Artificial Intelligence in Nailfold Capillaroscopy: A Scoping Review of Validation, Reproducibility, and Clinical Translation

Zahra Emrani[1], Seyed Kamaledin Setarehdan[1], Habibollah Jafari-Varzaneh[2], Alimohammad Fatemi[3], Abdolamir Karbalaie[4,*]

[1]Control and Intelligent Processing centre of Excellence, School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran
[2]Department of Computer Engineering, Na.C., Islamic Azad University, Najafabad, Iran
[3]Department of Rheumatology, Alzahra Hospital, Isfahan University of Medical Sciences, Isfahan, Iran
[4]Department of Clinical Science, Intervention and Technology (CLINTEC), Karolinska Institutet, 17177 Stockholm, Sweden
Corresponding author: Abdolamir Karbalaie, Department of Clinical Science, Intervention and Technology (CLINTEC), Karolinska Institutet. Email: abdolamir.karbalaie@ki.se

Supplementary Material

Supplementary Appendix A. Search strategy, information sources, and Covidence workflow

A.1 Search overview

A multidisciplinary search strategy was used to identify studies applying artificial intelligence (AI), machine learning, deep learning, computer vision, or automated image analysis to nailfold capillaroscopy (NFC), nailfold videocapillaroscopy (NVC), or nailfold microvascular images. The strategy combined three conceptual blocks: (1) nailfold or periungual terminology, (2) capillaroscopy or microcirculation terminology, and (3) AI/computer-vision terminology. The syntax was adapted for each database while preserving the same conceptual structure.

All database searches were completed on 28 March 2026. Records were exported from each database in RIS, PubMed, or EndNote XML format and imported into Covidence for de-duplication, screening, full-text/source assessment, extraction workflow management, and PRISMA-flow tracking. Each database export was imported separately and labelled using the corresponding Covidence source field. No additional database update is represented in the 3,059-record Covidence denominator unless explicitly documented in the repository search-history files.

A.2 Search platforms and Covidence import counts

Supplementary Table S1: Search platforms, export formats, and Covidence import counts.

Source	Platform	Export format	Records imported
PubMed	NLM PubMed	PubMed/RIS	560
Ovid MEDLINE	Ovid	RIS	639
Scopus	Elsevier Scopus	RIS	895
Web of Science Core Collection	Clarivate Web of Science	RIS/EndNote	940
IEEE Xplore	IEEE Xplore	RIS	25
Total			**3,059**

A.3 PubMed search

```
(nailfold OR nail-fold OR periungual OR "nail fold")
AND
(capillaroscopy OR videocapillaroscopy OR "video capillaroscopy" OR NVC
OR "nailfold capillary" OR "nailfold microcirculation")
AND
("artificial intelligence" OR AI OR "machine learning" OR "deep learning"
OR "computer vision" OR algorithm* OR automated OR segmentation
OR detection OR classification OR CNN OR "convolutional neural network"
OR U-Net OR YOLO OR ResNet OR DenseNet OR "Vision Transformer")
```

A.4 Ovid MEDLINE search

```
1. (nailfold* or nail-fold* or periungual or "nail fold*").ti,ab,kf.
2. (capillaroscop* or videocapillaroscop* or "video capillaroscop*" or NVC
   or "nailfold capillar*" or "nailfold microcirculation").ti,ab,kf.
3. ("artificial intelligence" or AI or "machine learning" or "deep learning"
   or "computer vision" or algorithm* or automat* or segment* or detect*
   or classif* or CNN or "convolutional neural network*" or U-Net or YOLO
   or ResNet or DenseNet or "Vision Transformer*").ti,ab,kf.
4. 1 and 2 and 3
```

A.5 Scopus search

```
TITLE-ABS-KEY(
  (nailfold OR "nail-fold" OR periungual OR "nail fold")
  AND
  (capillaroscop* OR videocapillaroscop* OR "video capillaroscop*"
```

```
   OR NVC OR "nailfold capillar*" OR "nailfold microcirculation")
  AND
  ("artificial intelligence" OR "machine learning" OR "deep learning"
   OR "computer vision" OR "image analysis" OR algorithm* OR automat*
   OR segment* OR detect* OR classif* OR CNN
   OR "convolutional neural network*" OR "neural network*" OR "U-Net"
   OR YOLO OR ResNet OR DenseNet OR EfficientNet OR "Vision Transformer*")
)
```

A.6 Web of Science Core Collection search

```
TS=(
  (nailfold OR "nail-fold" OR periungual OR "nail fold")
  AND
  (capillaroscop* OR videocapillaroscop* OR "video capillaroscop*"
   OR NVC OR "nailfold capillar*" OR "nailfold microcirculation")
  AND
  ("artificial intelligence" OR "machine learning" OR "deep learning"
   OR "computer vision" OR "image analysis" OR algorithm* OR automat*
   OR segment* OR detect* OR classif* OR CNN
   OR "convolutional neural network*" OR "neural network*" OR "U-Net"
   OR YOLO OR ResNet OR DenseNet OR EfficientNet OR "Vision Transformer*")
)
```

A.7 IEEE Xplore search

```
("nailfold" OR "nail-fold" OR "periungual" OR "nail fold")
AND
("capillaroscopy" OR "videocapillaroscopy" OR "video capillaroscopy"
 OR "nailfold capillary" OR "nailfold microcirculation")
AND
("artificial intelligence" OR "machine learning" OR "deep learning"
 OR "computer vision" OR "image analysis" OR "algorithm" OR "automated"
 OR "segmentation" OR "detection" OR "classification" OR "CNN"
 OR "convolutional neural network" OR "U-Net" OR "YOLO" OR "ResNet"
 OR "DenseNet" OR "EfficientNet" OR "Vision Transformer")
```

Because IEEE Xplore can be sensitive to long Boolean queries, supplementary shorter searches were

also used when needed:

```
"nailfold capillaroscopy" AND "deep learning"
"nailfold capillary" AND YOLO
"nailfold capillary" AND "computer vision"
"nailfold capillaroscopy" AND "machine learning"
"nailfold microhemorrhage" AND segmentation
```

A.8 Covidence workflow

All exported records were imported into Covidence at the title/abstract screening stage. The source field was assigned according to the database of origin: PubMed, Ovid MEDLINE, Scopus, Web of Science, or IEEE Xplore. Covidence was used to identify duplicate records, and additional duplicate or companion reports were checked manually when records represented the same study family. After de-duplication, titles and abstracts were screened against the predefined eligibility criteria. Records judged eligible or uncertain progressed to full-text/source assessment. Full-text/source exclusion reasons were recorded in Covidence using predefined categories, including abstract-only or insufficient methods/results, non-primary publication, duplicate or companion report, wrong modality or no eligible AI method, and no extractable outcome. The final evidence denominator counted linked study families once, using the most complete eligible peer-reviewed report.

A.9 Supplementary searching and citation checking

Citation searching and targeted manual checks were used to clarify study families, identify companion records, and retrieve contextual reports. These supplementary checks were not treated as an additional database source unless records were formally imported into Covidence and included in the PRISMA denominator. Google Scholar and other broad discovery tools were used only for supplementary checking and background verification, not as a primary source contributing to the 3,059 imported-record denominator.

Supplementary Appendix B. Evidence rules

1. Full peer-reviewed journal articles with original, extractable AI-NFC/NVC results were included in the journal primary track.

2. Full peer-reviewed conference proceedings were included in a separate primary track when dataset, model, validation, and results were sufficiently reported.

3. Abstract-only reports, reviews, commentaries, protocols, book chapters, and preprint-only reports were excluded from the primary denominator and retained only when useful for context or citation

tracing.

4. Multiple reports from the same underlying study were linked and counted once using the most complete eligible peer-reviewed report.

5. Feature-level ML using NFC/NVC-derived variables was retained but distinguished from direct image/video modelling.

6. Reported metrics were interpreted within task families and were not pooled across incompatible tasks.

7. Validation and reproducibility indicators were reported individually and were not combined into a composite quality score, maturity tier, or risk-of-bias rating.

8. AI-specific reporting frameworks such as TRIPOD+AI, CLAIM, DECIDE-AI, CONSORT-AI, and SPIRIT-AI were used in the Discussion as future reporting guidance, not as formal study-level appraisal tools in the present review.

Supplementary Appendix C. PRISMA and evidence-set reconciliation

C.1 Review-flow and evidence-set counts

Supplementary Table S2: Final review-flow and evidence-set counts.

Stage / evidence set	**Count**
Records imported	3,059
Web of Science	940
Scopus	895
Ovid MEDLINE	639
PubMed	560
IEEE Xplore	25
Duplicates removed	1,754
Records screened	1,305
Title/abstract exclusions	1,202
Full-text/source reports assessed	103
Reports excluded from the primary evidence set	60
Primary studies included	43
Peer-reviewed journal articles	37
Full peer-reviewed conference proceedings	6
Contextual reports retained for background	18

C.2 Characteristics of included primary studies

Supplementary Table S3: Characteristics of the 43 included primary studies in the scoping synthesis. Study IDs S01–S43 are sequential identifiers assigned only to the final included primary studies.

Study ID	Year / source	Application domain	Principal task	Model family	Study size	Validation / key limitation
S01	2014; Medical Image Computing and Computer-Assisted Intervention (MICCAI 2014), LNCS 8673	SSc/Raynaud spectrum	Detection/quantification	Layered machine learning using random forests: RF vesselness classifier, RF orientation/width regressors, RF apex-candidate classifier/regressor, HOG features and distal-row probability classification	Participant count not reported at individual level; dataset comprised nailfold mosaics from healthy controls, primary Raynaud's phenomenon and systemic sclerosis groups; 990 manually annotated nailfold mosaics overall; 80 fully annotated images generated 450 training ROIs; validation set 456 images (104 HC, 83 PR, 269 SSc); test set 455 images (104 HC, 83 PR, 268 SSc)	Vessel/orientation/width models trained using 80 fully annotated images; validation images used to tune distal-row and candidate-apex classification; final evaluation performed on an unseen 455-image test set Conference proceeding with limited clinical detail; participant-level counts and independent external validation were not reported; expert disagreement was substantial; method used classical random-forest-based image analysis rather than modern deep learning and provided group-level biomarker differences rather than a validated patient-level diagnostic tool.
S02	2016; Proceedings of the 2016 International Conference on Image Processing, Computer Vision, and Pattern Recognition, IPCV 2016	SSc/Raynaud spectrum	Classification/prediction	Bilateral enhancer preprocessing; multi-dimensional LBP variance texture descriptors; one-against-one multiclass linear SVM; majority voting across fingers for patient classification	12 subjects: three patients per class for control, early, active and late scleroderma-pattern groups; NC images from 3 to 4 fingers per subject; exact image count not reported	Leave-one-out cross-validation on a patient basis repeated 12 times; SVM cost parameter optimized using cross-validation Conference proceeding; very small dataset; no DOI reported; device and magnification not reported; no external validation; limited methodological and reproducibility details; not deep learning; results should be interpreted cautiously.

Study ID	Year / source	Application domain	Principal task	Model family	Study size	Validation / key limitation
S03	2020; Microvascular Research	General/physiological microcirculation	Segmentation	Res-Unet18 and Res-Unet34 compared with original U-Net; ResNet residual blocks integrated into U-Net encoder-decoder architecture	60 people aged 20–50 years: 52 healthy subjects, 3 with hypertension and 5 with hypotension; 30 cases used for training and 20 for testing; 30 training images and 20 test images; ROI cropped to 512×512; training set augmented to 20,664 images	Testing on 20 held-out images with quantitative and qualitative comparison Small dataset; no external validation; not a disease-diagnosis model; air bubbles and reflected light could be misclassified as capillary areas; increasing network depth increases parameters and may affect performance.
S04	2020; Sensors	General/physiological microcirculation	Dynamic/flow analysis	Deep-learning semantic segmentation based on U-Net; video stabilization using image registration/translation vectors; spatiotemporal maps; Radon transform and local maxima detection for WBC event detection	3 healthy subjects for validation videos; additional segmentation dataset image-level only; Segmentation model trained/validated on 1,358 capillary images with ground-truth labels: 950 training and 408 validation images; validation experiment used 3 videos, 2 capillaries per video, 900 frames per video	Segmentation training for 100 epochs with Adam, binary cross-entropy, learning rate 0.001 and batch size 3; validation accuracy reported by mean IoU; WBC counting compared across manual, conventional and DNN segmentation with/without stabilization Very small video validation set with only 3 healthy subjects; no patient/neutropenia validation; results largely qualitative for WBC event counting; no external validation; selected best capillaries for analysis.
S05	2020; IEEE Access	General/physiological microcirculation	Segmentation	DA-CapNet; compared with adaptive Gaussian thresholding, SegNet and original U-Net	7 healthy participants aged 20–35 years; 40 nailfold capillary images; 30 training and 10 testing images; three salient capillaries annotated in each image; augmented to 3000 training samples	30 training and 10 testing images; no cross-validation or external validation Very small dataset; healthy volunteers only; only three salient capillaries annotated per image; no external validation; limited clinical disease validation.

Study ID	Year / source	Application domain	Principal task	Model family	Study size	Validation / key limitation
S06	2022; International Journal of Online and Biomedical Engineering (iJOE)	Diabetes/metabolic	Classification/prediction	YOLOv2 and YOLOv3 object-detection architectures plus an exponential statistical model using rarefaction, avascularity, tortuosity, hemorrhage, elongation and wide-capillary features	Healthy and diabetic subjects; exact participant count not clearly reported; About 600 capillary images from healthy and diabetic subjects were collected; paper also reports 600 training images, 230 test images, and augmentation to 1018 images	Training/test dataset evaluation; YOLO average precision reported on training dataset; manual test-set evaluation and statistical-model evaluation on 230 test images Method combines DL detection with a hand-crafted statistical model; participant-level details are limited; automatic test-set evaluation was not possible according to authors; no external validation; low-cost imaging may reduce image quality.
S07	2022; Applied Sciences	General/physiological microcirculation	Segmentation	DAFM-Net: U-shape CNN backbone with dual attention fusion module and group normalization; compared with U-Net, U-Net+CAM, U-Net+SAM, U-Net+CBAM and U-Net+SSA	Participant count not clearly reported; dataset mainly described at image level; 30 nailfold capillaroscopy images of nailfold microhemorrhages with professional ground-truth masks; manually cropped to 256×256 pixels	Ablation and segmentation experiments on the collected dataset; exact independent train/test split not clearly reported Very small dataset of 30 images; images partly collected from internet/references; acquisition details and train/test split are limited; no external validation; no disease-level clinical classification.
S08	2022; Intelligent Automation & Soft Computing	General/physiological microcirculation	Image quality/preprocessing	Laplacian operator for blur detection; t-distributed stochastic neighbor embedding (t-SNE) for data visualization/cluster distribution; deconvolution sharpening	Not reported; 600 images extracted from two recorded videos; included sharp and blurry NC images; text reports 82 sharper images and 518 blurry images	Comparison of t-SNE cluster visualization with Laplacian/statistical sharpness distribution; no clinical validation No disease-level validation; no external validation; limited to low-cost device image quality; participant details not reported; t-SNE is mainly visualization rather than predictive modeling.

Study ID	Year / source	Application domain	Principal task	Model family	Study size	Validation / key limitation
S09	2022; IEIE Transactions on Smart Processing and Computing	General/physiological microcirculation	Dynamic/flow analysis	K-nearest neighbors; decision tree; random forest; PCA feature reduction	2 subjects / 2 videos; 490 event-data segments: 279 with WBC and 211 without WBC; 392 training and 98 testing segments	8:2 train/test split; no independent external validation Only two videos/two subjects; event data were re-recorded from displayed videos; fixed-length event representation; no external validation; target is WBC presence rather than capillaroscopic disease-pattern assessment.
S10	2022; Microcirculation	General/physiological microcirculation	Dynamic/flow analysis	Conventional U-Net for vessel segmentation; Radon transform; frequency spectrum analysis; custom MATLAB pipeline	36 healthy subjects: young adults aged 15–24 and middle-aged adults aged 41–61; 104 U-shaped capillaries analyzed: 74 from young subjects and 30 from middle-aged subjects; videos acquired for about 10 seconds	No ML train/test validation for diagnostic prediction; applied to healthy young vs middle-aged group comparison AI component is limited mainly to vessel segmentation; downstream flow analysis is deterministic image processing; healthy participants only; no disease classification or external validation; segmentation model training details are limited.

Study ID	Year / source	Application domain	Principal task	Model family	Study size	Validation / key limitation
S11	2022; Clinical and Experimental Rheumatology	SSc/Raynaud spectrum	Detection/quantification	RetinaNet object-detection models in MMDetection for detecting capillaries/haemorrhages and classifying capillary shape/size; Stacked DenseNet key-point detection models for apical diameter and limb-width measurements; auxiliary visual model for enlarged/giant classification when calibration unavailable	Patients with Raynaud's phenomenon, primary and secondary; exact participant count not reported; 2713 manually annotated and validated images; 15% held out for testing; object annotations included 23753 capillaries, 4426 tortuosities, 640 ramifications, 1081 giant capillaries and 2100 haemorrhages; measurement dataset included 15352 training and 1690 test capillaries with three measurements	15% image test set for object-detection models; key-point measurement models tested on 1690 annotated capillaries; performance compared with manual annotations/measurements No independent external validation; exact participant count not reported; performance was lower for ramifications due to limited examples; measurements require calibration for physical size interpretation; some low-magnification images are not suitable for detailed analysis; software creators are among the authors.
S12	2022; Proceedings of SPIE: Computational Biophysics and Nanobiophotonics	General/physiological microcirculation	Dynamic/flow analysis	UNet for semantic capillary segmentation; GoogLeNet or ResNet18 as CNN feature extractor; BiLSTM/LSTM regression network for RBC velocity approximation	Patients with rheumatic diseases and healthy volunteers were mentioned, but exact number was not reported in this paper; Videos recorded at 800×800 pixels and 150 fps; frames cropped to 224×224 pixels; about 300 cropped frames used for UNet training; 3 capillaries used for training/validation/test procedures	Three cropped capillary sequences were used for training, validation and testing; the test set was obtained from vessel/capillary 3; exact split proportions were not reported Conference proceeding with limited methodological detail; exact participant/sample counts and split sizes were not reported; the semi-automatic method requires deterministic velocity-label generation and estimates mean RBC velocity but not accurate beat-to-beat RBC pulsations.

Study ID	Year / source	Application domain	Principal task	Model family	Study size	Validation / key limitation
S13	2023; Rheumatology	SSc/Raynaud spectrum	Detection/quantification	U-Nets for apex candidate generation and vessel segmentation; ResNet34 for candidate classification and width estimation; logistic regression for subject-level SSc probability	309 subjects across groups: group A 111 for training, group B 132 high-resolution test, group C 66 low-cost microscope test; roughly half SSc and half normal/PRP/healthy controls; Group A: 910 mosaics; group B: high-resolution test mosaics/images; group C: low-cost microscope images; patch datasets used for U-Net and ResNet34 training	Trained on group A; tested on independent group B and low-cost device group C; bootstrap confidence intervals Focused on SSc vs normal/PRP/healthy controls; microhaemorrhages were not included; wider external validation across centres/devices and other diseases remains needed.
S14	2023; Rheumatology	SSc/Raynaud spectrum	Classification/prediction	Vision Transformer	234 EUSTAR patients with established SSc and 55 VEDOSS patients; Approximately 17,126 NFC images; five-fold cross-validation; 464-image reliability set	5-fold cross-validation plus 464-image reliability set compared with rheumatologists Single institution; single capillaroscope type; no healthy controls or primary RP cohort; routine clinical labels may be imperfect; no longitudinal assessment; final diagnosis still requires experienced observer judgment.

Study ID	Year / source	Application domain	Principal task	Model family	Study size	Validation / key limitation
S15	2023; Clinical and Experimental Rheumatology	SSc/Raynaud spectrum	Detection/quantification	Deep convolutional neural network-based automated software; first-phase model for locating/counting capillaries and microhaemorrhages; second-phase model for apical/limb measurements; auxiliary visual model for dilated/giant capillaries	Raynaud's phenomenon patients; 1,164 NVC images externally validated; algorithm training dataset described as 2,713 manually annotated and validated NVC images from primary and secondary RP patients	External validation against expert consensus; separate validation images not used for training Consensus was not achieved in a substantial number of images; abnormal-shape category had few cases; validation is image/object-level rather than direct patient-level diagnosis.
S16	2023; Microvascular Research 150 (2023) 104593	General/physiological microcirculation	Detection/quantification	Improved YOLOv5 using 9mosaic augmentation, SPPCSPC, bilinear interpolation and EIOU loss; compared with original YOLOv5, YOLOv6 and Simple-Faster R-CNN	30 healthy volunteers; 1788 nailfold images captured; 200 high-quality images selected for deep-learning experiments	Training/testing/validation split of 6:2:2; 150 epochs for model comparison Healthy-volunteer, single-site study with a selected image subset; no disease-classification cohort or independent external validation.
S17	2023; Journal of Diabetes	Diabetes/metabolic	Classification/prediction	CNNs from the ResNet family including ResNet-50, ResNet-101, ResNeXt-50 and ResNeXt-101 selected through hyperparameter optimization	120 adults aged 35–75 years: 60 with diabetes, including 30 type 1 and 30 type 2 diabetes, and 60 without diabetes; recruitment stratified by diabetes and cardiovascular disease status; 5326 nailfold images reported in Results/Table 1, mean 44.4 images per participant; abstract reports 5236 images	Five-fold patient-level cross-validation with non-overlapping patient subsets; all images from a patient kept within the same fold Small proof-of-concept study; selected population with many male and White participants; no external validation; cardiovascular and retinopathy outcomes were history-based rather than actively measured; internal image-count inconsistency appears between abstract and results.

Study ID	Year / source	Application domain	Principal task	Model family	Study size	Validation / key limitation
S18	2023; IAES International Journal of Artificial Intelligence	Diabetes/metabolic	Detection/quantification	YOLOv3 object detector with SqueezeNet backbone; YOLOv2 comparison; statistical scoring model for classification	Participant count not reported; 600 training images and 205 test images; training images split among healthy, diabetic and hypertensive subjects; hypertensive subset was small	600 training images and 205 test images; no cross-validation or independent external validation reported Low-cost USB microscope caused poor contrast in some images; participant-level details were not reported; hypertensive sample was too small for classification; no external validation; downstream classification used a statistical model rather than an end-to-end diagnostic AI model.
S19	2023; The Journal of Pediatrics	Neonatal ROP	Classification/prediction	Machine-learning pixel-level vessel classifier using Ilastik; REAVER vascular analysis in MATLAB; mixed-effects linear regression	32 analyzable infants born before 33 weeks gestation; 37 recruited and 5 excluded for poor image quality; 1386 nailfold capillary network images reported in abstract; mixed-effects model used 1368 images from 32 infants; paired time-point subgroup included 25 infants	No conventional AI train/test split for clinical prediction; Ilastik segmentation trained on 15 labeled images; paired time-point and mixed-effects regression analyses performed Small neonatal cohort with only 3 infants requiring ROP therapy; ROI selection was manual; segmentation was trained on a small set of labeled images; no independent external validation; this is biomarker prediction rather than conventional rheumatologic NVC pattern classification.

Study ID	Year / source	Application domain	Principal task	Model family	Study size	Validation / key limitation
S20	2023; Biomedical Physics & Engineering Express	General/physiological microcirculation	Dynamic/flow analysis	Modified U-Net spatial stream; dense optical-flow temporal stream; two-stream fusion network; CBAM variants tested; ST image construction and HoughLinesP line detection for velocity estimation	22 volunteers; About 30000 nailfold capillary images collected; 539 selected images split about 9:1 into training and test datasets; 51 additional independent validation images; five 10-second sequences from the same volunteer/finger/field of view used for application validation	539 selected images randomly split approximately 9:1 into training and test datasets; 51 independent validation images with no overlap; application validation used repeated five-time-point recordings of the same volunteer/finger/field of view Volunteer-only dataset; no disease cohort or clinical diagnostic validation; no independent external device/population validation; ST images remained noisy and trajectory detection errors affected velocity estimation; validation of blood-flow velocity was limited.
S21	2023; 2023 3rd International Conference on Bioinformatics and Intelligent Computing (BIC 2023), ACM	General/physiological microcirculation	Segmentation	U-Net backbone; Attention U-Net with CBAM, CA, ECA and SE modules; residual U-Net; Res_CBAM_Unet selected as best model	22 volunteers; 613 selected nailfold capillary images plus 51 independent validation images; augmented dataset contained 3678 images; DRIVE retinal vessel dataset was also used for cross-domain generalization testing	613 images were augmented to 3678 and randomly split into training/test sets at 9:1; 51 independent nailfold images were used as a separate validation dataset; DRIVE training/test split used for cross-domain vessel-segmentation testing Conference proceeding with a relatively small private volunteer dataset; no independent external nailfold dataset or clinical diagnostic validation; acquisition device and clinical spectrum were not fully reported; remaining limitations include noise sensitivity and dependence on image quality.

Study ID	Year / source	Application domain	Principal task	Model family	Study size	Validation / key limitation
S22	2024; Pediatric Research	Dermatomyositis/JDM	Classification/prediction	Lightweight explainable CNN model named NFC-Net with integrated gradients for visual attribution	142 children: 111 patients with juvenile dermatomyositis and 31 healthy controls; 1120 NFC images from JDM patients and 321 NFC images from healthy controls; disease-activity modeling used JDM images with DAS total, muscle and skin activity labels	Stratified 5-fold cross-validation; image-level testing and patient-level aggregation Proof-of-concept study with limited sample size; image quality and acquisition conditions varied; no external/multicentre validation; DAS has limitations for disease quiescence; persistent capillary damage may reduce disease-activity prediction accuracy.
S23	2024; Diagnostics	SSc/Raynaud spectrum	Detection/quantification	YOLOv8s object-detection model deployed in a Tkinter/Firebase single-board-computer system	800 subjects were identified; images from 23 subjects were excluded; nailfold capillary images from 777 patients/subjects were collected; 777 nailfold capillary images; 80% used for training and 20% for testing	80% training and 20% testing split No disease-specific classification; normal vs abnormal status was not collected; different capillary morphology types were not examined; no external validation; training requires more computing power.
S24	2024; Reumatologia	SSc/Raynaud spectrum	Classification/prediction	ResNet-34 deep residual neural network for normal/pathological classification; Darknet-YOLOv3 for capillary detection/counting; fast.ai/PyTorch and GPU-optimized environment	Not clearly reported at participant level; 200 capillaroscopic images: 100 normal and 100 pathological images	10-fold cross-validation with validation and test sets Pilot study; small dataset; no external validation; limited binary classification; larger database needed for multiple capillary features; requires GPU resources; data remain with authors.

Study ID	Year / source	Application domain	Principal task	Model family	Study size	Validation / key limitation
S25	2024; Microvascular Research	General/physiological microcirculation	Segmentation	Transformer–U2Net; compared with U-Net, Res-Unet, U2-Net, CBAM ablation and other published segmentation methods	22 healthy volunteers aged 18–30 years; 687 images captured; 466 filtered out for artifacts/overexposure/blur; 221 normal images retained; 200 high-quality images used for deep-learning experiments	Random 6:2:2 split into training, test and validation sets; models trained for 500 epochs Healthy-volunteer dataset only; no disease cohort or clinical diagnostic validation; no external validation; dataset creation and annotation are resource-intensive; image quality depends heavily on acquisition equipment.
S26	2024; Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine	Diabetes/metabolic	Segmentation	CNN using U-Net for capillary segmentation, compared with local thresholding, Otsu thresholding and morphological-operation-based segmentation; active contour and contour approximation used for morphology measurement	118 participants: 80 normal/non-diabetic and 38 diabetic participants; 118 nailfold capillaroscopy image samples	CNN/U-Net segmentation compared with thresholding and morphological segmentation methods; no clear CNN train/validation/test split or cross-validation design reported The study does not clearly report the CNN training/validation/test split, segmentation annotation protocol, external validation, or a supervised diagnostic classification model; participant age distribution was also imbalanced, with few diabetic participants aged 20–50 years.

Study ID	Year / source	Application domain	Principal task	Model family	Study size	Validation / key limitation
S27	2025; Diagnostics	SSc/Raynaud spectrum	Classification/prediction	MobileNetV3Large, ResNet152V2, Xception, VGG-19, InceptionV3 and NASNetLarge with ImageNet transfer learning	50 SSc patients and 30 healthy controls; SSc classified by 2013 ACR/EULAR criteria; 1,280 images collected; 191 unclassifiable and 112 low-quality excluded; 977 usable images; balanced final dataset 732 images: 245 normal, 174 active, 164 early and 149 late; train 513, validation 147, test 72	70/20/10 train/validation/test split with five repeated training/evaluation runs Images without rheumatologist agreement and low-quality images were excluded, creating possible selection bias; single centre/device; relatively small final test set; no external validation; risk of overfitting despite repeated runs.
S28	2025; Scientific Reports	General/physiological microcirculation	Classification/prediction	CE-NFCNet using EfficientNet-B0 with cascade transfer learning; compared with SE-NFCNet and CC-NFCNet	225 participants aged 18–65 years; 81 males and 144 females; healthy volunteers or diseased participants; 361 NFC images collected; 73 excluded; final dataset 288 images: 16 normal and 272 abnormal; 255 images used for domain-specific pretraining, 21 for target fine-tuning and 12 for final testing	3-fold cross-validation on target fine-tuning set plus independent 12-image test set Very small test set of 12 images; severe class imbalance with only 16 normal images; abnormal class is heterogeneous and not disease-specific; no external validation; perfect performance should be interpreted cautiously.

Study ID	Year / source	Application domain	Principal task	Model family	Study size	Validation / key limitation
S29	2025; Artificial Life and Robotics	General/physiological microcirculation	Classification/prediction	CNN with transfer learning from DRIVE fundus vessel images; compared with training from scratch and ImageNet-based transfer learning	6 healthy male subjects in their 20s; DRIVE fundus dataset used for source-domain pretraining; 80 nailfold images; 64 training and 16 validation images; 485 annotated vascular regions: 390 untwisted and 95 twisted; 522 fundus images used for pretraining	4:1 train/validation split; experiments repeated five times with different random seeds Healthy-only target dataset; small dataset; non-expert annotations; no disease labels; no external validation; clinical usefulness remains unvalidated.
S30	2025; Rheumatology	SSc/Raynaud spectrum	Classification/prediction	CatBoost-based CAPI-Detect algorithm with three hierarchical ML models	1780 capillaroscopies initially; 1724 valid for manual/software analysis; 1490 with partial/full consensus; 515 with full consensus; 48,199 images and 596,611 capillaries initially; 47,233 images and 585,404 capillaries valid for analysis; 298 capillaroscopies reserved for validation	Training/validation split about 80/20; validation on partial+full consensus and full-consensus-only subsets Consensus-labeled retrospective/routine images; no fully independent external validation beyond the source hospitals; algorithm assigns capillaroscopic pattern, not a direct SSc clinical diagnosis; dependent on software-extracted variables and complete capillaroscopy image sets.
S31	2025; Computers in Biology and Medicine	SSc/Raynaud spectrum	Classification/prediction	GoogLeNet; ResNet-50; DenseNet; Sugeno fuzzy integral model integration	Patient count not reported; 196 nailfold capillary images: 86 normal and 110 abnormal	80% training and 20% testing split Small single-centre image dataset; patient count and detailed labeling protocol were not clearly reported; no external validation; binary normal/abnormal classification rather than clinically detailed disease-pattern staging.

Study ID	Year / source	Application domain	Principal task	Model family	Study size	Validation / key limitation
S32	2025; Microvascular Research	SSc/Raynaud spectrum	Classification/prediction	Machine learning: RF, SVM, LR, LGB, XGB, KNN using handcrafted features; deep learning: pretrained ResNet-18, DenseNet-121 and VGG-16 from TIMM library	100 SCLEROCAP patients; 50 had SSc signs on nailfold capillaroscopy; Approximately 747 finger-level NC images: right hand 179 SSc-positive and 188 negative; left hand 191 SSc-positive and 189 negative	70% training and 30% testing; 5-fold cross-validation for hyperparameter testing Only the first 100 SCLEROCAP patients were used; image quality was heterogeneous; no independent external validation; findings need confirmation on the full SCLEROCAP image set and larger labeled datasets.
S33	2025; Rheumatology	Dermatomyositis/JDM	Classification/prediction	Unsupervised hierarchical clustering analysis; not an automated image-analysis model	76 dermatomyositis patients; 26 anti-synthase antibody syndrome patients and 33 systemic sclerosis controls; Exact number of NVC images/videos not reported; NVC data from 76 DM patients and comparator groups	Hierarchical clustering of 61 DM patients with survival outcomes; k=2 selected; no supervised train/test split Retrospective single-centre study; NVC features were manually evaluated rather than automatically extracted; exact image/video counts not reported; no external validation; clustering was exploratory and not a validated diagnostic model.
S34	2025; Frontiers in Immunology	Other autoimmune/inflammatory	Classification/prediction	Unsupervised k-means clustering; deep embedded clustering model; factor analysis; not an image-segmentation or image-classification AI model	72 analyzed subjects: 18 PsA, 16 psoriasis, 19 rheumatoid arthritis and 19 healthy controls; NVC assessment of digits 2–5 of both hands; exact number of NVC images not reported; US and NVC variables averaged across assessed digits	K-means and DEC exploratory clustering; 10 random initializations used for clustering validation; no train/test diagnostic validation for a supervised model Small pilot single-centre cohort; clustering was exploratory and feature-based; NVC parameters were manually/clinically assessed rather than automatically extracted from images; DEC added no value; no external validation or diagnostic prediction model was developed.

Study ID	Year / source	Application domain	Principal task	Model family	Study size	Validation / key limitation
S35	2025; 2025 IEEE 22nd International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), Hammamet, Tunisia	General/physiological microcirculation	Classification/prediction	SVM; ANN with one hidden layer of 10 neurons; K-NN with k=9; hand-crafted morphological feature extraction	103 patients: 75 females and 28 males; age information not collected; 309 nailfold capillaroscopy images: 135 healthy samples and 174 unhealthy samples	Random image-level split with 66% training and 33% testing; no patient-level split or cross-validation reported Conference proceeding with small private single-centre dataset; random image-level splitting likely caused data leakage and overfitting; no external validation, no patient-level split, manual ROI selection, no age data, and simplified healthy/unhealthy labels rather than diagnosis-specific classification.
S36	2025; Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (KDD 2025)	Diabetes/metabolic	Dataset/benchmark	Improved YOLOv8 and other object detectors for morphology detection; CNNs including VGG-11, ResNet18, ResNet50 and MobileNet plus Vision Transformer for diabetes classification; benchmarked YOLOv3-v11, Faster R-CNN, EfficientDet and RT-DETR++	202 participants: 126 T2DM patients, mean age 63.79±8.81 years, and 76 healthy controls, mean age 23.94±2.26 years; 6695 images total: 3283 images from diabetic subjects and 3412 images from healthy/non-diabetic participants according to the abstract and dataset section; 1279 refined images for morphology detection after filtering duplicates and low-quality images; 120 videos from 8 participants	8:1:1 train/validation/test split by participant for morphology and diabetes tasks; benchmark experiments on NVIDIA RTX 4090 using PyTorch 2.1.1 and CUDA 11.8 Conference proceeding; healthy controls were much younger than T2DM patients; dataset may be biased; the text contains an internal class-count inconsistency in one subsection; external clinical validation and stronger disease-level diagnostic evaluation are lacking; annotation inter-rater reliability is not reported.

Study ID	Year / source	Application domain	Principal task	Model family	Study size	Validation / key limitation
S37	2026; Microvascular Research	General/physiological microcirculation	Image quality/preprocessing	Improved MIMO-UNet with Semantic Residual Feedback Block and Blur Region Attention-Aware module; compared with baseline MIMO-UNet, DeblurGAN, MPT and FFTformer	50 male and female volunteers; 400 paired clear/blurred nailfold capillary image pairs; 280 training, 60 validation and 60 test pairs; 70 capillaries used for static-parameter consistency validation	Train/validation/test split of 280/60/60 image pairs; no independent external validation It is mainly a preprocessing/restoration study rather than diagnostic classification or disease-pattern recognition; volunteers were apparently healthy; no external clinical validation; clinical impact on final NVC interpretation was not directly tested.
S38	2026; RMD Open	Other autoimmune/inflammatory	Classification/prediction	XGBoost multiclassification model; SHAP interpretability	396 participants after quality control: 65 controls, 229 RA and 102 SLE; 600 NVC images: 117 control, 337 RA and 146 SLE images	7:3 train/test split; fivefold cross-validation; bootstrap 95% confidence intervals ML was based on manually annotated NVC features rather than fully automated image analysis; no independent external validation; test performance was moderate for RA and SLE; limited to RA/SLE/control classification.
S39	2026; Neural Computing and Applications	SSc/Raynaud spectrum	Detection/quantification	YOLOv8; YOLOv9; YOLOv10; YOLOv11; YOLOv12	Original participant count not reported; dataset-level study; 321 annotated nailfold capillaroscopy images; 224 training, 64 validation and 33 test images	Training/validation/test split 224/64/33; 5-fold cross-validation; identical training settings across models Small public dataset; limited disease-spectrum and clinical metadata; no prospective or multicentre clinical validation; performance metrics are modest; external real-world generalizability remains uncertain.

Study ID	Year / source	Application domain	Principal task	Model family	Study size	Validation / key limitation
S40	2026; Computer Methods and Programs in Biomedicine	General/physiological microcirculation	Dynamic/flow analysis	Two-stage YOLOv8n object-detection framework for vessel localization and WBC detection; Flow-Guided Feature Aggregation for temporal consistency; ByteTrack for multi-object tracking and counting	22 volunteers: 12 males and 10 females; Vascular dataset included 57373 frames; WBC dataset included 7225 frames from 22 subjects	Subject-combination-based train/validation/test splits to reduce leakage; vessel datasets blood_vessel0-2 and WBC datasets WBC0-3 evaluated; WBC1 selected as final WBC model Evaluation used 22 volunteers, a single custom device and manual image annotations rather than absolute leukocyte counts from hematology analyzers; no diseased hematologic/immune cohorts or external validation were included; reported WBC F1 values are internally inconsistent.
S41	2026; Revista da Associação Médica Brasileira	SSc/Raynaud spectrum	Classification/prediction	Pretrained Vision Transformer google/vit-base-patch16-224 from Hugging Face; ImageNet-trained model applied without fine-tuning or domain-specific training	No original patient cohort; 104 anonymized public images were analyzed; 104 images: 23 normal and 81 pathological nailfold capillaroscopy images	No training/validation/testing of the AI model; retrospective evaluation of an off-the-shelf pretrained ViT on 104 images Small retrospective public-image dataset; no domain-specific training or fine-tuning; no external prospective validation; public image sources may be heterogeneous; the article contains an obvious placeholder/error in the conclusion text, although methods/results report 104 images.
S42	2026; Transactions of the Korean Institute of Electrical Engineers	General/physiological microcirculation	Segmentation	Six semantic segmentation models were evaluated: U-Net, SegNet, Attention U-Net, UNet3+, DeepLabV3+, and ResUNet++; U-Net achieved the best performance.	not reported; not reported	Internal evaluation only; no external validation reported. Dataset size and participant diversity were limited; participants were from a single age group; smartphone camera quality may affect image quality.

Study ID	Year / source	Application domain	Principal task	Model family	Study size	Validation / key limitation
S43	2026; Medicina Clínica	ATTRwt amyloidosis	Classification/prediction	CatBoost Machine-learning-based models using software-extracted NVC variables	32 ATTRwt patients; 19 had repeat NVC 12 months after TTR-stabilizing therapy; 99 HFpEF controls; 51 ATTRwt NVC examinations implied from 32 baseline plus 19 follow-up NVCs; 30 NVCs used for model training and 21 for validation; 99 HFpEF controls included for amyloidosis-profile discrimination	Small sample size; internal validation only; possible overfitting due to repeated NVCs from some patients; elderly cohort with comorbidities; no external validation.

C.3 Reported performance metrics and main contributions

Supplementary Table S4: Reported performance metrics and main contribution of the 43 included primary studies. Study IDs S01–S43 correspond to the same studies listed in Supplementary Table S3.

Study ID	Study	Metrics reported	Main result and interpretation
S01	Michael Berks et al. (2014): An automated system for detecting and measuring nailfold capillaries	Precision, recall, F-measure, distal/non-distal accuracy, Cohen's kappa, Wilcoxon rank-sum tests for group differences	Against human-observer consensus on the 455-image test set, the automated system achieved 64.1±1.8 precision, 80.9±0.7 recall and 71.5±1.1 F-measure for capillary detection, with 93.6±1.0% accuracy and Cohen's kappa=0.690±0.034 for distal/non-distal labelling; automated measurements showed significant group differences between SSc, PR and HC for capillary density, median apical width and tortuosity, except PR vs HC vessel density. This foundational study showed that a fully automated layered machine-learning pipeline can extract quantitative nailfold capillaroscopy biomarkers with expert-comparable capillary detection and clinically meaningful group-level differences.
S02	Niraj P. Doshi et al. (2016): MD-LBPV texture analysis for nailfold capillaroscopy image classification	Finger classification accuracy; patient classification accuracy; comparison with earlier multi-scale LBPV approach	The MD-LBPV approach achieved 73.17% finger-level classification accuracy and correctly classified 10 of 12 patients, giving 83.33% patient-level accuracy; the earlier LBPV approach correctly classified 8 patients and produced one rejected diagnosis. This study is relevant because it applies a classical ML pipeline to automated scleroderma-pattern classification in nailfold capillaroscopy images and improves over an earlier texture-analysis method.
S03	Shupeng Liu et al. (2020): Segmenting nailfold capillaries using an improved U-net network	Accuracy; Dice score; precision-recall curve; ROC/AUC; training accuracy and loss	Res-Unet18 achieved mean accuracy 91.72% and mean Dice score 97.66%, outperforming U-Net accuracy 86.24% and Dice 97.29%. The reported AUC for Res-Unet18 was 0.9027, and qualitative examples showed better segmentation of crossed and low-contrast capillaries. Relevant because it uses deep learning directly for nailfold capillary segmentation and supports automated quantitative assessment beyond manual capillaroscopy.

Study ID	Study	Metrics reported	Main result and interpretation
S04	Byeonghwi Kim et al. (2020): Automated White Blood Cell Counting in Nailfold Capillary Using Deep Learning Segmentation and Video Stabilization	Mean IoU validation accuracy; event-count agreement/qualitative comparison with expert counts; prediction consistency	Deep-learning segmentation achieved 91.11% validation accuracy by mean IoU. The stabilized DNN method produced WBC event counts close to expert medians and correctly detected event locations, outperforming conventional segmentation and methods without stabilization. Relevant because it applies deep learning directly to nailfold capillary image/video analysis and extends beyond rheumatologic SSc diagnosis to noninvasive WBC monitoring.
S05	Yuli Sun Hariyani et al. (2020): DA-CapNet: Dual Attention Deep Learning Based on U-Net for Nailfold Capillary Segmentation	IoU/Jaccard index; precision; recall	DA-CapNet achieved the best overall segmentation performance, with IoU 0.6431 compared with 0.218 for adaptive Gaussian thresholding, 0.556 for SegNet and 0.559 for U-Net; precision and recall were also reported as best for DA-CapNet in box-plot comparisons. This study is relevant because it is a direct deep-learning nailfold capillary segmentation paper using a U-Net attention architecture.
S06	Suma K V et al. (2022): Deep Learning Approach to Nailfold Capillaroscopy Based Diabetes Mellitus Detection	Average precision; accuracy; sensitivity; specificity; precision-recall curves	YOLOv3 outperformed YOLOv2, with AP 0.74 for normal, 0.80 for elongated, 0.85 for wide, 0.99 for tortuosity and 0.91 for hemorrhages. The final proposed diabetes model achieved accuracy 88.2%, sensitivity 89% and specificity 89%, higher than the earlier DCT/WT model accuracy of 73.3%. Relevant because it applies deep-learning object detection directly to NFC images for diabetes mellitus, a non-SSc application.
S07	Ruiqi Liu et al. (2022): Nailfold Microhemorrhage Segmentation with Modified U-Shape Convolutional Neural Network	IoU; Dice score; qualitative segmentation comparison	DAFM-Net with group normalization achieved IoU 78.03% and Dice 87.34%. U-Net baseline achieved IoU 54.37% and Dice 63.89%; attention mechanisms improved performance, and DAFM with GN performed best. Relevant because it applies deep learning directly to nailfold microhemorrhage segmentation, an abnormality relevant across diseases beyond only systemic sclerosis.
S08	Hung-Hsiang Wang et al. (2022): Applying t-SNE to Estimate Image Sharpness of Low-cost Nailfold Capillaroscopy	Laplacian score; t-SNE visualization/cluster distribution; qualitative comparison before/after deconvolution	The t-SNE visualization showed a sharp-image distribution close to the Laplacian/statistical method; deconvolution increased Laplacian sharpness scores compared with original images and the software sharpening function. Relevant as an AI/ML-related NFC image-quality-control paper, but it is not a diagnostic, segmentation or clinical classification model.

Study ID	Study	Metrics reported	Main result and interpretation
S09	Yuli Sun Hariyani et al. (2022): Event-based White Blood Cell Classification from Nailfold Capillaries	Accuracy; precision; recall; F1-score	Random forest performed best with accuracy 75.51%, F1-score 75.08%, precision 75.71% and recall 75.51%; KNN and decision tree performed lower. This study is relevant as an AI/ML application using nailfold capillary data, but it is not a conventional NVC morphology or disease-pattern analysis study.
S10	Tomoya Niizawa et al. (2022): Automated capillary flow segmentation and mapping for nailfold video capillaroscopy	Morphological parameters; flow speed; coefficient of variation; frequency-spectrum AUC; t-test/Wilcoxon; correlation coefficients	Curved-segment diameter and straight-segment interval differed significantly between young and middle-aged subjects; arterial-side flow speed was faster in middle-aged subjects than young subjects, about 0.48 mm/s vs 0.26 mm/s; high-frequency power spectrum correlated with flow speed. This study is relevant because it uses U-Net-based automated segmentation and computational mapping of nailfold capillary flow dynamics, expanding AI-assisted NVC beyond static morphology.
S11	Borja Gracia Tello et al. (2022): The challenge of comprehensive nailfold videocapillaroscopy practice: a further contribution	mAP; AP at IoU 0.50 and 0.75; AR; precision; recall; key-point measurement success rate	Object detection for capillaries/haemorrhages by shape achieved AP 0.471 at IoU 0.50:0.95, AP 0.758 at IoU 0.50 and AR 0.625; capillary size detection achieved AP 0.515, AP 0.811 at IoU 0.50 and AR 0.663; with confidence >=0.40 and IoU >=0.50, capillary detection precision was 83.84% and recall 92.44%; apex measurement succeeded in 88% and limb measurements in 84% of visible test examples. This study is important because it reports a clinically oriented deep-learning software framework for comprehensive NVC analysis across multiple devices, supporting standardized automated capillary counting, classification and measurement.
S12	Alexey V. Kornaev et al. (2022): Application of deep convolutional and long short-term memory neural networks to red blood cells motion detection and velocity approximation	Accuracy of mean velocity approximation over 10 seconds for train, validation and test sets	Best non-overfitting setting used ResNet18 feature extraction with sequence length 15 and BiLSTM with 20 hidden units, achieving 96.2% train, 92.3% validation and 96.7% test accuracy for mean velocity approximation over 10 s; point-wise velocity/pulsation estimation was less accurate. This study is relevant because it applies deep learning directly to nailfold videocapillaroscopy data for capillary segmentation and RBC flow/velocity estimation, extending AI-based NFC analysis beyond morphology-only assessment.

Study ID	Study	Metrics reported	Main result and interpretation
S13	Praveen Gurunath Bharathi et al. (2023): A deep learning system for quantitative assessment of microvascular abnormalities in nailfold capillary images	ROC-AUC; equal-error-rate sensitivity/specificity; expert sensitivity and specificity	Group B AUC 97% with equal sensitivity/specificity 91%; group C AUC 95% with equal sensitivity/specificity 89%; expert consensus achieved sensitivity 82% and specificity 73%. Highly relevant because it provides a complete automated quantitative deep-learning NVC pipeline and subject-level SSc probability.
S14	Alexandru Garaiman et al. (2023): Vision transformer assisting rheumatologists in screening for capillaroscopy changes in systemic sclerosis: an artificial intelligence model	AUC; sensitivity; specificity; accuracy; human agreement; inference time	Mean AUC across folds ranged from 81.8% to 84.5% for different NFC changes; on the reliability set AUC was 92.6% for giant capillaries, 90.2% for enlarged capillaries, 86.7% for capillary loss, 85.1% for microhaemorrhages and 88.6% for scleroderma pattern; inference time was 0.19 s per image. Relevant because it applies a modern transformer architecture directly to NVC image screening and compares performance with rheumatologists.
S15	B.C. Gracia Tello et al. (2023): External clinical validation of automated software to identify structural abnormalities and microhaemorrhages in nailfold videocapillaroscopy images	Overall concordance; PPV; NPV; sensitivity; specificity	Consensus among $\geq$3 experts was achieved in 86.9% of images and 75.8% of these were correctly predicted by the algorithm; consensus among $\geq$4 experts occurred in 52.0% of cases and 87.1% matched the algorithm; NPV and specificity were $>$89% for all categories. Highly relevant because it externally validates an automated deep-learning NVC tool for detecting clinically important capillary abnormalities.
S16	Hao Yin et al. (2023): Automated nailfold capillary density measurement method based on improved YOLOv5	Precision; recall; AP@0.5; validation object loss; distal capillary count; density agreement with manual measurement	Proposed method reached AP@0.5 of 85.2%, improving over original YOLOv5, YOLOv6 and Simple-Faster R-CNN by about 4.92–4.93%, 5.24% and 107%, respectively. Density estimates were generally consistent with manual measurements under normal, stray-light, many-tiny-vessel and overexposed conditions. Peer-reviewed AI study proposing improved YOLOv5 detection, distal-capillary filtering with the 90° method, and automated nailfold capillary-density measurement.

Study ID	Study	Metrics reported	Main result and interpretation
S17	Reema Shah et al. (2023): Nailfold capillaroscopy and deep learning in diabetes	AUROC; AUPR; sensitivity; specificity; bootstrap 95% confidence intervals	Model best identified diabetes with AUROC 0.84, AUPR 0.84, sensitivity 0.82 and specificity 0.75. Performance was AUROC 0.81 for type 1 diabetes and 0.88 for type 2 diabetes; HbA1c $\geq$6.5% AUROC was 0.77. Models showed weaker performance for hypertension, albuminuria, retinopathy and cardiovascular events, although cardiovascular event prediction among patients with diabetes reached AUROC 0.65 and AUPR 0.72. Relevant because it applies CNN-based deep learning directly to nailfold capillaroscopy images for diabetes-related classification beyond systemic sclerosis.
S18	Suma Kuncha Venkatapathiah et al. (2023): Deep learning based object detection in nailfold capillary images	Class-wise accuracy; precision-recall curves; statistical-model accuracy, specificity and sensitivity	YOLOv3 class accuracies were normal 0.54, wide 0.85, hemorrhages 0.91, tortuosity 0.99 and elongated 0.79, outperforming YOLOv2 for most classes. The downstream statistical model achieved accuracy 88.2%, specificity 89% and sensitivity 89% for healthy vs diabetic classification. This study is relevant because it applies deep-learning object detection directly to nailfold capillary images and extracts clinically interpretable capillary morphology features.
S19	Daniel York et al. (2023): Nailfold Capillaroscopy: A Promising, Noninvasive Approach to Predict Retinopathy of Prematurity	VLD, VBPD, VSC, VAF; paired/unpaired tests; sensitivity/specificity for VLD thresholds; mixed-effects regression	VLD, VBPD, VSC and VAF were significantly higher in infants who later developed ROP at both time points. VLD cutoff >14100 at T1 or >11900 at T2 detected ROP severity score $\geq$7 with 100% sensitivity and 75% specificity. VLD cutoff >15100 at T1 or T2 detected all 3 infants requiring ROP therapy; discussion reports 100% sensitivity and 65% specificity for therapy prediction. This study is relevant because it uses machine-learning-assisted nailfold capillary segmentation and vascular quantification as a noninvasive biomarker outside systemic sclerosis.

Study ID	Study	Metrics reported	Main result and interpretation
S20	Shupin Chen et al. (2023): Blood flow characterization in nailfold capillary using optical flow-assisted two-stream network and spatial-temporal image	Accuracy; Dice; IoU; VOE; loss; blood-flow velocity statistics	Proposed TwoStreamNet achieved loss 0.1772, accuracy 94.01%, Dice 0.8099, IoU 0.6806 and VOE 0.3194, outperforming U-Net with accuracy 92.83%, Dice 0.7463, IoU 0.5952 and VOE 0.4048. Median velocity in the repeated same-vessel validation decreased from 734.76 μm/s at 10 AM to 35.55 μm/s at 2 PM. This study is relevant because it applies deep learning directly to nailfold capillary image sequences for vessel segmentation and combines AI-assisted segmentation with ST-image-based blood-flow velocity quantification.
S21	Shupin Chen et al. (2023): Nailfold Capillary Segmentation Based on U-Net and Attention Mechanism	Accuracy, Dice, IoU, VOE; sensitivity/recall, specificity and Dice on DRIVE dataset	Best nailfold segmentation result was obtained with Res_CBAM_Unet plus CLAHE clipLimit=2, achieving accuracy 94.31%, Dice 0.8297, IoU 0.7090 and VOE 0.2910; on DRIVE, Res_CBAM_Unet achieved accuracy 97.13%, sensitivity 0.8478, specificity 0.9822 and Dice 0.8287. This study is relevant because it applies modern deep-learning segmentation methods directly to nailfold capillary images and evaluates U-Net variants with attention and residual mechanisms.
S22	Peyman Hosseinzadeh Kassani et al. (2024): Artificial intelligence for nailfold capillaroscopy analyses – a proof of concept application in juvenile dermatomyositis	AUROC; accuracy; sensitivity; specificity; precision; patient-level accuracy	NFC-Net distinguished JDM from healthy controls with AUROC 0.93, accuracy 0.91, sensitivity 0.85, specificity 0.90 and precision 0.95. Patient-level JDM/control accuracy was 0.833. For total DAS activity prediction, AUROC was about 0.75–0.76 and accuracy 0.74; AUROC was 0.69 for muscle DAS and 0.73 for skin DAS. Relevant because it applies explainable deep learning directly to NFC images outside systemic sclerosis, extending AI-based nailfold analysis to juvenile dermatomyositis.
S23	Seda Arslan Tuncer et al. (2024): YOLOv8-Based System for Nail Capillary Detection on a Single-Board Computer	F1 score; mAP50; precision; recall; capillary count; density; average thickness	YOLOv8s achieved F1 score 0.83 and mAP50 0.882 for capillary detection. Example outputs automatically reported capillary counts, density per mm and average capillary thickness, supporting repeatable measurement on the SBC-based interface. Relevant because it applies YOLOv8 directly to nailfold capillary images and implements an end-to-end deployable detection/measurement system.

Study ID	Study	Metrics reported	Main result and interpretation
S24	Olga Elżbieta Brzezińska et al. (2024): Automatic assessment of nailfold capillaroscopy software: a pilot study	Sensitivity; specificity; IoU; PCDI / detection precision	For correct vs pathological classification, sensitivity/specificity were 89.0%/89.4% in training and 89.0%/86.9% in validation. For capillary counting, IoU was 50.82%, precision of correctly detected images 56.37%, precision of ground-truth images detected within ROI 78.03%, and precision of real images detected within ROI 96.48%. Relevant because it applies AI directly to nailfold capillaroscopy image classification and capillary counting, but current cohort is mainly SSc/scleroderma-spectrum focused.
S25	Qianyao Ye et al. (2024): Improved nested U-structure for accurate nailfold capillary segmentation	Accuracy; Dice; IoU; recall; loss	Transformer–U2Net achieved accuracy 0.9813, Dice 0.8852 and IoU 0.8039; it outperformed U-Net, Res-Unet and U2-Net, and Table 3 reports better performance than Ronneberger et al., Liu et al., Qin et al. and Chen et al. comparison methods. This study is relevant because it develops a modern Transformer-enhanced deep-learning model for direct nailfold capillary image segmentation.
S26	Sowmiya Elumalai et al. (2024): Analysis of microvascular pattern in diabetes mellitus condition using the nailfold capillaroscopy images	MSE; PSNR; SSIM for preprocessing; Jaccard index and Sorensen/Dice coefficient for segmentation; capillary diameter, density, distribution, shape and hemorrhage presence for morphology	Wiener filtering performed best for preprocessing with MSE 0.002, PSNR 42.117 and SSIM 0.905. CNN/U-Net segmentation performed best among tested segmentation methods with Jaccard index 0.78 and Sorensen/Dice coefficient 0.85. Among diabetic participants, 60% had tortuous capillaries, 15% cross-linked capillaries, 20% giant capillaries and 5% hemorrhage. This study is relevant because it applies CNN/U-Net-assisted nailfold capillary segmentation and quantitative morphology analysis to diabetes-related microvascular changes, extending AI-based NFC applications beyond systemic sclerosis.
S27	Müçteba Enes Yayla et al. (2025): Deep Learning Performance in Analyzing Nailfold Videocapillaroscopy Images in Systemic Sclerosis	Loss; accuracy; precision; recall; F1 score; ROC-AUC; per-class accuracy	InceptionV3 performed best with accuracy 98.95%, precision 98.94%, recall 98.80%, F1 score 98.88% and ROC-AUC 99.99%; all models achieved accuracy 90.62–98.95% and ROC-AUC 98.99–100%. Relevant because it directly evaluates multiple deep-learning models for automated SSc NVC pattern staging.

Study ID	Study	Metrics reported	Main result and interpretation
S28	Mona Ebadi Jalal et al. (2025): Abnormality detection in nailfold capillary images using deep learning with EfficientNet and cascade transfer learning	Accuracy; precision; recall; F1 score; ROC-AUC; PR curve; average precision; confusion matrix	CE-NFCNet achieved accuracy, precision, recall, F1 score and ROC-AUC of 1.00 on the 12-image test set; SE-NFCNet and CC-NFCNet each achieved 0.67 accuracy, precision, recall and F1 score, with ROC-AUC 0.83. Relevant because it applies deep learning directly to NFC abnormality detection and addresses small/imbalanced datasets through cascade transfer learning.
S29	Rio Ishiguro et al. (2025): Shape classification of nailfold capillaries using convolutional neural network	Precision; recall	Fundus-based transfer learning achieved the best performance with precision 0.929±0.014 and recall 0.935±0.017, outperforming ImageNet-based transfer and nailfold-only training. This study is relevant because it applies CNN-based transfer learning directly to nailfold capillary morphology classification.
S30	Gema M. Lledó-Ibáñez et al. (2025): CAPI-Detect: machine learning in capillaroscopy reveals new variables influencing diagnosis	AUPRC; ROC-AUC; precision; recall; accuracy; match/discrepancy analysis	Accuracy for distinguishing SSc from non-SSc was 0.912, SSc-pattern subtype classification 0.812 and normal vs non-specific 0.746 in partial/full consensus validation; in full-consensus-only validation, accuracies improved to 0.910, 0.925 and 0.933; ROC-AUC for SSc vs non-SSc was about 0.97–0.98. This study is highly relevant because it combines automated capillaroscopy feature extraction with machine learning for clinically meaningful NVC pattern classification.
S31	Chiao-Chi Ou et al. (2025): Classification method for nailfold capillary images using an optimized Sugeno fuzzy ensemble of convolutional neural networks	Accuracy; recall; precision; F1-score	The Sugeno fuzzy ensemble achieved accuracy 85%, recall 81.82%, precision 90% and F1-score 85.17%; individual GoogLeNet, ResNet and DenseNet accuracies were 73.33%, 67.96% and 70.83%, respectively. This study is relevant because it applies CNN-based deep learning directly to nailfold capillaroscopy images for automated normal/abnormal classification.
S32	Lutfi Ozturk et al. (2025): Analysis of nailfold capillaroscopy images with artificial intelligence: Data from literature and performance of machine learning and deep learning from images acquired in the SCLEROCAP study	F1-score; accuracy; AUC; Wilcoxon signed-rank test	Machine-learning F1 scores were highest with RF, LGB and XGB, around 0.75–0.79; best ML result was LGB with combined variables, F1 0.79±0.05. Deep learning performed best with DenseNet-121: accuracy 0.94±0.02, F1 0.94±0.02 and AUC 0.95±0.03. This study is relevant because it directly compares machine-learning and deep-learning approaches on expert-labeled nailfold capillaroscopy images for SSc-landscape detection.

Study ID	Study	Metrics reported	Main result and interpretation
S33	Hui Xu et al. (2025): The role of nailfold video-capillaroscopy in the assessment of dermatomyositis	Cluster membership; P values for clinical associations; survival comparison	Abnormal NVC was found in 73.7% of DM patients. Hierarchical clustering separated DM patients into two clusters; cluster 1 had higher likelihood of ILD (P=0.035), while survival did not differ significantly between clusters (P=0.719). This study is relevant only as a borderline feature-level ML study because it uses unsupervised clustering on NVC-derived features, not direct AI-based automated image analysis.
S34	Giacomo Cafaro et al. (2025): Combined semiquantitative nail-enthesis complex ultrasonography and capillaroscopy in psoriasis and psoriatic arthritis	Silhouette score; cluster membership; Bartlett's test; KMO; factor eigenvalues/loadings; group comparisons using Kruskal-Wallis and chi-square tests	K-means outperformed DEC with silhouette score 0.27 vs 0.17. Clustering separated many healthy controls from PsA subjects and suggested overlap/continuum between psoriasis and PsA. Significant group differences were found for total GS BUNES, plate GS BUNES, bed GS BUNES, Wortsman classification and NVC tortuosities. This study is relevant only as a borderline AI/ML feature-level analysis because it applies unsupervised learning to NVC-derived parameters, but it does not perform automated AI image analysis of nailfold capillaroscopy images.
S35	Dorra Boudaouara et al. (2025): Automated Classification of Nailfold Capillaroscopy Images Using Morphological Feature Extraction	Accuracy, precision, recall, F1-score, ROC/AUC	SVM and ANN both achieved 100% accuracy, precision, recall and F1-score with AUC=1.00; K-NN with k=9 achieved 78.43% accuracy, 78.78% precision, 78.43% recall, 78.60% F1-score and AUC≈0.7686; the authors explicitly noted that the 100% results are likely unrealistic and suggest overfitting/data leakage. This study is relevant because it uses machine-learning classifiers on extracted nailfold capillaroscopy morphology features for automated healthy/unhealthy microvascular abnormality classification.
S36	Hang Thi Phuong Nguyen et al. (2025): Capillary Dataset: A Dataset of Nail-fold Capillaries Captured by Microscopy for Diabetes Detection	mAP@50; FPS; parameter count; top-1 accuracy; F1 score	Improved YOLOv8 achieved the best morphology detection performance with mAP@50 0.799, 2,668,266 parameters and 714 FPS; ResNet18 achieved 99.7% top-1 accuracy on the 1×1 diabetes dataset and 99.9% on the 3×3 dataset; pre-trained ViT achieved 98.56% accuracy and F1 score on 3×3 merged images. This paper is valuable because it introduces a public nailfold capillaroscopy dataset with images/videos and benchmark models for diabetes-related nailfold capillary analysis beyond systemic sclerosis.

Study ID	Study	Metrics reported	Main result and interpretation
S37	Peiqing Guo et al. (2026): Wide-field nailfold capillary image deblurring method based on an improved MIMO-UNet	PSNR; MSE; SSIM; MAE; APE; static diameter measurement error	Proposed method achieved PSNR 32.779, MSE 0.00098 and SSIM 0.975; compared with original MIMO-UNet, PSNR increased by about 3.82–3.83%, SSIM by 0.22%, and MSE decreased by 60.2%; static parameter differences for apical, arterial and venous diameters were all below 0.995 μm. This study is relevant because it applies deep learning directly to nailfold capillary images and improves image quality for downstream quantitative analysis.
S38	Jie Li et al. (2026): Machine learning-based multiclass model for autoimmune disease diagnosis and classification through nailfold videocapillaroscopy features	Macro AUC; class-specific AUC; accuracy; sensitivity; specificity; F1-score; PPV; NPV; MCC; SHAP	Macro AUC was 0.96 in training and 0.80 in testing; test class-specific AUCs were 0.92 for controls, 0.74 for RA and 0.75 for SLE; macro average accuracy was 0.90 in training and 0.73 in testing; key SHAP features included papilla shape, RBC aggregation, number of capillary loops, crossed loops and SVP. This study is relevant because it applies machine learning to NVC features for autoimmune disease classification beyond systemic sclerosis.
S39	Mohamed Tawfik et al. (2026): A Comparative evaluation of deep learning-based capillaroscopy image analysis using YOLOv8–YOLOv12 models for microvascular diagnostics	mAP@0.5; mAP@0.5–0.95; precision; recall; inference time; ANOVA	YOLOv9 achieved the best overall accuracy/precision: mAP@0.5 0.5155, mAP@0.5–0.95 0.2871, precision 0.5057 and recall 0.5359. YOLOv11 had the highest recall 0.5610. YOLOv8 was fastest with inference time about 4.06 ms/image. This study is relevant because it benchmarks modern YOLO object-detection models directly on nailfold capillaroscopy images for automated microvascular feature detection.
S40	Fan Zhang et al. (2026): Noninvasive real-time dynamic monitoring of white blood cells based on microscopic imaging and deep learning	mAP50; mAP50-95; precision; recall; F1-score; confusion matrix; processing speed	Best vessel detector blood_vessel0 achieved mean mAP50 0.812 and mean mAP50-95 0.380 across validation and test sets. Best WBC detector WBC1 achieved mean mAP50 0.710 and mean mAP50-95 0.265. Reported best F1 is internally inconsistent: the text/figure reports peak F1 0.79, while Table 5/conclusion reports 0.83. The complete pipeline processed about 100 frames/s. This study is relevant because it applies deep learning directly to nailfold microcirculation videos for noninvasive real-time cell-level monitoring, extending AI-based nailfold analysis beyond static morphology and systemic sclerosis.

Study ID	Study	Metrics reported	Main result and interpretation
S41	Neşe Çabuk Çelik et al. (2026): Comparison of artificial intelligence and rheumatologists in nailfold capillaroscopic evaluation	Clinical applicability; accuracy; sensitivity; specificity; Cohen's kappa	Pretrained ViT produced 0% clinically applicable capillaroscopy predictions and failed to generate meaningful medical classifications. Rheumatologist consensus achieved 94.2% accuracy, 97.5% sensitivity, 82.6% specificity and Cohen's kappa 0.827. Individual rheumatologist accuracy was 86.5% and 79.8%. This study is relevant as a negative validation/comparison study showing that general-purpose pretrained ViT models are not clinically useful for nailfold capillaroscopy without domain adaptation.
S42	Youngjin Park et al. (2026): Detection of Nailfold Capillaries Using Smartphone Microscopy and Deep Learning Analysis	Dice Similarity Coefficient; Accuracy	U-Net showed the best performance among the six segmentation models, with a Dice Similarity Coefficient of 96.10% and an Accuracy of 93.23%. This study proposes a mobile-based nailfold capillary analysis system combining a 3D-printed smartphone microscope with deep-learning-based segmentation, demonstrating the feasibility of accessible and real-time capillary assessment.
S43	L Pérez Abad et al. (2026): Unlocking the potential of nailfold videocapillaroscopy in diagnosing and staging wild-type transthyretin amyloidosis: A preliminary approach	Accuracy	The ML models achieved accuracies of 0.81 and 0.90 for predicting disease severity according to NT-proBNP and Cheng score, respectively; an additional model distinguishing an amyloidosis-suggestive profile from HFpEF controls achieved accuracy 0.73. This preliminary study suggests that ML-based analysis of NVC variables may support early diagnosis and staging of ATTRwt, extending AI-assisted nailfold videocapillaroscopy beyond systemic sclerosis.

C.4 Study-level validation and reproducibility indicators

Supplementary Table S5: Study-level validation and reproducibility indicators for the 43 included primary studies. Study IDs S01–S43 correspond to Supplementary Tables S3 and S4. These indicators are reported individually and are not combined into a validated quality or risk-of-bias score.

Study ID	Study	External validation	Multicentre	Expert comparison	Public code	Public data	Participant-level split
S01	Michael Berks et al. (2014)	No	No	Yes	NR	NR	NR
S02	Niraj P. Doshi et al. (2016)	No	No	No	NR	NR	NR
S03	Shupeng Liu et al. (2020)	No	No	No	NR	NR	NR
S04	Byeonghwi Kim et al. (2020)	No	No	Yes	NR	NR	NR
S05	Yuli Sun Hariyani et al. (2020)	No	No	No	NR	NR	NR
S06	Suma K V et al. (2022)	No	No	No	NR	NR	NR
S07	Ruiqi Liu et al. (2022)	No	No	No	NR	No	NR
S08	Hung-Hsiang Wang et al. (2022)	No	No	No	NR	NR	NR
S09	Yuli Sun Hariyani et al. (2022)	No	No	No	NR	NR	NR
S10	Tomoya Niizawa et al. (2022)	No	No	No	No	NR	NR
S11	Borja Gracia Tello et al. (2022)	No	Yes	Yes	No	NR	NR
S12	Alexey V. Kornaev et al. (2022)	No	No	No	NR	NR	NR
S13	Praveen Gurunath Bharathi et al. (2023)	No	NR	Yes	Yes	No	Yes
S14	Alexandru Garaiman et al. (2023)	No	No	Yes	Yes	No	NR
S15	B.C. Gracia Tello et al. (2023)	Yes	Yes	Yes	NR	NR	Yes
S16	Hao Yin et al. (2023)	No	No	Yes	NR	NR	NR
S17	Reema Shah et al. (2023)	No	No	No	Yes	NR	Yes
S18	Suma Kuncha Venkatapathiah et al. (2023)	No	No	No	NR	NR	NR
S19	Daniel York et al. (2023)	No	No	No	NR	NR	NR
S20	Shupin Chen et al. (2023)	No	No	No	NR	No	NR
S21	Shupin Chen et al. (2023)	No	No	No	NR	NR	NR

Study ID	Study	External validation	Multicentre	Expert comparison	Public code	Public data	Participant-level split
S22	Peyman Hosseinzadeh Kassani et al. (2024)	No	No	No	NR	No	Yes
S23	Seda Arslan Tuncer et al. (2024)	No	No	No	NR	No	NR
S24	Olga Elżbieta Brzezińska et al. (2024)	No	No	Yes	NR	No	NR
S25	Qianyao Ye et al. (2024)	No	No	No	NR	No	NR
S26	Sowmiya Elumalai et al. (2024)	No	NR	No	NR	NR	NR
S27	Müçteba Enes Yayla et al. (2025)	No	No	No	NR	No	NR
S28	Mona Ebadi Jalal et al. (2025)	No	No	No	NR	No	NR
S29	Rio Ishiguro et al. (2025)	No	No	No	NR	NR	NR
S30	Gema M. Lledó-Ibáñez et al. (2025)	No	Yes	Yes	NR	No	NR
S31	Chiao-Chi Ou et al. (2025)	No	No	No	NR	NR	NR
S32	Lutfi Ozturk et al. (2025)	No	Yes	Yes	NR	No	NR
S33	Hui Xu et al. (2025)	No	No	No	NR	NR	NR
S34	Giacomo Cafaro et al. (2025)	No	No	No	NR	No	NR
S35	Dorra Boudaouara et al. (2025)	No	No	No	NR	NR	No
S36	Hang Thi Phuong Nguyen et al. (2025)	No	No	No	Yes	Yes	Yes
S37	Peiqing Guo et al. (2026)	No	No	No	NR	NR	NR
S38	Jie Li et al. (2026)	No	No	No	NR	NR	NR
S39	Mohamed Tawfik et al. (2026)	No	No	No	No	Yes	NR
S40	Fan Zhang et al. (2026)	No	No	No	NR	NR	NR
S41	Neşe Çabuk Çelik et al. (2026)	No	No	Yes	NR	No	NR
S42	Youngjin Park et al. (2026)	No	NR	No	NR	NR	NR
S43	L Pérez Abad et al. (2026)	No	NR	No	NR	NR	NR

C.5 Validation and reproducibility indicators by AI task family

Supplementary Table S6: Validation and reproducibility indicators stratified by principal AI task family. Values indicate the number of studies with explicit affirmative reporting under the review's strict coding criteria.

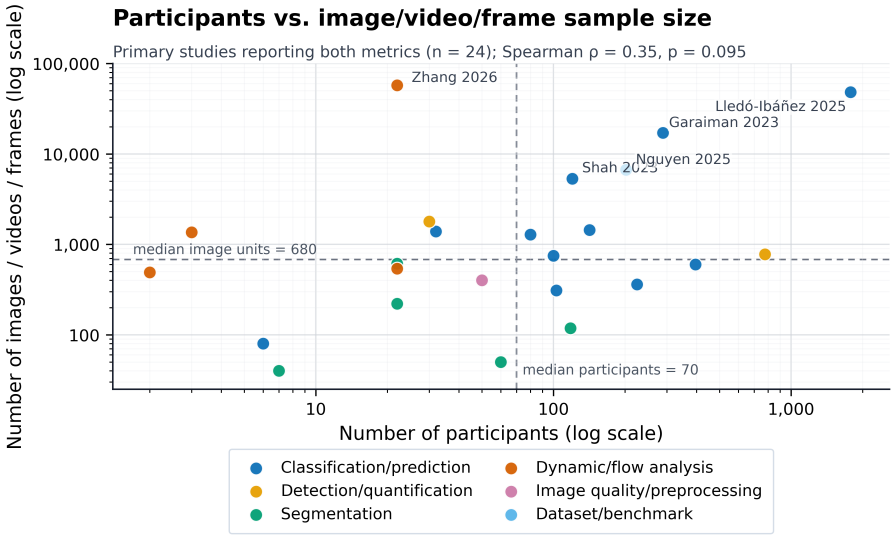
Principal AI task family	Studies	External validation	Multicentre	Expert comparison	Public code	Public data	Participant-level split
Classification/prediction	19	0	2	5	2	0	2
Detection/quantification	8	1	2	5	1	1	2
Segmentation	7	0	0	0	0	0	0
Dynamic/flow analysis	6	0	0	1	0	0	0
Image-quality/preprocessing	2	0	0	0	0	0	0
Dataset/benchmark	1	0	0	0	1	1	1
Total	**43**	**1**	**4**	**11**	**4**	**2**	**5**

Note. Indicators are descriptive evidence-mapping variables and should not be interpreted as a validated risk-of-bias, study-quality, or deployment-readiness score. Each study was assigned one principal task family according to its main reported contribution; some studies performed more than one technical function.

Supplementary Appendix D. Sample-size and denominator analyses

D.1 Participant and image/video/frame denominators

As an exploratory denominator check, we compared participant counts with the image/video/frame denominator among the 24 primary studies that reported both metrics. The relationship was weak-to-moderate and not statistically significant (Spearman $\rho = 0.35$, $p = 0.095$; Supplementary Figure S1). Because still images, video clips, extracted frames, and annotated frames are not equivalent sampling units, this coefficient should not be interpreted as a standardised estimate of effective sample size or as evidence of statistical independence. The analysis is retained only to document denominator heterogeneity and to motivate separate reporting of participant counts and technical image-unit counts. Detailed study-level denominators are provided in Supplementary Table S8, with task-family summaries in Supplementary Table S7.



Supplementary Figure S1: **Exploratory relationship between participant sample size and image/video/frame denominator among studies reporting both metrics (n = 24).** Each point represents one primary study and is coloured by the principal AI task family. Both axes are log-scaled because reported denominators varied by several orders of magnitude. Dashed grey lines show the median participant count and median image/video/frame denominator among the 24 studies. The median participant count in this 24-study subset (70) differs from the 30-participant median reported in the main text for the full set of 31 studies that reported a participant count, because the two values are computed over different subsets. Because the image-based units combine still images, video clips, extracted frames, and annotated frames, the Spearman coefficient ($\rho = 0.35$, $p = 0.095$) is descriptive only and should not be interpreted as a standardised estimate of effective sample size.

D.2 Sample-size distribution by AI task family

Supplementary Table S7: Sample-size distribution by principal AI task family among studies reporting both participant and image/video/frame counts ($n = 24$).

Principal AI task family	Studies	Participants, n		Image/video/frame units, n		Median units
	n	Median	Min–max	Median	Min–max	per participant
Classification/prediction	11	120	6–1,780	1,280	80–48,199	13.3
Detection/quantification	2	403.5	30–777	1,282.5	777–1,788	30.3
Segmentation	5	22	7–118	118	40–613	5.7
Dynamic/flow analysis	4	12.5	2–22	948.5	490–57,373	348.8
Image quality/preprocessing	1	50	50–50	400	400–400	8.0
Dataset/benchmark	1	202	202–202	6,695	6,695–6,695	33.1

Note. This table includes only studies that reported both participant counts and image/video/frame counts. The median is the central value within each task family; min–max shows the smallest and largest reported values. For task families with an even number of studies, the median may be fractional because it is calculated as the average of the two middle values. Image/video/frame units were not harmonised across studies and may refer to still images, video clips, extracted frames, or annotated frames. The median units per participant is descriptive and should not be interpreted as a standardised sampling rate. AI = artificial intelligence.

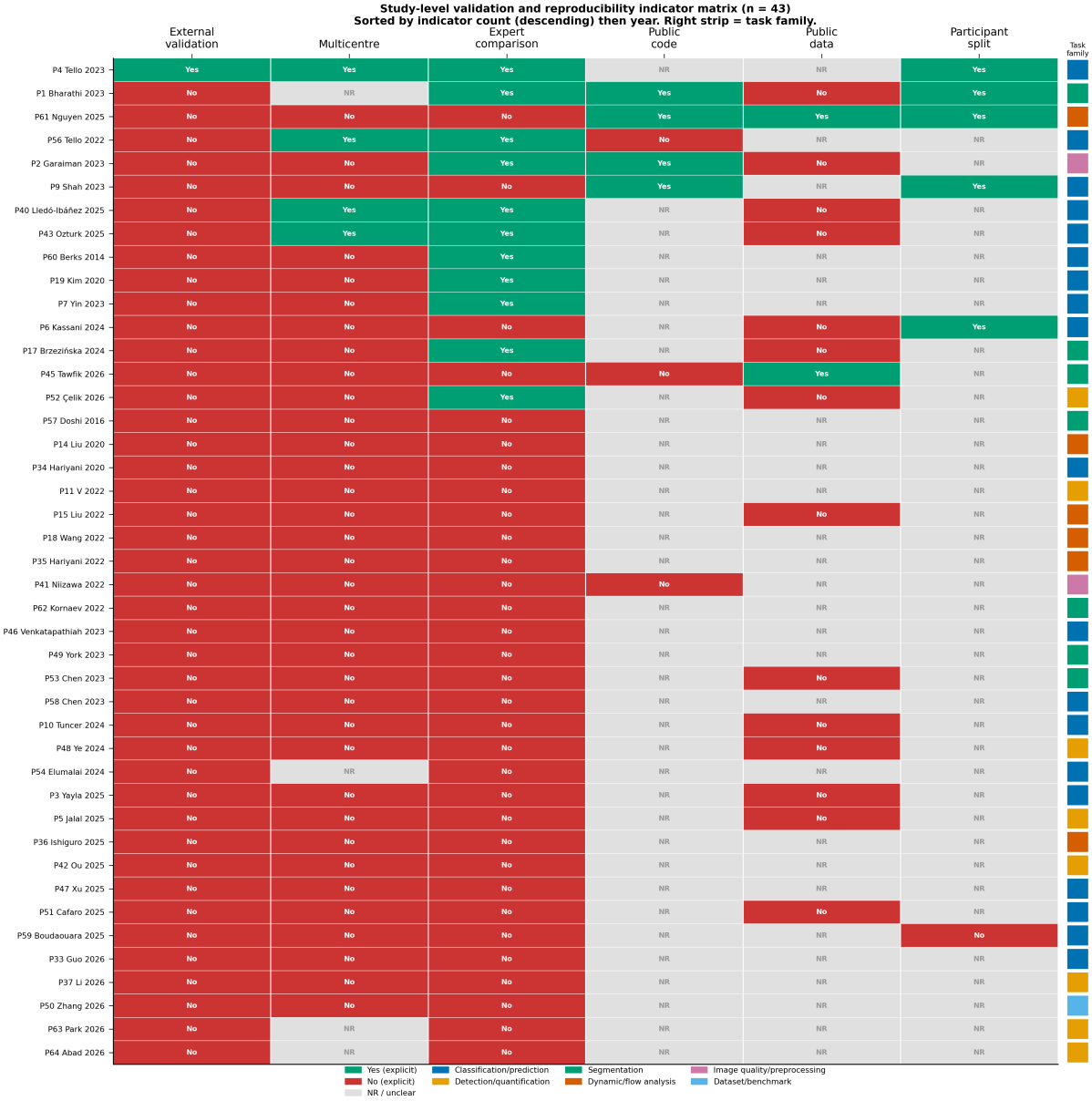
D.3 Study-level participant and image-unit denominators

Supplementary Table S8: Study-level participant and image/video/frame sample sizes among primary studies reporting both metrics ($n = 24$).

No.	Study	Year	Principal AI task family	Participants, n	Image units, n	Image units/participant
1	Lledó-Ibáñez	2025	Classification/prediction	1,780	48,199	27.1
2	Li	2026	Classification/prediction	396	600	1.5
3	Garaiman	2023	Classification/prediction	289	17,126	59.3
4	Jalal	2025	Classification/prediction	225	361	1.6
5	Kassani	2024	Classification/prediction	142	1,441	10.1
6	Shah	2023	Classification/prediction	120	5,326	44.4
7	Boudouaura	2025	Classification/prediction	103	309	3.0
8	Ozturk	2025	Classification/prediction	100	747	7.5
9	Yayla	2025	Classification/prediction	80	1,280	16.0
10	York	2023	Classification/prediction	32	1,386	43.3
11	Ishiguro	2025	Classification/prediction	6	80	13.3
12	Tuncer	2024	Detection/quantification	777	777	1.0
13	Yin	2023	Detection/quantification	30	1,788	59.6
14	Elumalai	2024	Segmentation	118	118	1.0
15	Liu	2020	Segmentation	60	50	0.8
16	Chen	2023a	Segmentation	22	613	27.9
17	Ye	2024	Segmentation	22	221	10.0
18	Hariyani	2020	Segmentation	7	40	5.7
19	Chen	2023b	Dynamic/flow analysis	22	539	24.5
20	Zhang	2026	Dynamic/flow analysis	22	57,373	2,607.9
21	Kim	2020	Dynamic/flow analysis	3	1,358	452.7
22	Hariyani	2022	Dynamic/flow analysis	2	490	245.0
23	Guo	2026	Image-quality/preprocessing	50	400	8.0
24	Nguyen	2025	Dataset/benchmark	202	6,695	33.1

Note. This table includes only the primary studies that reported both participant counts and image/video/frame counts. "Image units" refers to the image-based denominator reported by each study and may represent still images, video clips, extracted video frames, or annotated frames. The image-units-per-participant value was calculated as the reported number of image units divided by the reported number of participants and is intended only as a descriptive indicator of dataset density. Because image units were not harmonised across studies, this ratio should not be interpreted as a standardised sampling rate or as the number of statistically independent observations. Rows are grouped by principal AI task family and ordered by participant count within each task family. Duplicate author–year labels, such as Chen 2023a and Chen 2023b, should be aligned with the final bibliography. AI = artificial intelligence.

D.4 Study-level validation and reproducibility matrix



Supplementary Figure S2: **Study-level validation and reproducibility indicator matrix ($n = 43$).** Studies sorted by indicator count, descending, then year. The colour strip on the right edge indicates the principal task family. Yes = green, explicit; No = red, explicit; NR/unclear = grey. The matrix is descriptive and does not constitute a validated maturity or risk-of-bias score.

References

[1] Michael Berks, Phil Tresadern, Graham Dinsdale, Andrea Murray, Tonia Moore, Ariane Herrick, and Chris Taylor. An automated system for detecting and measuring nailfold capillaries. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2014), LNCS 8673*, 2014. doi:10.1007/978-3-319-10404-1_82.

[2] Niraj P. Doshi, Gerald Schaefer, and Iakov Korovin. MD-LBPV texture analysis for nailfold capillaroscopy image classification. In *Proceedings of the 2016 International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV 2016)*, 2016.

[3] S. Liu, Y. Li, J. Zhou, et al. Segmenting nailfold capillaries using an improved u-net network. *Microvascular Research*, 132:104011, 2020. doi:10.1016/j.mvr.2020.104011.

[4] B. Kim, Y.S. Hariyani, Y.H. Cho, and C. Park. Automated white blood cell counting in nailfold capillary using deep learning segmentation and video stabilization. *Sensors*, 20(24):7101, 2020. doi:10.3390/s20247101.

[5] Yuli Sun Hariyani, Heesang Eom, and Cheolsoo Park. DA-CapNet: Dual attention deep learning based on U-Net for nailfold capillary segmentation. *IEEE Access*, 2020. doi:10.1109/ACCESS.2020.2965651.

[6] K. Venkatapathiah, S.S. Selvan, P. Nanda, M. Shetty, V.M. Swamy, and K. Awasthi. Deep learning approach to nailfold capillaroscopy based diabetes mellitus detection. *International Journal of Online and Biomedical Engineering (iJOE)*, 18(6):95–109, 2022. doi:10.3991/ijoe.v18i06.27385.

[7] R. Liu, J. Tian, Y. Li, et al. Nailfold microhemorrhage segmentation with modified u-shape cnn. *Applied Sciences*, 12(10):5068, 2022. doi:10.3390/app12105068.

[8] Hung-Hsiang Wang and Chih-Ping Chen. Applying t-sne to estimate image sharpness of low-cost nailfold capillaroscopy. *Intelligent Automation & Soft Computing*, 32(1):237–254, 2022. doi:10.32604/iasc.2022.020665.

[9] Yuli Sun Hariyani and Cheolsoo Park. Event-based white blood cell classification from nailfold capillaries. *IEIE Transactions on Smart Processing and Computing*, 2022. doi:10.5573/IEIESPC.2022.11.4.270.

[10] Tomoya Niizawa, Kota Yokemura, Tomoya Kusaka, Takuma Sugashi, Ichiro Miura, Keiji Kawagoe, and Kazuto Masamoto. Automated capillary flow segmentation and mapping for nailfold video capillaroscopy. *Microcirculation*, 2022. doi:10.1111/micc.12753.

[11] Borja Gracia Tello, Eduardo Ramos Ibáñez, Patricia Fanlo Mateo, Luis Sáez Cómet, Elena Martínez Robles, Juan José Ríos Blanco, Begoña Marí Alfonso, Gerard Espinosa Garriga, José Todolí-Parra, Norberto Ortego-Centeno, José Luis Callejas Rubio, Mayka Freire Dapena, Adela Marín Ballvé, Albert Selva-O'Callaghan, Alfredo Guillén del Castillo, Carmen Pilar Simeón-Aznar, and Vicent Fonollosa Pla. The challenge of comprehensive nailfold videocapillaroscopy practice: a further contribution. *Clinical and Experimental Rheumatology*, 2022. doi:10.55563/clinexprheumatol/6usce8.

[12] Alexey V. Kornaev, Viktor V. Dremin, Elena P. Kornaeva, and Mikhail V. Volkov. Application of deep convolutional and long short-term memory neural networks to red blood cells motion detection and velocity approximation. In *Computational Biophysics and Nanobiophotonics*, volume 12194, page 121940C. International Society for Optics and Photonics, SPIE, 2022. doi:10.1117/12.2626040.

[13] Praveen Gurunath Bharathi, Michael Berks, Graham Dinsdale, Andrea Murray, Joanne Manning, Sarah Wilkinson, Maurizio Cutolo, Vanessa Smith, Ariane L. Herrick, and Chris J. Taylor. A deep learning system for quantitative assessment of microvascular abnormalities in nailfold capillary images. *Rheumatology (Oxford)*, 62(6):2325–2329, 2023. doi:10.1093/rheumatology/kead026.

[14] Alexandru Garaiman, Farhad Nooralahzadeh, Carina Mihai, Nicolas Perez Gonzalez, Nikitas Gkikopoulos, Mike Oliver Becker, Oliver Distler, Michael Krauthammer, and Britta Maurer. Vision transformer assisting rheumatologists in screening for capillaroscopy changes in systemic sclerosis: an artificial intelligence model. *Rheumatology (Oxford)*, 62(3):1140–1150, 2023. doi:10.1093/rheumatology/keac541.

[15] B.C. Gracia Tello, E. Ramos Ibáñez, L. Saez Comet, A. Guillén Del Castillo, C.P. Simeón Aznar, A. Selva O'Callaghan, G. Espinosa, G. Lledó, M. Freire Dapena, E. Martinez Robles, J.J. Ríos, J.A. Todolí Parra, B. Marí Alfonso, N. Ortego Centeno, A. Marín Ballvé, J.L. Callejas Rubio, V. Fonollosa Plá, and P. Fanlo. External clinical validation of automated software to identify structural abnormalities and microhaemorrhages in nailfold videocapillaroscopy images. *Clinical and Experimental Rheumatology*, 41(8):1605–1611, 2023. doi:10.55563/clinexprheumatol/m6obl3.

[16] Hao Yin, Zhiwei Wu, An Huang, Jiaxiong Luo, Junzhao Liang, Jianan Lin, Qianyao Ye, Mugui Xie, Cong Ye, Xiaosong Li, and Yanxiong Wu. Automated nailfold capillary density measurement method based on improved YOLOv5. *Microvascular Research*, 150:104593, 2023. doi:10.1016/j.mvr.2023.104593.

[17] Reema Shah, Jeremy Petch, Walter Nelson, Karsten Roth, Michael D. Noseworthy, Marzyeh Ghassemi, and Hertzel C. Gerstein. Nailfold capillaroscopy and deep learning in diabetes. *Journal of Diabetes*, 15(2):145–151, 2023. doi:10.1111/1753-0407.13354.

[18] Suma Kuncha Venkatapathiah, Sethu Selvi Selvan, Pranav Nanda, Manisha Shetty, Vikas Mallikarjuna Swamy, and Kushagra Awasthi. Deep learning based object detection in nailfold capillary images. *IAES International Journal of Artificial Intelligence*, 12(2):931–942, 2023. doi:10.11591/ijai.v12.i2.pp931-942.

[19] Daniel York, Gustave H. Falciglia, Elizabeth Managlia, Xiaocai Yan, Hawke Yoon, Aaron Hamvas, David Kirchenbuechler, Constadina Arvanitis, and Isabelle G. De Plaen. Nailfold capillaroscopy: A promising, noninvasive approach to predict retinopathy of prematurity. *The Journal of Pediatrics*, 2023. doi:10.1016/j.jpeds.2023.113478.

[20] Shupin Chen, Dan Wei, Shenming Gu, and Zhangru Yang. Blood flow characterization in nailfold capillary using optical flow-assisted two-stream network and spatial-temporal image. *Biomedical Physics & Engineering Express*, 2023. doi:10.1088/2057-1976/acdb7c.

[21] Shupin Chen, Dan Wei, Shenming Gu, and Zhangru Yang. Nailfold capillary segmentation based on U-Net and attention mechanism. In *2023 3rd International Conference on Bioinformatics and Intelligent Computing (BIC 2023), ACM*, 2023. doi:10.1145/3592686.3592747.

[22] Peyman Hosseinzadeh Kassani, Louis Ehwerhemuepha, Chloe Martin-King, Ryan Kassab, Ellie Gibbs, Gabrielle Morgan, and Lauren M. Pachman. Artificial intelligence for nailfold capillaroscopy analyses – a proof of concept application in juvenile dermatomyositis. *Pediatric Research*, 95:981–987, 2024. doi:10.1038/s41390-023-02894-7.

[23] S.A. Tuncer, M. Yildirim, T. Tuncer, and M.K. Mulayim. YOLOv8-based system for nail capillary detection on a single-board computer. *Diagnostics*, 14(17):1843, 2024. doi:10.3390/diagnostics14171843.

[24] Olga Elżbieta Brzezińska, Krzysztof Andrzej Rychlicki-Kicior, and Joanna Samanta Makowska. Automatic assessment of nailfold capillaroscopy software: a pilot study. *Reumatologia*, 62(5):346–350, 2024. doi:10.5114/reum/194040.

[25] Qianyao Ye, Hao Yin, Jianan Lin, Junzhao Liang, Mugui Xie, Cong Ye, Bin Zhou, An Huang, Zhiwei Wu, Xiaosong Li, and Yanxiong Wu. Improved nested U-structure for accurate nailfold capillary segmentation. *Microvascular Research*, 2024. doi:10.1016/j.mvr.2024.104680.

[26] Sowmiya Elumalai, Nirmala Krishnamoorthi, Naveen Periyasamy, Mohamed Farazullah, Kiran Raj, and Shriraam Mahadevan. Analysis of microvascular pattern in diabetes mellitus condition using the nailfold capillaroscopy images. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 2024. doi:10.1177/09544119231224510.

[27] M.E. Yayla, A. Aydin, M. Kılıcaslan, M. Kalkan, M.S. Güzel, A. Shikhaliyeva, S. Sezer, E. Uslu, A. Ateş, and T.M. Turgay. Deep learning performance in analyzing nailfold videocapillaroscopy images in systemic sclerosis. *Diagnostics*, 15(22):2912, 2025. doi:10.3390/diagnostics15222912.

[28] Mona Ebadi Jalal, Omar S. Emam, Cristian Castillo-Olea, Begonya García-Zapirain, and Adel Elmaghraby. Abnormality detection in nailfold capillary images using deep learning with efficientnet and cascade transfer learning. *Scientific Reports*, 15:3094, 2025. doi:10.1038/s41598-025-85277-8.

[29] Rio Ishiguro, Sae Kawasaki, Ryosuke Imai, Kota Toyama, Masato Takahashi, and Norimichi Tsumura. Shape classification of nailfold capillaries using convolutional neural network. *Artificial Life and Robotics*, 2025. doi:10.1007/s10015-025-01052-y.

[30] Gema M Lledó-Ibáñez, Luis Sáez Comet, Mayka Freire Dapena, Miguel Mesa Navas, Miguel Martín Cascón, Alfredo Guillén del Castillo, Carmen Pilar Simeon, Elena Martinez Robles, José Todolí Parra, Diana Cristina Varela, Génesis Maldonado, Adela Marín, Laura Pérez Abad, Jimena Aramburu, Laura Vela, Eduardo Ramos Ibáñez, and Borja del Carmelo Gracia Tello. Capidetect: machine learning in capillaroscopy reveals new variables influencing diagnosis. *Rheumatology*, 64(6):3667–3675, 06 2025. doi:10.1093/rheumatology/keaf073.

[31] Chiao-Chi Ou, Yun-Chi Liu, Kuo-Ping Lin, Tsai-Hung Yen, and Wen-Nan Huang. Classification method for nailfold capillary images using an optimized Sugeno fuzzy ensemble of convolutional neural networks. *Computers in Biology and Medicine*, 2025. doi:10.1016/j.compbiomed.2025.109975.

[32] L. Ozturk, C. Laclau, C. Boulon, M. Mangin, E. Braz-Ma, J. Constans, L. Dari, and C. Le Hello. Analysis of nailfold capillaroscopy images with artificial intelligence: data from literature and performance of machine learning and deep learning from images acquired in the SCLEROCAP study. *Microvascular Research*, 157:104753, 2025. doi:10.1016/j.mvr.2024.104753.

[33] Hui Xu and Jie Qian. The role of nailfold video-capillaroscopy in the assessment of dermatomyositis. *Rheumatology*, 2025. doi:10.1093/rheumatology/keae677.

[34] Giacomo Cafaro, Roberto Bursi, Valentina Valentini, Katharina Hansel, Carlo Perricone, Vincenzo Venerito, Onelia Bistoni, Manuela Sebastiano, Fabiana Topini, Luca Stingeni, Roberto Gerli, and Elena Bartoloni. Combined semiquantitative nail-enthesis complex ultrasonography and capillaroscopy in psoriasis and psoriatic arthritis. *Frontiers in Immunology*, 2025. doi:10.3389/fimmu.2024.1505322.

[35] Dorra Boudaouara, Randa Boukhris, Alima Damak, and Chifa Damak. Automated classification of nailfold capillaroscopy images using morphological feature extraction. In *2025 IEEE 22nd Inter-*

national Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), Hammamet, Tunisia, 2025. doi:10.1109/STA66620.2025.11364855.

[36] Hang Thi Phuong Nguyen and Hieyong Jeong. Capillary dataset: A dataset of nail-fold capillaries captured by microscopy for diabetes detection. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (KDD 2025)*, 2025. doi:10.1145/3711896.3737425.

[37] Peiqing Guo, Bin Zhou, Shou Feng, Qirui Sun, Jiaxiong Luo, Jianan Lin, Yaozhang Huang, Lei Li, Yanxiong Wu, and Chao Wang. Wide-field nailfold capillary image deblurring method based on an improved MIMO-UNet. *Microvascular Research*, 2026. doi:10.1016/j.mvr.2026.104916.

[38] Jie Li, Congcong Jian, Jiaojiao Zhao, Zhuming Yin, Xue Li, Chaochun Song, Hongjun Zhou, Wenzhang Chen, Qin Liu, Jie Zhang, Zhengzhong Tang, Xiangyi Xiong, Yuqi Liu, Jun Zhou, Xue Yang, Jianhong Wu, Tingting Wang, Qinghua Zou, Yong Zhang, Hua'e Shu, and Fanxin Zeng. Machine learning-based multiclass model for autoimmune disease diagnosis and classification through nailfold videocapillaroscopy features. *RMD Open*, 2026. doi:10.1136/rmdopen-2025-006393.

[39] Mohamed Tawfik and Wael Badawy. A comparative evaluation of deep learning-based capillaroscopy image analysis using YOLOv8–YOLOv12 models for microvascular diagnostics. *Neural Computing and Applications*, 2026. doi:10.1007/s00521-026-11847-0.

[40] Fan Zhang, Bingchen Yu, Jianwei Zuo, Rui Xu, Kai Pang, Wei Jin, and Jiajia Luo. Noninvasive real-time dynamic monitoring of white blood cells based on microscopic imaging and deep learning. *Computer Methods and Programs in Biomedicine*, 2026. doi:10.1016/j.cmpb.2026.109238.

[41] Neşe Çabuk Çelik and Elif Altunel Kılınç. Comparison of artificial intelligence and rheumatologists in nailfold capillaroscopic evaluation. *Revista da Associação Médica Brasileira*, 2026. doi:10.1590/1806-9282.20251389.

[42] Youngjin Park, Sehyun Kye, and Ojin Lee. Detection of nailfold capillaries using smartphone microscopy and deep learning analysis. *Transactions of the Korean Institute of Electrical Engineers*, 2026. doi:10.5370/KIEE.2026.75.1.166.

[43] L. Pérez Abad, M.Á. Aibar Arregui, J. Aramburu Llorente, J. Ramón y Cajal Calvo, A. Andrés Gracia, P. Revilla Martí, S. Atienza Ayala, C. Lahuerta Pueyo, A. Campos Sáenz de Santamaría, E. Ramos Ibañez, and B.D.C. Gracia Tello. Unlocking the potential of nailfold videocapillaroscopy in diagnosing and staging wild-type transthyretin amyloidosis: A preliminary approach. *Medicina Clínica*, 2026. doi:10.1016/j.medcli.2025.107283.