheumatol/6usce8.
17. Dosovitskiy A, Beyer L, Kolesnikov A *et al.* An image is worth 16 × 16 words: transformers for image recognition at scale. arXiv:2010.11929 [cs]. 2021. <http://arxiv.org/abs/2010.11929> (29 November 2021, date last accessed).
18. Cuenat S, Couturier R. Convolutional Neural Network (CNN) vs Vision Transformer (ViT) for Digital Holography. 2022. <http://arxiv.org/abs/2108.09147> (18 July 2022, date last accessed)
19. Ingegnoli F, Herrick AL, Schioppo T *et al.*; the European League against Rheumatism (EULAR) Study Group on Microcirculation in Rheumatic Diseases and the Scleroderma Clinical Trials Consortium. Reporting items for capillaroscopy in clinical research on musculoskeletal diseases: a systematic review and international Delphi consensus. *Rheumatology (Oxford)* 2021;60:1410–8.
20. Cutolo M, Herrick AL, Distler O *et al.*; CAP Study Investigators. Nailfold videocapillaroscopic features and other clinical risk factors for digital ulcers in systemic sclerosis: a multicenter, prospective cohort study. *Arthritis Rheumatol* 2016;68:2527–39.
21. Smith V, Vanhaecke A, Herrick AL *et al.*; EULAR Study Group on Microcirculation in Rheumatic Diseases. Fast track algorithm: how to differentiate a 'scleroderma pattern' from a 'non-scleroderma pattern'. *Autoimmun Rev* 2019;18:102394.
22. Andrew YN. Preventing 'overfitting' of cross-validation data. In: *Proceedings of the 14th International Conference on Machine Learning*, 1997.
23. Li F, He H. Assessing the accuracy of diagnostic tests. *Shanghai Arch Psychiatry* 2018;30:207–12.
24. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med* 2012;22:276–82.
25. Smith V, Beeckman S, Herrick AL *et al.*; on behalf of the EULAR Study Group on Microcirculation. An EULAR study group pilot study on reliability of simple capillaroscopic definitions to describe capillary morphology in rheumatic diseases. *Rheumatology (Oxford)* 2016;55:883–90.
26. Hinchcliff M, Fischer A, Schioppo E *et al.* Pulmonary Hypertension Assessment and Recognition of Outcomes in Scleroderma (PHAROS): baseline characteristics and description of study population. *J Rheumatol* 2011;38:2172–9.