

1 **Title: Artificial Intelligence Algorithms in Nailfold**
2 **Capillaroscopy Image Analysis: A Systematic Review**

3
4 **Author list:** Omar S. **Emam**^{1*}; Mona **Ebadi Jalal**^{1,2}, Begonya **Garcia-Zapirain**³, Adel S.
5 **Elmaghraby**^{1,2}

- 6 1. Hive AI Innovation Studio, University of Louisville, KY, US
7 2. Department of Computer Science and Engineering, University of Louisville, KY, US
8 3. eVida Lab, Faculty of Engineering, University of Deusto, Bilbao, Spain

9
10 ***Corresponding Author:** Omar S. Emam

11 Hive AI Innovation Studio, University of Louisville, Louisville, Kentucky, US

12 Email: osemam01@louisville.edu | Alternative Email: omarsamehemam@gmail.com

13
14 **Author Contributions:** OSE and MEJ performed systematic search, study selection, data
15 curation and extraction. BGZ (Co-PI) and ASE (PI) supervised and guided the flow process,
16 ensured validation and quality of the manuscript, and were resorted to in cases of inter-author
17 disagreement. All four authors were involved in conceptualization, writing, review and editing.

18
19 **Running Title:** A Systematic review on AI in nailfold capillaroscopy

20 **Keywords:** Artificial Intelligence, Machine Learning, Deep Learning, Nailfold
21 Capillaroscopy, Video-capillaroscopy, Microscopic Angioscopy.

22 **Financial Disclosure Statement:** None of the authors have any conflicts of interest.

23 **ABSTRACT**

24 **Background** Non-invasive imaging modalities offer a great deal of clinically significant
25 information that aid in the diagnosis of various medical conditions. Coupled with the never-
26 before-seen capabilities of Artificial Intelligence (AI), uncharted territories that offer novel
27 innovative diagnostics are reached. This systematic review compiled all studies that utilized AI
28 in Nailfold Capillaroscopy as a future diagnostic tool.

29 **Methods and Findings** Five databases for medical publications were searched using the
30 keywords artificial intelligence, machine learning, deep learning and nailfold capillaroscopy to
31 return 105 studies. After applying the eligibility criteria, 10 studies were selected for the final
32 analysis. Data was extracted into tables that addressed population characteristics, AI model
33 development and nature and results of their respective performance. We found supervised deep
34 learning approaches to be the most commonly used ($n = 8$). Systemic Sclerosis was the most
35 commonly studied disease ($n = 6$). Sample size ranged from 17,126 images obtained from 289
36 participants to 50 images from 50 participants. Ground truth was determined either by experts
37 labelling ($n = 6$) or known clinical status ($n = 4$). Significant variation was noticed in model
38 training, testing and feature extraction, and therefore the reporting of model performance. Recall,
39 precision and Area Under the Curve were the most used metrics to report model performance.
40 Execution times ranged from 0.064 to 120 seconds per image. Only two models offered future
41 predictions besides the diagnostic output.

42 **Conclusions** AI has demonstrated a truly remarkable potential in the interpretation of Nailfold
43 Capillaroscopy by providing physicians with an intelligent decision-supportive tool for improved

44 diagnostics and prediction. With more validation studies, this potential can be translated to daily
45 clinical practice.

46 **1 INTRODUCTION**

47 Capillaries are the tiniest and most numerous blood vessels that link the body's
48 arterial system to its venous system. They branch superficially and deeply into all body tissues to
49 provide nutrients and remove waste products (1, 2). This healthy microenvironment can get
50 severely dysfunctional due to the pathological processes in numerous abnormal conditions (3);
51 for example: a) systemic diseases such as diabetes (4, 5), or metabolic syndrome, b) auto-
52 immune inflammatory pathologies such as vasculitis and dermatomyositis (6), and c) connective
53 tissue diseases (7) such as systemic sclerosis (SSc) (8, 9) systemic lupus erythematosus (10, 11),
54 and Raynaud's phenomenon (12, 13). Capillaries in the retina, tongue, or nailfolds provoke a
55 particular medical interest due to their ease of accessibility and examination using common non-
56 invasive tools that yield clinically significant information. In other words, the diagnosis and
57 follow up of internal systemic conditions can be performed without the need to resort to invasive
58 approaches.

59 Nailfold Capillaroscopy (NFC), which is a type of microscopic angioscopy, is a
60 technique of visualizing the capillaries in the nailfold area (14). It examines the ultimate
61 capillary endings in the finger as capillaries loop to turn back around. To visualize these terminal
62 capillaries in that thin layered area of the skin, the subject as well as the surrounding
63 environment are prepared and then a microscopic lens can be used to directly observe the
64 capillaries. Nailfold Video Capillaroscopy (NVC) (15, 16) uses a more advanced scope with a
65 camera that offers far superior resolution, clarity, and the ability to record a video and take
66 photos (6). These images are then analyzed by trained experts to differentiate normal healthy

67 capillaries from faulty pathological ones according to certain specific criteria. This information
68 can aid in diagnosis, progression and severity assessment (17), disease staging (18-22), follow-
69 up and perhaps prediction of certain medical conditions. Unfortunately, such convenient classical
70 methods of manual analyses imply subjectivity, prolonged analyses time, and ambiguity of
71 findings (23-25).

72 The recent trends utilizing Artificial Intelligence (AI), especially in medicine and
73 biomedical sciences, seem to offer a highly sought-after outturn as a superior alternative (26, 27).
74 To illustrate, many deep learning models have demonstrated the ability to objectively analyze
75 images from NFC/NVC with higher, or at least comparable, efficiency as capillaroscopy experts
76 in a significantly lower time (5, 15, 28-41). Advances in computer vision algorithms allow the
77 extraction, quantification, and accurate analysis of far more features compared to human experts.
78 These innovations present an unprecedented potential to link a multitude of variables to diseases
79 and, consequently, draw future predictions (7). The information gained from such technology
80 would be crucial to inform decisions concerning patient education and treatment in many ways.
81 For example, predictive models can be used as an adjunct screening tool, diagnostic tools can
82 help accelerate clinical work, establish risk stratification, and reduce rates of misdiagnoses (42).
83 Moreover, the smart ‘learning’ nature of these tools will benefit from feedback and provide
84 explanations to further improve their accuracy (43). As such, these innovations claim the
85 potential to transform existing medical practices by enhancing the decision-making process using
86 such tools. In other words, is it practically feasible to improve the quality of care delivered to
87 patients using such cheap, easy, non-time-consuming means?

88 In this systematic review, we aim to give an overview on the state-of-the-art by
89 compiling all studies, to date, that utilized an AI algorithm to analyze output from NFC/NVC as

90 a tool to be used clinicians in medical practice. We describe the methodology used to conduct a
91 comprehensive search of the literature to encompass all novel studies addressing this topic. Next,
92 we presented the summarized results of the included reports. Finally, we divided the discussion
93 into four sections to organize the gathered evidence to answer our research questions with
94 commentary on a few important issues as follows: Section I: a brief background on the
95 significance of NFC in medicine, Section II: Challenges facing the manual method, Section III:
96 How AI is solving these challenges, and finally Section IV: Future directions and limitations.

97

98 **2 METHODS**

99 This systematic review was conducted in compliance with the Preferred Reporting Items
100 for Systematic Reviews and Meta-Analysis (PRISMA) (44). We formulated the research
101 questions, and explained their significance, in **Table 1**, to guide the following search and
102 screening steps.

103 **Table 1.** Research questions and their rationale

	Research Question	Rationale
Q1	Can AI successfully detect clinically significant changes in NFC images?	To ensure that the information captured is relevant and useful to clinicians
Q2	Is pre-analysis processing required for an optimized output?	To check the user-friendliness and practically consider the time elapsed
Q3	Is there a minimum number of variables/features to be extracted for an accurate enough output?	To understand how these algorithms work and study the “significant” variables/features
Q4	Is information from NFC/NVC alone enough? Or is combining them with other diagnostic modalities and data sources needed?	To maximize the potential for detection, data acquisition and interpretation
Q5	Would the “early” use of such technologies help in predicting future outcomes?	To consider the possibility of early interventions and thus better patient reported outcomes
Q6	What conditions/diseases can this technology be applied to?	To understand the scope and generalizability or individuality of these AI applications

104

105 **2.1 Data Collection**

106 **2.1.1 Search Strategy**

107 We performed an all-time search on December 14th, 2023, that was later updated in
108 March 2024, utilizing five electronic medical databases: PubMed (including MEDLINE),
109 EMBASE, CINAHL, and Web of Science using the keywords: "Nailfold", "Capillaroscopy",
110 "Artificial Intelligence", "Machine Learning" and "Deep Learning" to generate the search
111 string: (nail OR nailfold OR "nail-fold") AND (capillaroscopy OR "video capillaroscopy" OR
112 NVC OR "microscopic angioscopy" OR onychoscopy) AND ("artificial intelligence" OR AI OR
113 "machine learning" OR "deep learning" OR algorithm*). MeSH terms and boolean operators
114 were used where appropriate. **Fig.1** shows a PRISMA flow diagram that summarizes the
115 searching process.

116

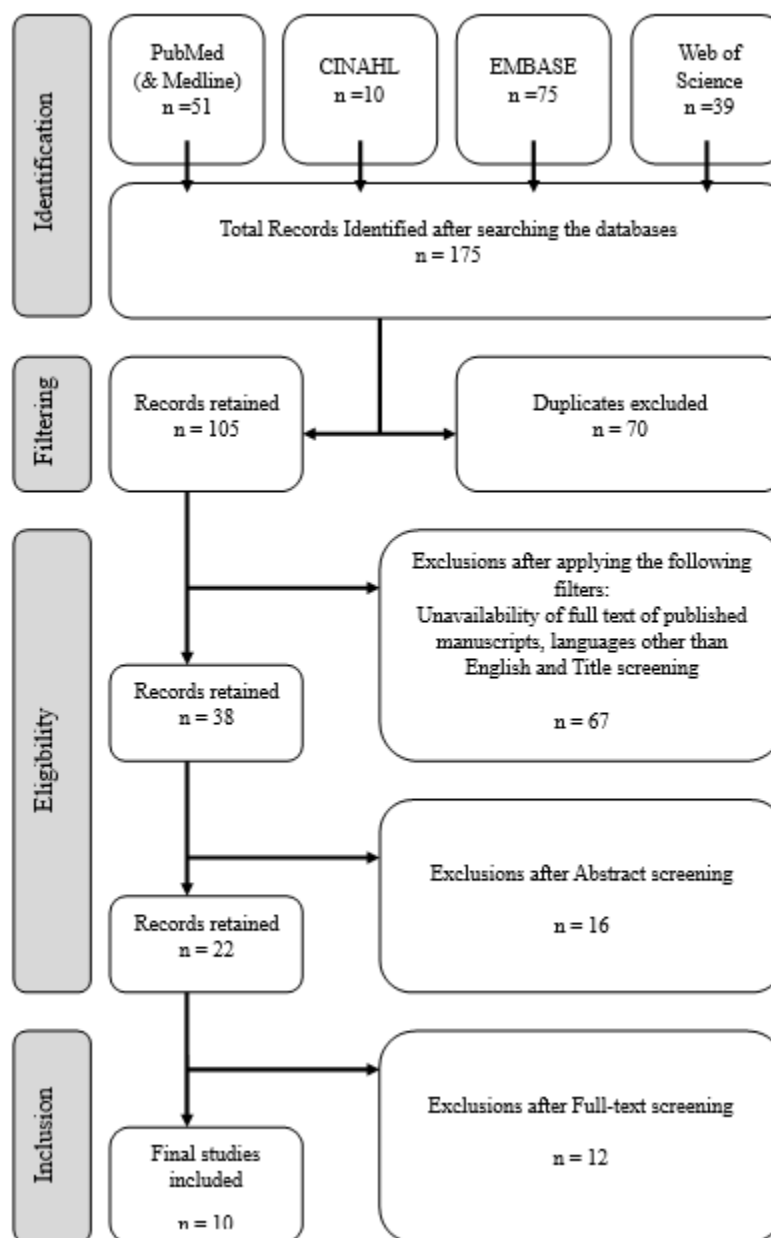
117 **2.1.2 Inclusion and Exclusion Criteria**

118 We limited the search results to availability of full texts in the English language through
119 our institutional access. We included all 1) peer reviewed journal articles, 2) both qualitative and
120 quantitative studies, 3) that presented an AI model, 4) to analyze NFC/NVC images, and 5) was
121 internally validated or tested on a real dataset.

122 We excluded unpublished data under review, conference abstracts and proceedings, and
123 grey literature such as short surveys, and letters to editors. Additionally, we excluded studies that
124 1) utilized a purely mathematical model/algorithm that is not categorized as AI, or 2) presented
125 technical aspects of the technology without applying it to a real data set as a diagnostic or
126 prognostic tool, or 3) included other techniques to study the nailfold capillaries other than NFC
127 such as optoacoustic imaging or ultrasound, whether independently or in fusion with NFC.

128

Fig.1 PRISMA Flow diagram summarizing the search and screening process



129

130 **2.1.3 Study Selection and Data Extraction Process**

131 Two authors independently performed the search and imported the results to the EndNote
132 v.20 reference manager, where duplicates were removed. Any conflicting judgements were
133 resolved by a third author. We then extracted data related to the study design, target population

134 characteristics and sampling, pre-procedural setting description, pre-analysis processing, AI
135 methods, model training and development, type of input and output, performance metrics,
136 limitations, strengths, and finally conclusions; that process was repeated for all final ten studies
137 included in the review.

138

139 **2.2 Quality Assessment and Risk of Bias**

140 Assessment of quality and the risk of bias in diagnostic accuracy studies is commonly
141 done using known validated tools such as STARD (45) and QUADAS (46), or tools like
142 TRIPOD (47) for prognostic studies as well. However, the novel nature of AI-related studies
143 created a demand for more relevant and appropriately updated tools that many authors sought to
144 meet by modifying or adding extensions to the aforementioned tools such as QUADAS-AI (48),
145 STARD-AI (49), and TRIPOD-AI (50). Unfortunately, these tools are still under development
146 analysis and there is currently no agreed upon gold standard tool to be used (51). For these
147 reasons we opted to refer to the updated QUADAS-2 tool (52) and complement the assessment
148 with the MINIMAR (53) and CAIR (54) tools that present a checklist for reporting AI studies to
149 healthcare providers, as outlined in supplementary document **S1**.

150 3 RESULTS

151 The search process returned 105 results that were then screened by title, abstract, and
152 finally after full-text reading using the eligibility criteria to arrive at the final list of ten studies as
153 summarized in **Fig. 1** (15, 28-32, 35, 40, 41, 55). Six out of the ten studies were published within
154 a year of writing this manuscript (in 2023) illustrating the novelty of the topic at hand (28-31, 41,
155 55). The geographical distribution of countries where these studies were conducted is shown in
156 **Fig. 1**. The following

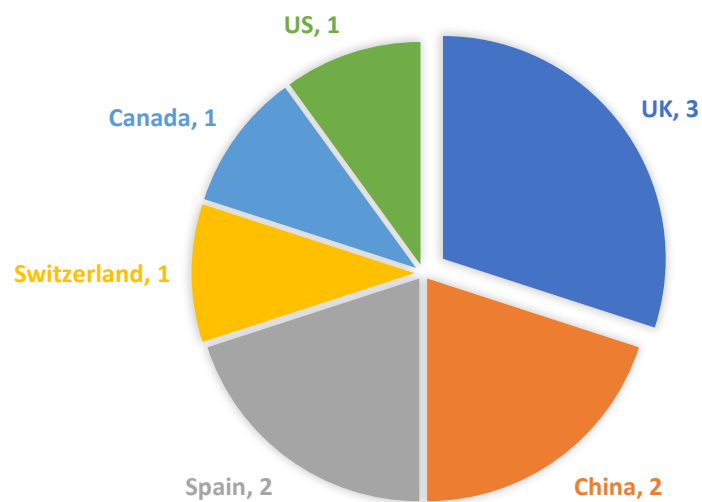


Fig. 1 Geographical distribution of the countries of included studies

157 **Table 2** shows a summary of the main highlights of all ten studies including population
158 characteristics.

159

Table 2. A summary of the main highlights and population characteristics of all ten studies.

Author and Year	Country	Target Population Characteristics					AI Model	
		Diseases addressed	Disease Duration	Database	Sample size	No. of Participants	Nature of Task	Detection Time
Yin et al 2023 (28)	China	HC only	—	N/A	1,788 images	30	DL -Capillary Segmentation	0.064 sec /image
Shah et al 2023 (55)	Canada	DM, complications* & HC	≥ 10 years	Boris Clinic at Hamilton Health Sciences	5,236 images	120 = 60 HC + 60 DM	DL - Classification	—
Garaiman et al 2023 (30)	Switzerland	SSc of varying severities	—	Local registries EUSTAR & VEDOSS	17,126 images	289	DL - Classification	0.19 secs /image
Bharati et al 2023 (31)	UK	Early SSC, PRD & HC	—	Salford Royal NHS Foundation Trust, a 3ry referral center	2,541 mosaics = 2,117 Hi-res (A & B) + 424 low-res (C)	309 = 155 SSC + 154 HC/PRD	DL - Mixed	—
Kassani et al 2023 (41)	US	aJDM & HC	3-33 years	JDM clinic of Lurie Children's Hospital of Chicago	Diagnostic: 1,441 images & Predictive: 1,760 images	142 = 111 aJDM + 31 HC	DL - Classification	—
Tello et al 2023 (29)	Spain	PRD, & SRP	—	5 Spanish hospitals	1,164 images	—	DL - Mixed	—
Tello et al 2022 (15)	Spain	PRD, & SRP	—	Built from 9 Spanish 3 ^{ry} hospitals (members of GEAS, SEMI & SEMIAS)	2,713 images with >18k measurements	—	DL - Mixed	—
Liu et al 2020 (32)	China	Htn & HC	—	N/A	50 **	60	DL - pixel Segmentation	—
Berks et al 2014 (35)	UK	Severe SSC, PRD & HC	—	3 ^{ry} referral center for SSC Patients	990 mosaics	—	ML - Mixed	—
Murray et al 2011 (40)	UK	SSC (LC, DC), PRD & HC	—	N/A	116 mosaics	116 = 46 HC + 21 PRD + 49 SSC	ML - Classification	2 min /image

160 ML = Machine Learning, DL = Deep learning, N/A = No available data or not reported by author, HC = Healthy
 161 Controls, DM = Diabetes Mellitus, PRD = Primary Raynaud's Disease, SRP = Secondary Raynaud's Phenomenon,
 162 SSc = Systemic Sclerosis, LC = Limited Cutaneous, DC = Diffuse Cutaneous, aJDM = Active Juvenile
 163 Dermatomyositis, Htn = Hypertension *Complications include: Cardiovascular event, retinopathy, albuminuria and
 164 hypertension, ** used data augmentation to generate 20,664 samples from the original 50.

165 Synthesis of Evidence

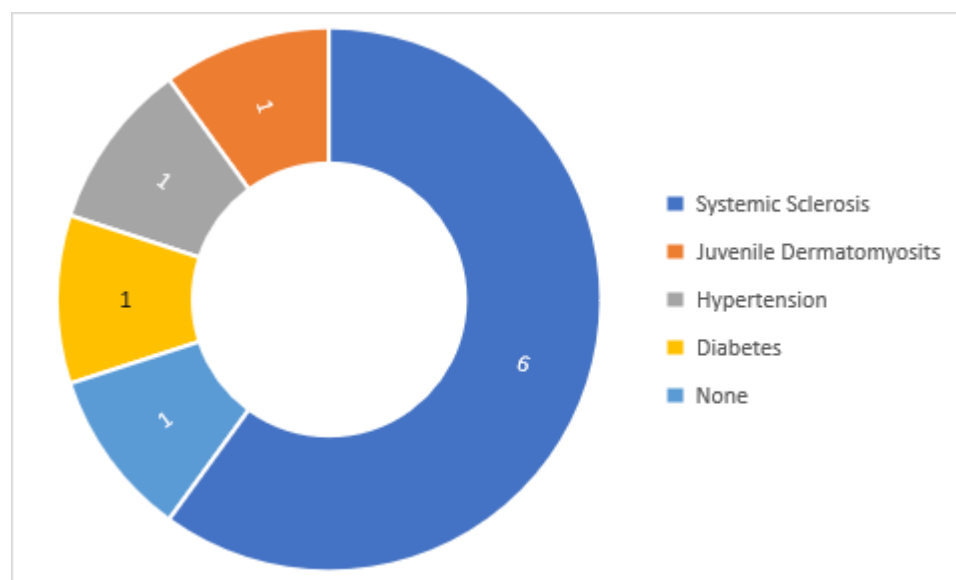
166 3.1 Study Designs

167 Consistent with most diagnostic accuracy studies, all ten investigators adopted a
168 non-experimental cross-sectional design with case-control selection (15, 28-32, 35, 40, 41, 55).
169 After recruiting participants via certain criteria, the developed AI algorithm was tested against
170 the ground truth to assess the model's performance.

171

172 3.2 Population Characteristics

173 Yin et al was the only author to include healthy volunteers as the sole subject of his
174 study (28). Alternatively, Garaiman et al was the only investigator to acknowledge limiting his
175 study to include only diseased patients due to ethical consent considerations (30). The remaining
176 eight studies included both diseased participants and normal controls (15, 29, 31, 32, 35, 40, 41,
177 55). Without including both normal and diseased subjects, an AI algorithm would not be capable
178 of predicting abnormalities. The diseases addressed by each study are shown in **Fig. 2**.



179

180

Fig. 2 Diseases addressed by the included studies

181 Among the six studies investigating systemic sclerosis (SSc), three authors
182 considered two variants as the normal controls: completely disease-free healthy participants and
183 others with a common benign condition known as primary Raynaud’s phenomenon (31, 35, 40);
184 also known as Raynaud’s disease (PRD) (56). Conversely, Garaiman et al and Tello et al were
185 satisfied with PRD only as normal controls (15, 29, 30). Interestingly, despite the recognized
186 significance of disease duration due to its pathological impact on the body, and, consequently,
187 the outcomes assessed by NFC, only three authors (40, 41, 55) reported the disease duration of
188 their respective participants.

189 Participants’ demographics were surprisingly not addressed in 50% of the studies
190 (15, 28, 29, 32, 35). Moreover, only Kassani et al and Shah et al reported racial backgrounds, and
191 thus, shedding light on a very significant - yet overlooked – factor: skin tone (41, 55). This
192 finding highlights a significant gap in addressing imbalances due to a particularly relevant factor
193 like skin-color, in addition to generally important variables such as sex, age, and other co-
194 morbidities, that may influence the algorithm’s learning and output. Among all studies that
195 investigated adults, Kassani et al was also unique in investigating a condition prevalent in the
196 children’s age group, juvenile dermatomyositis (JDM) (41).

197 **3.3 Sample Sizes**

198 The highest sample size was 17,126 images obtained from 289 participants, as
199 reported by Garaiman et al (30); which was more than three times the second highest sample size
200 of 5,236 from 120 participants reported by Shah et al (55). Liu et al developed his algorithm with
201 the least sample size of 50 images (32). **Fig. 3** Error! Reference source not found. shows a c
202 omparison of the total sample sizes of both images and participants across all included studies.

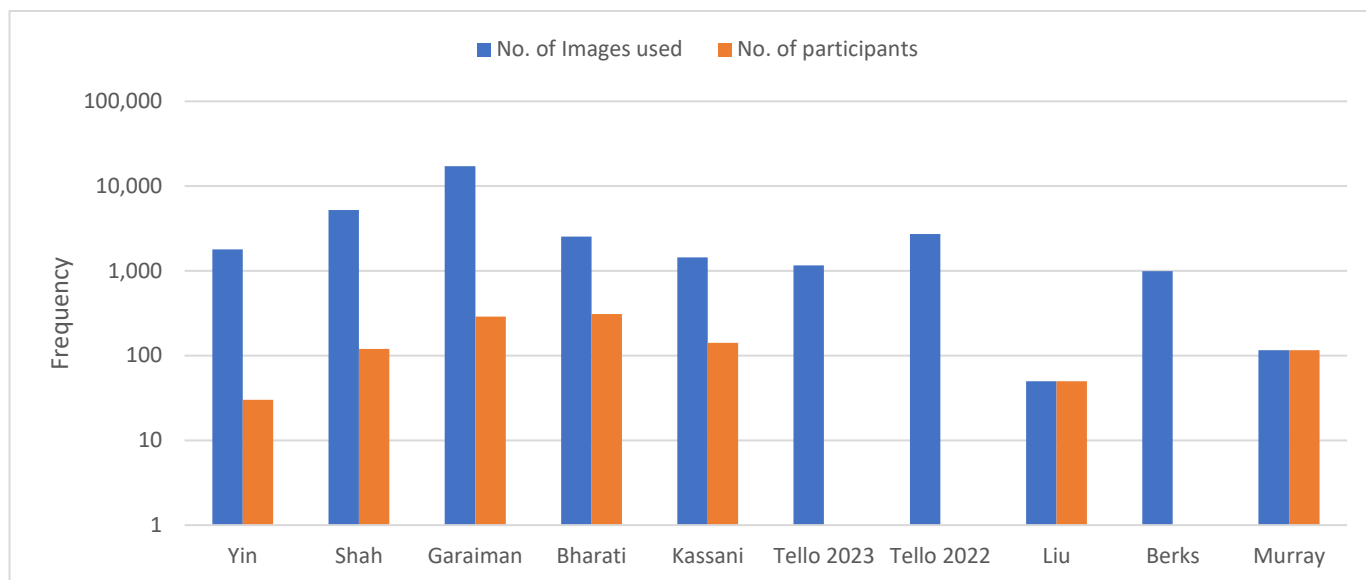


Fig. 3 Comparison of the total sample sizes across all included studies shown in logarithmic scale

203
204 Authors differed in their approach to obtaining sample images from participants, and
205 in how balanced the normal controls and main study groups were. For instance, five authors
206 obtained multiple images from multiple sections of multiple nailfolds spanning multiple
207 encounters (28, 30, 31, 41, 55). Conversely, both Liu et al and Murray et al obtained a single
208 image per participant (32, 40). However, Murray's image is a mosaic, which is a panoramic
209 wholesome image of the entire nailfold, compared to Liu's single image that represents a section
210 of the nailfold. Nonetheless, Liu et al used a data augmentation technique to generate a much
211 bigger sample size, derived from the original 30, that reached 20,664 images (32). Alternatively,
212 Tello et al chose images at random from a bigger pool of a previously prepared dataset without
213 reporting the number of participants (15, 29). Similarly, Berks et al used 990 mosaics without
214 reporting the number of fingers, participants, or encounters that contributed to that dataset (35).
215 Finally, Kassani et al had a sample of 1,441 images for their diagnosis model and a different
216 sample of 1,760 images for the predictive model (41).

217 **3.4 AI Algorithms Development**

218 **3.4.1 Model Architecture**

219 The majority of authors employed a deep learning model to accomplish their task
220 (15, 28-32, 41, 55), except for Berks et al and Murray et al who utilized a pipeline of machine
221 learning models (35, 40). All ten authors utilized a supervised machine learning approach where
222 the input images were labelled with the ground truth so that the model learns through the input-
223 output pairs.

224 **Table 3** summarizes the different architectures for each algorithm, learning approaches,
225 diagnostic models' development and performance.

226 The nature of tasks performed by the algorithms varied as well. The first approach
227 employed by both Yin et al and Liu et al was a 'segmentation-based' object detection for
228 capillaries (28, 32). Liu et al relied on a pixel-wise semantic segmentation to decide whether it
229 belongs to a capillary or not (32), while Yin et al took a more global prospective to detect a
230 capillary in the image rather than detecting it at the pixel-level (28). Four authors adopted the
231 second approach of 'classification' that aimed to classify the input image into one category out
232 of two or more categories/classes; either at the pixel, capillary, image, or global levels (30, 40,
233 41, 55). The remaining three authors took advantage of a multi-step mixed approach that relied
234 on detecting certain features first, then classifying them at one or more levels into the assigned
235 classes (15, 29, 31, 35).

Table 3 Comparing AI models across all ten studies

Author and Year	Pre-analysis processing	Features / variables used	Ground Truth	Model Development			Output		Results
				Training	Validation	Testing	Diagnostic Model	Predictive as well	
						Image level	Global level		
Yin et al 2023 (28)	N/A	Capillary detection and density calculation	Calculation by 3 experts	200 High Quality images	The remaining split Tr:V:T as 6:2:2	<i>Quantitative</i> (capillary count and density)	N/A	No	Average Precision 85.2%, AUPRC 85.2%
Shah et al 2023 (55)	N/A	Images as a whole + DM status	Clinical Status through chart review using specific defining criteria for each diagnosis	5-fold cross validation applied to each of the 5 diagnoses	<i>Not done.</i> Relied on mean of all 5 testing sets	Qualitative	Yes, Qualitative	Yes. To predict a history of CV event in DM patients	For DM: 0.82 Recall, 0.84 AUROC. 0.84 AUPRC
Garaïman et al 2023 (30)	N/A	Enlarged, Giant capillaries, Capillary loss, and Micro-hemorrhages	Majority vote-derived labels ≥ 3 out of 4 practicing rheumatologists	5-fold cross validation. The original dataset was split randomly into 5 equal groups. Tr:V:T was 3:1:1, repeated 5 times until every group had the chance to be in each set		Qualitative (SSC pattern or not)	N/A	No	AUROC 85.07 – 92.6% and Recall of 81.91% - 92.61%
Bharati et al 2023 (31)	Yes, rescaling to a resolution of 2.5 μ m	Apex detection, Mean shape score, Apical width, and Density	Clinical Status as SSC or Normal	Half of Group A = 456 mosaics of high-resolution	The remaining 455 from group A	<i>Quantitative</i> (count, density..etc.) AND Qualitative (SSC pattern or Normal)	Yes, using average measurements from all fingers	No	AUROC 97% (B), 95% (C). Recall of 91.5 (B), 78% (C)
Kassani et al 2023 (41)	Yes, size reduction	Images as a whole with ~28,000 trainable parameters	Known clinical status of the patient in the registry.	Data was divided into training and testing sets based on a 5-fold cross-validation	70 HC and 217 aJDM images	Qualitative (presence of JDM or not)	Yes, if \geq aggregate predictions from all pictures then it counts	Yes. Scores to JDM activity vs no activity	Recall 0.85, Precision 0.95, AUROC 0.93

Author and Year	Pre-analysis processing	Features / variables used	Ground Truth	Model Development			Output			Results
				Training	Validation	Testing	Diagnostic Model		Predictive as well	
							Image level	Global level		
Tello et al 2023 (29)	N/A	Abnormal shape, giant, dilation tortuosity, Microhemorrhages	A consensus of ≥ 3 and ≥ 4 out of 5 NFC experts	85% of the dataset		15% of the dataset	<i>Quantitative</i> AND Qualitative	N/A	No	>89% Specificity and NPV for all categories
Tello et al 2022 (15)	Ensured a magnification of 200x	Detecting capillaries by their shape, and size	Validated manual annotation by experts	85% of the dataset (2,306 images with 12,352 manually annotated and validated capillaries with multiple measurements)		15% of the dataset (407 images with 1,690 capillary)	<i>Quantitative</i> AND Qualitative	N/A	No	Mean Recall 85%, mean Precision of 72%
Liu et al 2020 (32)	Image enhancing, cropping, resolution adjustment, manual RoI selection	Detection of pixels that entail a capillary	Labels marked by a professional	On 30 NFC images *	On 20 NFC images		Qualitative (whether each pixel is a capillary or not)	N/A	No	Mean Accuracy 91.72%, mean dice score 97.66%
Berks et al 2014 (35)	Feature engineering to fuse images into a single mosaic	Capillary detection, apical width, orientation, tortuosity	Labels marked by a professional	80 images with 450 training RoI	456 images (104 HC, 83PRD and 268 SSc)	455 images (104 HC, 83 PRD and 268 SSc)	<i>Quantitative</i> (capillary count and vessel morphology measurements)	N/A	No	80.9 Recall, 64.1 Precision, F-measure of 71.5
Murray et al 2011 (40)	Combining images into mosaics, inversion, rotation, RoI selection...etc	Inter-capillary distance, width, tortuosity, derangement	Clinical diagnosis	Randomly split the dataset into training and testing. Used a linear set of SVM classifiers for training . Repeated that process > 60 trials			<i>Quantitative</i> AND Qualitative	N/A	No	Pearsons Correlation 75% correct match.

238 N/A = No available data or not reported by author, Tr:V:T = Training:Validation:Testing, HC = Healthy
239 Controls, aJDM = active Juvenile Dermatomyositis, * Data augmentation was used to generate 20,664
240 images, RoI = Region of Interest, PRD = Primary Raynaud’s Disease, SSc = Systemic Sclerosis, DM =
241 Diabetes Mellitus, CV = Cardiovascular, NPV = Negative Predictive, Value.

242

243 **3.4.2 Model Input**

244 Pre-analysis processing to optimize the model performance was carried out either
245 manually, automatically, or semi-automatically and reported in varying detail in six studies (15,
246 31, 32, 35, 40, 41). It included simple tasks such as sorting the images into “acceptable” or
247 “unusable” quality, rescaling the image resolution or size to a certain value, flipping, inversion,
248 rotation, or brightness and noise-level adjustment. It also encompassed more complex tasks such
249 as feature engineering to fuse multiple images of the same nailfold into a one panoramic
250 ‘*mosaic*’, selection of the Region of Interest (RoI), marking of certain ‘landmarks’ on the image,
251 or calculation of certain measurements to draw graphs that would later serve the model.

252 Eight studies reported the exact features to be extracted or calculated to be used by
253 the model in the segmentation/classification task (15, 28-32, 35, 40). As outlined in

254 **Table 3**, capillary detection itself appears to be the most frequently used feature, as it serves as
255 foundation for detecting and/or calculating other metrics such as: capillary count/density, inter-
256 capillary distance, capillary loss, apical width, enlargement, orientation, tortuosity, and
257 derangement. On the contrary, Kassani et al and Shah et al relied on their deep learning models
258 to find their own patterns to successfully classify NFC images, with Kassani's model providing a
259 visual explanation for its prediction (41, 55). **Fig. 5** depicts the process flow starting from image
260 acquisition until the final output by the AI model is produced.

261 **3.4.3 Ground Truth Determination**

262 Clinical status of the participant as recorded in the registry of the dataset, or charts of
263 electronic medical records, was used as the benchmark by four authors (31, 40, 41, 55). Although
264 this method eliminates a lot of subjectivity associated with manual labelling by experts and looks
265 more pragmatically at the entire process through the final end-goal, it is potentially biased with
266 documentation errors of a single expert opinion and lacks the more reliable conclusions of
267 multiple observers. That particular downside provided grounds for the remaining authors to
268 depend on expert annotations or labelling, whether at the capillary level or image level for
269 ground truth determination (15, 28-30, 32, 35). They differed in the number of experts recruited
270 (one to five experts) and in the background of the experts, whether a general NFC technician, a
271 vascular specialist, an internal medicine physician, or a rheumatology specialist who was either a
272 young resident or an experienced attending. Even so, the number of experts per se made a
273 difference. For example, an odd number of 3, as implemented by Yin et al, made it easier to
274 consider a decisive majority-vote in cases of disagreement (28). On the other hand, Tello et al
275 who relied on 5 experts reported significant interobserver disagreement that was achieved in a
276 non-negligible number of images (29). Authors calculated the accuracy, sensitivity and
277 specificity of expert's consensus and used these values as thresholds to compare the model's
278 performance against (30, 31, 40).

279

280 **3.4.4 Training, Validation and Testing**

281 It is essential to distinguish between three different terms that entail model
282 development: training, validation, and testing. 'Training' refers to the main process by which the
283 model "learns" to perform its task, usually through multiple iterations known as epochs.

284 ‘Validation’ usually refers to the process of internal validation, also known as *reliability testing*,
285 in which the model’s ‘knowledge’ is tested, like a mock exam. Finally, ‘Testing’ refers to the
286 actual assessment of the model’s performance on a set that the model was never exposed to
287 before. Conducting model testing adds more objectivity when measuring the model performance
288 to compare it against other models, or to the gold standard. Authors varied considerably, as
289 shown in

290 **Table 3** Comparing AI models across all ten studies their approaches to splitting the dataset into
291 Training: Validation: Testing (Tr:V:T) for model development and in their interchangeable usage
292 of the terms validation and testing.

293 Three authors explicitly stated using *k-fold* cross-validation in their models (30, 41,
294 55). Shah et al used 5-fold cross-validation so that each random set is used 4 times for training
295 and one time for testing by the end of the fifth round. This process was repeated for each of the 5
296 diagnoses (classes). They did not include additional testing using a separate dataset and they
297 used the mean values from all 5 testing sets to be considered the final testing result (55).
298 Garaiman et al also used a variation of 5-fold cross-validation where the original dataset was
299 randomly divided into 5 equal-sized subsets so that Tr:V:T was done in the ratio of 3:1:1. This
300 process was repeated 5 times until each subset had the chance to be in each set. Garaiman also
301 had a separate dataset of 464 images that was randomly selected from the first validation
302 subsample to test the model's performance against manual labeling by experts (30). Finally,
303 Kassani et al split the main dataset of 1,441 into stratified 5-fold cross-validation for both
304 training and validation. Then a separate data set of 287 of both normal and abnormal images was
305 used for testing the model (41).

306 Bharati et al separated the original dataset into 3 groups. Group A included high-
307 resolution mosaics from 10 fingers, half of which was used for training and the other half for
308 validation. Groups B (high-resolution mosaics from 10 fingers) and C (low-resolution mosaics
309 from 4 fingers) were used for the final testing of the model. Thus Tr:V:T was approximately
310 9:9:4 (31). Yin et al first isolated 200 high-quality normal images to train the model, then the
311 remaining 1,588 were divided into Tr:V:T as 6:2:2 (28). Conversely, Tello et al divided the
312 dataset into 85% for training and validation, and 15% for testing in both studies (15, 29).

313 Liu et al divided the dataset into 60% training (30 images) and 40% testing (20
314 images) and compensated for the small sample by generating 20,664 images for training, from
315 the original 30, using a data augmentation technique (32). Berks et al extracted 450 RoIs for
316 training out of a relatively small sample of 80 mosaics. An additional set of 910 images in the
317 dataset was split into two equally balanced sets representing all three classes; one half for
318 internal reliability and the other for final testing (35). Finally, Murray et al split a dataset of 116
319 mosaics randomly into training and testing sets. Then, they utilized a set of linear support vector
320 machine classifiers to train the model based on their labels and the associated features (40).

321 The methodology varied significantly between authors owing to the difference in
322 model architecture, type of learning, balance/imbalance in the datasets, pre-processing analysis
323 techniques, layers of networks in the model, single vs. multi-phase flow between these layers,
324 whether feature detection was first needed before processing, and the quantitative or qualitative
325 nature of output of the model. (15, 28-32, 35, 40, 41, 55). Additionally, authors varied in
326 reporting the specifications of hardware and software used.

327 **3.5 Models Performance**

328 **3.5.1 Nature of Output**

329 Three authors presented their model output solely in qualitative terms (30, 32, 55).
330 Shah et al had the model classify the image out of 5 classes representing healthy status, and four
331 different disease states (55). Gariaman's classification of the image was binary as either diseased
332 or non-diseased (30). Liu et al analyzed the image at the pixel level to classify whether it belongs
333 to a distal capillary or not, so that in the end it presents a binary pixel map showing capillaries or
334 their absence (32). In contrast, two authors reported solely quantitative reports. Berks et al
335 reported capillary count and vessel morphology (35). Yin et al identified capillary count and
336 density (28). Lastly, the remaining five studies combined both qualitative and quantitative results
337 (15, 29, 31, 40, 41).

338 These results were majorly presented at the image level. Only three authors have
339 also presented their results at the global, or participant-level (31, 41, 55). Bharati et al averaged
340 measurements from multiple mosaics of multiple fingers to give a subject-level probability at a
341 single visit (31). Alternatively, Kassani et al considered an aggregate value of $\geq 50\%$ to be the
342 threshold to consider a patient-level prediction successful (41).

343 Lastly, in addition to the main diagnostic model, only Kassani et al and Shah et al
344 had presented a second predictive model that provided additional clinically useful information
345 beyond merely diagnosing the presence or absence of the disease (41, 55). Kassani's predictive
346 model aimed to provide a score for disease activity, which translates to its future severity (41).
347 Similarly, Shah's model predicted a history of a complication (cardiovascular event) in patients
348 with the disease using only the NFC images and the disease status of that particular patient (55).

349 3.5.2 Reporting Metrics

350 Standardized metrics commonly used in computer vision to assess detection of
351 objects (57) overlap with similar metrics commonly used in medicine to describe the
352 performance of diagnostic tests. They usually entail a confusion matrix that shows the predicted
353 against the actual positives and negatives. From those values, many indices can be calculated
354 such as sensitivity, also known as *recall*, specificity, positive predictive value, also known as
355 *precision*, negative predictive value, and accuracy. Certain graphs that efficiently demonstrate
356 the effectiveness of these models are generated from these metrics and indices, such as Area
357 Under the Receiver Operating Curve (AUROC) and Area Under the Precision Recall Curve
358 (AUPRC). Other metrics reported included Intersection over Union (IoU), F1-scores, Pearson-
359 Correlation, and Dice-scores. The lack of standardization of reporting presents a challenge to
360 compare the performance of these models, which is an expected hindrance, given the novelty of
361 this subject. It's highly recommended that a unified approach be presented for future research so
362 that progress can be made. However, for the purposes of this review we present below some of
363 the significant results as reported by their respective authors.

364 *Recall* was reported in 7 studies, either for the final model's ability to classify an
365 NFC image or regarding a feature per se (15, 29-31, 35, 41, 55). It ranged from 93.87% by
366 Garaiman et al in detecting "giant capillaries" (30) down to 70.5% by Berks et al when the
367 model's ability to 'detect capillaries' was compared with observer no.2 as the ground truth (35).
368 *Precision* was reported in 5 studies (15, 28, 29, 35, 41). It's highest value of 0.95 at bootstrapped
369 confidence interval (CI) of 95% was reported by Kassani et al in classifying JDM patients from
370 healthy controls at the image level (41). Berks et al reported the lowest precision of 51.7 in
371 detecting capillaries when observer no.2's labels were used as the ground truth (35). *Specificity*

372 was reported in 5 studies (29-31, 41, 55). Bharati et al reported the highest specificity of 91% for
373 detecting a SSc pattern in high-resolution images of group B from normal controls (31). While
374 the lowest value of 62.2% was reported by Garaiman et al for detecting microhemorrhages (30).

375 *Accuracy* was reported in 4 studies only (30, 32, 35, 41). Berks et al reported the
376 highest accuracy of 93.6% in differentiating distal from non-distal capillaries when the ground
377 truth was determined by consensus of both expert observers (35). However, the lowest range of
378 accuracy of 85.5% - 93.5% was reported by Garaiman et al in delineating the presence of early,
379 active and late patterns of SSc (30). *Area Under the Curve (AUC)* was reported in 5 studies (28,
380 30, 31, 41, 55). The highest AUROC of 97% was reported by Bharati et al in predicting SSc
381 patterns from normal controls in high resolution images of group B (31). Shah et al reported the
382 lowest AUROC of 0.84 in detecting diabetes (55).

383 **3.5.3 General Descriptive Results**

384 The time taken by the algorithm to produce the final output was not reported by most
385 authors, except for Yin et al who reported 0.064 seconds for the capillary density calculation (28)
386 and Garaiman et al who stated that a report can be generated for each patient with 16 images in
387 ~3 seconds; given that labelling 1 image takes 0.19 seconds (30). Kassani et al reported the time
388 it takes to train 1 epoch of the model but not the final time elapsed, including any pre-analysis
389 processing time elapsed (41).

390 The overall performance of all models is summarized in

391 **Table 3.** Compared to the gold standard of manual experts labelling these models have
392 demonstrated equally consistent performance with faster timing (28, 35), or better detection of
393 features (30), or patterns (55). Several models outperformed other AI algorithms such as YOLO5
394 (28), MobileNet (41), U-Net and ResNet (32).

395 Some models were tested in less than ideal conditions, such as low-resolution
396 images (31), or under different lighting conditions (28) to mimic real-life scenarios where the
397 image acquisition process will not yield high-quality images yet the model would still be able to
398 function effectively. Two authors reported lower performance of their models compared to
399 experts in certain areas and they discussed possible explanations for such differences (29, 30).
400 Finally, it is worth noting that Tello et al was the only author to do an external validation study
401 of the previously developed algorithm at capillary.io (29).

402

403 **DISCUSSION**

404 To answer the proposed research questions, we sought to discuss the findings of this
405 review in four sections; given the wide variation in variables contributing to NFC analysis.
406 Section-I lays the background by presenting the NFC technique and its significance in medicine.
407 Section-II addresses the challenges facing the current methodology. Section-III presents the
408 solutions AI offers to overcome said challenges. Finally, Section-IV compiles some
409 considerations for future directions.

410 **Section I: NFC Technique and It's Significance in Medicine**

411 **I - 1 Rationale and Premise of NFC**

412 NFC is a simple non-invasive technique that looks at the microcirculation in the
413 fingers. Arterial blood carrying nutritive oxygenated blood from the heart reaches one of its

414 final destinations at the distal end of fingers/toes. NFC observes nailfolds closely through a
415 microscope to clearly visualize the terminal capillary networks. The distal row of capillaries is
416 seen as a convex hairpin loop that turns around to eventually form venules that carry the waste
417 products away from tissues and back to the heart (1). This healthy microenvironment is tightly
418 regulated and maintained so that significant changes to the capillaries array or morphology do
419 not develop over a short period of time (3). Accordingly, chronic pathological conditions
420 eventually distort the normal homeostasis and induce microvascular changes revealing the long-
421 standing tissue damage (2). Examples of such systemic conditions affecting the whole body
422 where NFC findings were correlated to disease status or progression include many
423 rheumatological diseases such as Systemic Lupus Erythematosus (SLE) (10, 11, 58), Systemic
424 Sclerosis (SSc) (9, 25, 59-61), Ankylosing Spondylitis (AS) (62), inflammatory diseases such as
425 vasculitis (63-65), Idiopathic Inflammatory Myopathies such as polymyositis and Juvenile
426 Dermatomyositis (JDM) (6, 41), inflammatory arthritis (66), dermatological such as
427 dermatomycosis and psoriasis, and finally components of metabolic syndrome such as diabetes
428 (55) and hypertension (32).

429 **I - 2 Advantages and Significance in Medicine**

430 NFC offers countless perks to detect such systemic manifestations. Most
431 importantly, is the fact that it's a non-invasive procedure that yields very valuable information
432 that can help in the diagnosis, prognosis, staging and follow up of patients. Additionally, the
433 equipment needed is relatively cheap, and mobile, guaranteeing broad accessibility for healthcare
434 facilities. Not only that, but the NFC technique to obtain images is by itself user friendly and
435 easy to implement. Coupling such perks with a considerable reliability in the manual analysis of
436 NFC images has rendered the technique the gold standard in assessing nailfold capillaries (67).

437 To illustrate the significance of NFC, studies of abnormal changes in the nailfold
438 capillaries were found to be associated with skin involvement and duration of untreated JDM
439 (68, 69). Likewise, abnormal NFC changes constitute two out of the nine scoring points to fulfill
440 the 2013 ACR/EULAR classification criteria of SSc (70) and is important in stratifying SSc
441 patients into early, active and late (71-73).

442 It is also used to differentiate between types of Raynaud's *Syndrome* (RS). RS is an
443 episodic color change in the fingers with or without pain in response to cold. It can present as a
444 primary benign condition with no evidence of an underlying disease, known as Primary
445 Raynaud's Syndrome OR Raynaud's *disease* (PRD). Or it can be a secondary symptom
446 associated with other medical conditions, like SSc (56), one of the most common causes of
447 Secondary Raynaud Syndrome or Raynaud's *phenomenon* (SRP). Trombetta et al found that a
448 quantitative change in the capillary diameter is predictive of progression of PRD into SRP (17).
449 Dolijanovic et al followed a cohort of 250 PRD patients over six years and found that most
450 participants had normal findings with only 10 out of all subjects (4%) would show SSc pattern 6
451 months before expressing a particular disease (74). Although, this study concluded a
452 considerable lack of reliability in predicting future progression to certain diseases from such non-
453 specific capillary changes in that study population, it did highlight an important finding, and that
454 is if a SSc pattern was found, it would highly correlate to future development of SSc with high
455 specificity and precision.

456 **I - 3 Preprocedural Preparation**

457 Before the procedure, participants are prepared by avoiding caffeine, smoking, stress,
458 and cosmetic procedures on fingers/toes for about 3 weeks before the test day. Then, they are
459 seated in an upright position where the heart is at the level of the nailfold, in a quiet room with a

460 stable temperature of >20 °C for about 15-20 min to provide heat adaptation and endure mental
461 comfort. Next, the nails are cleaned, and an immersion oil is applied to improve visibility and
462 translucency by reducing diffuse reflections (75).

463 The apparatus is then set after deciding the following key aspects. The model and
464 type of the capillaroscopy device is determined and whether it will be fixed or hand-held. Then
465 the magnification is set depending on the camera resolution, its angle and physical distance from
466 the nailfold (contact vs on-contact). Note that calibration to account for such distance might be
467 required; either manually or automatically. Next, the finger/toe to be examined is placed and the
468 light source, its intensity and angle are adjusted for optimal brightness and least reflections.

469 **I – 4 NFC Image Acquisition**

470 Traditional NFC relies on a camera to capture photos of the nailfold. NVC is an
471 updated alternative that records a video instead, then extracts screenshots of frames with good
472 capillary visibility. Some authors like Murray et al described a software that allowed high-
473 magnification panoramic mosaics to be constructed from a video without movement artifacts
474 (40). After images are obtained, they might be manipulated and *pre-processed* to varying
475 degrees, either manually, or using additional software, before they are finally fed into the AI
476 model.

477 **Section II: Current Challenges Facing the Manual Approach**

478 The current golden standard encounters several significant challenges, owing to the
479 sheer variance in variables and circumstances required to obtain NFC images and analyze them.
480 We also comment on how certain authors have approached that area and consider how AI can
481 potentially overcome these obstacles.

482 **II – 1. Lack of standardization of the procedure (technique homogenization)**

483 **II – 1.1 Preprocedural Matters**

484 Starting with patient selection, most studies were unsuccessful in reporting
485 significant demographics related to the selected sample. At the top of the list, differences in skin
486 tone and pigmentation. This vital factor could point towards a potential bias in acquiring images
487 of fair-skinned participants only or in the reliability and validation of the technique itself in dark-
488 skinned patients, propagating error to be carried forward when training and developing the
489 Machine Learning (ML) or Deep Learning (DL) models. With regards to the detailed
490 technicalities of patient preparation before image acquisition, only Garaiman et al intentionally
491 stated referring to the international Delphi consensus described by Ingegnoli et al (75); taking a
492 step forward towards unifying the pre-procedural setting.

493 Consider the two following apparatuses that were described by some authors: a
494 contact hand-held Oplia capillaroscope versus a non-contact fixed Dino-Lite microscope. The
495 first setup permits a broader-angle adjustment and higher zoom to improve visualization, but it
496 runs the risk of unsteady operator hands, and thus image blur due to movement artifacts. On the
497 contrary, the second setup eliminates movement artifacts and can provide a more panoramic
498 view of the entire nailfold, yet it requires higher magnifications and resolution to compensate for
499 the increased distance between the camera lens and the nailfold surface. Low magnifications
500 such as 50x taken by dermatoscopes can provide a broader view of the nailfold, but are not
501 detailed enough to reliably discern capillaries when compared to the more ideally desired
502 magnification of 200x (76).

503 A capillary's morphology varies between hands and feet, finger to finger, and even
504 central to peripheral nailfold sectors (77). According to Cutolo et al, the gold standard technique
505 is to capture at least two adjacent fields of 1mm in the center of the nailfold at 200x (76).

506 Dinsdale et al reported that examination of all 8 fingers is needed, excluding thumbs. Otherwise
507 missing some abnormalities will reduce sensitivity(78). Conversely, Murray et al reported high
508 model performance using only the 4th (ring) finger in the non-dominant hand (40).

509 **II – 1.2 Procedural Considerations**

510 Surprisingly, the duration of the whole process from image acquisition till final
511 output was not reported by many authors. The breakdown of training time, pre-processing time
512 and execution time is also recommended to be reported in future studies to calculate the net time
513 elapsed. Considering how much time pre-analysis processing consumes, it's variation from
514 single simple tasks to complex multi-step editing might necessitate user expertise and training. It
515 is of vital importance that authors report the software used, its version, specifications and tasks
516 performed so that appropriate comparisons can be concluded. For instance, shah et al fed the DL
517 model with images only, to receive the final output (55). On the contrary, Tama et al (79) and
518 Doshi et al (80) adopted complex multi-step approaches such as binarization, skeleton extraction
519 and segmentation, and enhancement operations with alpha-trimmed filter to address non-uniform
520 lighting combined with an iterative rule-based skeletonization procedures.

521 Additionally, Other aspects are not agreed on unanimously, yet. For example,
522 AlGindya et al reported that according to the literature, the green channel of an RGB image
523 shows high contrast between capillaries and the background (81), and therefore used the green
524 filtered images for input. Contrarywise, Liu et al relied on the greyscale because they found that
525 there were no significant differences between the greyscale and the green channel image (32).
526 Another example reported by Liu et al, is that they claimed vertical flipping of an image might
527 change its semantic information in the object hierarchies, in contrast to the horizontal flipping
528 (82).

529 Finally, we emphasize the need to report the specifications of hardware used, in
530 addition to the software. Such factors might be limiting in terms of computational costs, timing,
531 and even critical decisions such as feeding multiple measurements as input versus relying on a
532 single average value, hence determining the level of processing, pixel-wise, capillary level, or
533 image level.

534 **II – 2. Image Ambiguity**

535 The ambiguity in images is usually a result of its poor quality and/or presence of
536 artifacts that could be due to a wide range of factors. Camera-related factors include low
537 contrast/resolution images (mostly in low-cost devices), high image noise, lighting issues such as
538 reflections on the oil, non-uniform lighting, and extremes of brightness. It could be due to
539 physical factors such as air bubbles in the immersion oil, dust on lenses, dirt on/in fingernails, or
540 blurring due to movement of patient fingers during imaging or examiner’s hand if it’s a hand-
541 held device. Lastly, such variance may be disease-related such as too much fibrosis as in the late
542 stages of SSc, or due to presence of non-delineable structures. Poor reporting and lack of
543 standardization in tolerance thresholds to all these procedural parameters will result in increased
544 image heterogeneity and accordingly, higher interobserver variability that will eventually be
545 transferred to the ML model.

546 **II – 3. The Subjective Nature of Analyses by Human Experts**

547 Manual analysis of NFC images, whether by clinicians or trained personnel, exhibits
548 considerable subjectivity that massively influences the gold-standard technique as well as the
549 input to ML/DL models, and therefore their predictions. Firstly, *operator bias* is demonstrated
550 through their reliance on intuition to select fields, capillaries and in classifying them instead of
551 examining all capillaries in each image. It also manifests as rough estimates of features instead of

552 relying on accurate measurements and indices. Finally, numerous cognitive biases could develop
553 if analysis is done, for example, right after reviewing a patient chart. Other expected
554 shortcomings due to human operators include the need for training to raise expertise (23) in
555 addition to user, owing to the time consuming tasks to identify, label structures, categorize them
556 and calculate indices; especially when eight fingers are examined to maintain a high sensitivity
557 (78).

558 Inter-individual variability is yet another significant aspect due to the multi-variable
559 nature of the task. Despite testing its reliability, high interobserver variability still poses a
560 considerable bias especially when images get less clear/more ambiguous (67, 83). Saez et al in
561 the GEAS survey found considerable heterogeneity between capillaroscopy experts, particularly
562 when considering morphological differences and not just categorical normal versus abnormal
563 patterns (84). Similarly, Garaiman et al reported high agreements especially regarding giant
564 capillaries and microhemorrhages, as well as regarding assessment of patterns (e.g. SSc Pattern),
565 but not very much regarding capillary loss and enlargement (30). Such high inter-rater variance
566 presents a challenge, per se, when the performance of the algorithm is judged against experts as
567 the decision of which expert to be regarded as the benchmark is made. Thus authors like Tello et
568 al 2023 (85) and Garaiman et al (30) reported an acceptable decent performance by the algorithm
569 despite most experts performing better than the algorithm. Furthermore, the agreement is not
570 only low with regards to the grading of each image, but also in selecting which areas to be
571 evaluated (35). That was the basis for a more standardized criteria called the ‘fast track
572 algorithm’ that was developed to help ease and standardize the grading process (60, 72, 86, 87).

573 **II – 4. Lack of Agreement on Which Features to be Extracted?**

574 Current practice only looks at the distal row of nailfold capillaries. Some general
575 features of a normal nailfold include a transparent skin with clearly visible capillaries, absent
576 pericapillary oedema, visible subpapillary venous plexus (in up to 30% of healthy people), and
577 similar-looking capillaries that are regular in arrangement, mostly straight and perpendicular to
578 the nailfold (88).

579 Most of the features considered for analysis are qualitative and they can be a singular
580 feature such as a U-shaped hairpin-convexity constituting the normal capillary morphology. Or it
581 could be the absence of singular features such as tortuosity, ramifications, neo-angiogenesis, 3-
582 point crossing, non-convex tip, and hemorrhages. (89, 90).

583 An additional way of analysis is through a collection of certain features, known as a
584 pattern, that defines a certain abnormal condition from the normal population. For example,
585 Raynaud's Syndrome is an episodic color change in the fingers with or without pain in
586 response to cold. Raynaud's syndrome can present as a primary condition with no evidence of
587 an underlying disease, known as Primary Raynaud's Syndrome OR Raynaud's disease (PRD).
588 Or it can be a secondary symptom associated with other medical conditions, like SSc (56), one
589 of the most common causes of Secondary Raynaud Syndrome or Raynaud's phenomenon
590 (SRP). To distinguish PRD from SRP, Mannarino et al (91) used the following pattern: altered
591 arrangement of capillary loop, decrease in the number of capillaries, and abnormal ramifications.
592 Diagnosing SRP is significant because it is considered a reliable early parameter for diagnosing
593 early SSc that has been clinically validated. One important way of doing that distinction,
594 according to Bharati et al is that if PRD are negative clinically and serologically, they are likely
595 normal (31).

596 Another very common and significant example relates to SSc. It is a potentially
597 lethal autoimmune disease characterized by 3 hallmarks: Microangiopathy, production of
598 disease-specific autoantibodies and deposition of extracellular matrix proteins resulting in tissue
599 fibrosis. (70, 92) The microangiopathy usually manifests as low capillary density, high capillary
600 dimensions (dilation or giant), and abnormal morphology and hemorrhage (73, 93, 94). Smith et
601 al defined SSc pattern as very low capillary density (<3 capillaries.mm), or the presence of giant
602 capillaries (72). Murray et al added high tortuosity and derangement; that is disorganization in
603 the overall direction of capillaries (40).

604 The same disease, like SSc, could show different patterns across different stages of the
605 same disease. that is typically used to help monitor improvement or disease progression on
606 follow up. For example, early, active and late/severe patterns of SSc have been identified and
607 validated in clinical studies, commonly the one described by Cutolo et al and standardized by
608 Smith et al. (17, 25, 95, 96). The importance of recognizing such patterns helps monitor disease
609 progression or improvement on follow up. This finding presents an opportunity for early
610 intervention before severe organ involvement could occur, or potentially predicting future
611 disease status. Finally, Other less understood non-specific set of features are as known as
612 disease-associated changes (15, 30, 85)

613 A few quantitative features have also been described such as avascular areas defined
614 as distance between 2 loops > 500 μ m, that normally should be absent (88). Interestingly, it was
615 demonstrated that calculation of the mean score of such capillary loss could be reliably reduced
616 from 32 NVC images (four fields per finger for eight fingers of the patient analyzed) to eight
617 NVC images (one field per finger for eight fingers of the patient analyzed). This finding can save

618 valuable patient time as well computational costs, stressing on the significance of disclosing all
619 aspects of the process to help deduce such conclusions (97).

620 Another commonly calculated feature is capillary density (normally = 9-13 in 1
621 linear mm). Density is related to tissue perfusion and microvascular function. The less normal
622 conditions are the lower the density. Finally, many measurements can be calculated for each
623 capillary, including: arterial limb diameter, venous limb diameter, apical loop diameter (7)
624 (normally < 20 μm (25)) and total width. From these calculations, two important descriptions
625 emerge, namely a dilated capillary (typically 20-50 μm), or a giant/mega capillary (typically >50
626 μm) (25) that are normally absent in a healthy individual.

627 **II – 5. Lack of Standardization of such Features**

628 **II – 5.1 Defining a Parameter or Pattern**

629 Unfortunately, the aforementioned features lack unanimous agreement on a validated
630 and clinically relevant definition. To demonstrate, let's consider the capillary density. Generally
631 speaking, Neubauer et al described a normal density as 9-13 per linear mm (88), while Cutolo et
632 al relied on capillary loss as <7 capillaries per mm (98). Alternatively, in et al didn't rely on a
633 specific criteria and they rather compared their model's performance to that of the experts count
634 in normal healthy volunteers (28). On the other hand, Bharati et al were more specific in their
635 definition as they counted all capillaries from left-most to right-most within each mosaic image
636 of the entire nailfold and divided that number by the distance in mm between the same left-most
637 to right-most (31).

638 Another variation lies in the approach to detecting a terminal capillary loop/apex in
639 the distal line of capillaries to differentiate it from other detected vascular structures. For
640 instance, some authors rely on the direct observation method compared to the 90° degrees

641 method described by Hofstee et al (67) and used by Yin et al (28). Alternatively, Karbalaie et
642 al developed the semi-automated Elliptic Broken Line (EBL) method (99). Recently, Bharati et
643 al used the fully automated deep learning model U-Nets for apex candidate generation, followed
644 by ResNet34 for candidate classification (31).

645 The difficulties due to lack of such standardization is not unclear, as expressed by
646 Cutolo et al, regarding the reliability of simple capillaroscopic definitions in describing capillary
647 morphology in rheumatic diseases (89). In an attempt to overcome these differences some
648 authors proposed potential standards to be followed whether in defining a single parameter or a
649 pattern. For instance, Jones et al proposed a taxonomy for the morphology of nailfold capillaries
650 (38). Neu-bauer Geryk et al described a possible standardized technique with proposed normal
651 values to define a normal for healthy subjects (88). Smith et also developed well-defined criteria
652 that was adopted by the EULAR society to describe normal ($<20\ \mu\text{m}$), dilated ($20\text{-}50\ \mu\text{m}$), and
653 (giant/mega $>50\ \mu\text{m}$) capillaries (25). Smith et al and Cutolo et al attempted to standardize the
654 defining criteria for SRP in SSc and how it differs from PRD (25, 72, 73). Cutolo et al also
655 initiated a similar proposition for a different disease like SLE (11).

656 **II – 5.2 Clinical Relevance and Validation of Features**

657 Surely, studying these features without any clinical relevance considerations or
658 reliability testing will render the entire process purely academic and hinder its progression to a
659 practically applicable tool. Such studies will emphasize the ‘more significant’ features so that
660 they get assigned a higher weight. As an example, Trombetta et al found that a capillary diameter
661 of $> 30\ \mu\text{m}$ is an independent predictor for the progression of PRD patients to SRP (17).
662 Similarly, capillary density was found to be a significant quantitative parameter in studies of
663 conditions like diabetes (100), connective tissue diseases, pulmonary hypertension in SSc

664 patients, and chronic kidney disease (99) that would relate to a particular aspect of these
665 diseases. Understanding the implications of this matter, the SCLEROCAP study by Boulon et al
666 was conducted to assess the reliability of Maricq and Cutolo’s classification of capillaroscopic
667 features that can stage SSc patients, or rather predict their status (101). Accordingly, efforts to
668 propagate and disseminate the validation of these features should be encouraged to guide future
669 research in towards a more patient-oriented direction, whether these features were relating to the
670 study of condition at hand, or other unknown conditions. Moreover, certain features might be
671 deliberately, and safely, ignored, like what Bharati et al did. They disregarded microhemorrhages
672 from being input into the model since they saw the difficulty to define it with lack of its
673 relevance to diagnosing SSc (31).

674 For that reason, authors like Tullo et al (29) decided to detect a feature such as
675 *tortuosity*, even though it had no current validating studies, to understand possible ways it could
676 prove relevant. Typically, it has been looked at with a quantitative approach, calculating the
677 percentage of tortuous capillaries to discern normal healthy individuals (<5%) from other
678 conditions that could potentially manifest at certain higher cutoffs such as 40% or 70% (102).

679 **II – 5.3 Decisions in Conflicting Matters**

680 Defining hierarchical principles in labelling a capillary is an important step that
681 needs to be addressed especially after these labels will be used to train ML models. For example,
682 Tello et al (103) prioritized classifying a capillary as being ‘giant’ over ‘tortuous’, if both
683 features were present in the same capillary. Additionally, Garaiman et al noticed sub-optimal
684 consistency in NFC image labels owing to the simple fact the different attending physicians
685 have analyzed images of the same patient at different time points, thus affecting the ground truth
686 of the developed model (30). Another decision would arise if the feature at hand is not clearly

687 visualized. For example, Tello et al relied on the width of limbs as an acceptable alternative (in
688 case the apical diameter wasn't clear enough) to detect dilated/giant capillaries(29).

689 Furthermore, when experts, and henceforth, ML models, aim to assign patient-level
690 diagnoses, the approaches differ. For example, some experts would label a participant as
691 abnormal if 'any' NFC image from anyone of the finger is labelled as abnormal. Conversely,
692 Bharati et al averaged measurements from all participants to achieve global-level labelling (31),
693 while Kassani et al considered a participant diseased if the aggregate from all pictures was \geq
694 50% (41) .

695 **Section III: How AI Is Solving These Challenges**

696 **III – 1 Evolution of AI beyond simple automation**

697 Artificial Intelligence (AI) was first described as a term by John McCarthy in 1956
698 as the science and engineering of making intelligent machines. Despite the difficulty of defining
699 exactly the aspects of human thinking, rationalization and decision-making processes, AI aims to
700 simulate one or more of these complex processes, at least partly. Automation of any process
701 relies on algorithms which are mathematical operations or code, that can be simple or a multi-
702 step series, that are fed a particular type of input to process in a precisely defined manner and
703 give a certain output. If it was given any new or unusual input that it's not programmed to
704 process, it will fail to deal with it. AI, however, is more than a simple automated code, that
705 performs 'smart' functions, in an attempt to mimic human reasoning. ML, which is a subtype of
706 AI, takes it a step further, by learning how to deal with new information on its own without much
707 prior programming, exactly like a human child learning. DL is a further subtype of ML that tries
708 to learn entirely on its own by interacting with its environment in certain ways and developing a
709 'brain' of its own that is a black box to its developer.

710 A common way of teaching these machines how to learn is the supervised approach
711 through input-output pairs. In our case, this means showing the model a normal NFC image and
712 telling it that it is normal or showing it a particular feature and telling (labelling) it what it is.
713 Another would be to let the model figure out differences in input on its own by feeding it only
714 the images, known as unsupervised. Other methods employ a mixture of both include semi-
715 supervised and reinforcement learning whether using a human agent in the loop or not. Although
716 ML is a smart enough machine that can do incredibly difficult tasks, DL models seem to look at
717 things through a ‘fresh set of eyes’ that offers an unparalleled and novel approach to analyses.

718 **III – 2 Superiority of AI**

719 Overcoming the challenges that come with NFC image ambiguity (as discussed in
720 section II-2) had posed a huge obstacle for experts to interpret these images. AI models can
721 detect minute structures that are invisible to the naked eye and find relations between an
722 incredible number of parameters within a very short time frame. Such qualities allow it to
723 circumvent such heterogeneities, or even edit them before analyses in a fully automated, and
724 possibly real-time, manner.

725 AI has also eliminated a lot of (not all) the errors attributed to human operation.
726 Provided with decent hardware, an AI model can compute an insanely massive number of
727 parameters at the same time to produce a consistent more objective output within seconds, if not
728 less. Its training is less exhaustive and consuming compared to human training, and it would be
729 able to analyze so much more with less resources. Such capabilities can potentially present a
730 more objective output that relies on certain criteria, compared to human experts. As an example
731 of how AI models can reduce operator bias, can be demonstrated as many experts usually rely on
732 their subjective intuition to select a few capillaries within a sector of an NFC image of a single

733 finger to classify it rough manner. An AI model, will instead look at all the capillaries in the
734 entire nailfold of all fingers, take precise measurements and draw indices related to all of them,
735 extract much more features and issue a more comprehensive and thorough judgment in a
736 fraction of the time taken by experts. Not only that, but its output wouldn't be influenced by
737 reading patient's charts, as might happen with experts.

738 Additionally, future predictions about the potential risk of developing a certain
739 condition can be drawn. Current models have demonstrated the potential to fill the gap in
740 identifying predictive features correlating to actual diseases progression (33).

741 **Section IV: AI Considerations and Future Directions**

742 **IV -1 Current Limitations and Challenges**

743 Thorough documentation and comprehensive reporting of the entire process, from
744 patient selection and pre-procedural preparation to image acquisition and analyses cannot be
745 overstated. It is essential to recognize that as long as humans exist, their biases will too. The
746 question therefore becomes how to minimize propagating these biases to AI. For example, the
747 scarcity of datasets in dark skinned population will render these AI models inefficient in
748 performing the same task compared light-skinned populations.

749 Similarly, errors and biases in the preprocessing, as pointed out by Murray et al may
750 shift towards a particular parameter more than others, and eventually skew the mode towards
751 over/under estimations (40). Moreover, standardizing the measurement units' whether real or
752 arbitrary will close the gap towards a more objective feature extraction and therefore, output. An
753 additional drawback is the limited open-access datasets that newer models can use for training.

754 Condition-specific factors may also limit the predictive models' accuracy. For
755 example, Kassani et al reported difficulty in predicting JDM disease activity due to difficulty in

756 measuring disease quiescence (and limitations in doing so using the manual DAS itself, as it fails
757 to sensitively capture disease activity compared to other recent biomarkers) and the fact the
758 capillary damage may persist with inactive disease(41). That example enforces the fact that
759 biases are carried forward with the AI, and that the burden of providing such answers falls on the
760 developer, not the AI model. That is not to say that AI models don't share the burden of ethical
761 responsibility to be explainable, but that it highlights the usually overlooked human bias.

762 **IV-2 The Future of AI in NFC**

763 AI models can be further enhanced by feeding them with more information that just
764 an NFC image, such as a patient's electronic medical record, similar to what Sheh et al did (55).
765 Impressive results already established when NFC was fused with other modalities such as
766 ultrasound (104, 105), doppler sonography (106), laser scanning microscopy(107), and
767 optoacoustic imaging (108, 109).

768

769 **CONCLUSION**

770 AI models have demonstrated a truly remarkable potential as a clinical decision-
771 supportive tool. Owing to the novel nature of this technology, it is of the utmost significance for
772 authors to report future NFC-related studies in the most comprehensive way possible,
773 particularly population demographics and execution times. That is to overcome propagating
774 human biases to AI models, standardize the NFC methodology and reporting metrics to allow for
775 comparisons and conclusions to be made. ML and DL models succeeded in producing a fully
776 automated and objective quantitative output that will form the basis for future prediction and
777 patient-reported outcomes research. Fusion of NFC with other technologies like doppler laser
778 and optoacoustic imaging to enhance the extraction of features and the precision of measured

779 values will further increase the sensitivity and specificity of such tools to be very efficient in
780 daily clinical practice. Future studies should focus on the deployment and provision of full
781 functionality for these types of applications using explainable AI. Quality assessment standards
782 and ethical considerations still present a big challenge in reporting and in testing the safety of
783 these techniques. Finally, with more external validation studies across multiple different settings
784 in different populations these tools will revolutionize diagnostics through NFC image
785 interpretation.

786

787 **OTHER INFORMATION**

788 **FUNFING AND CONFLICT OF INTERESET**

789 The authors of this review declare no potential conflicts of interest regarding research,
790 authorship, or publication. This review was not supported by any source of financial aid.

791 **ACKNOWLEDGMENTS**

792 **Figure 1** was created using MS Word

793 **Figures 2, 3 and 4** were created using MS Excel

794 **Figure 5** were created using MS Paint

795 **LEGENDS**

796 **Table 1.** Research questions and their rationale.

797 **Table 2.** Compares study design and population characteristics across all ten studies.

798 **Table 3.** Compres AI model architecture, development and performance across all ten studies.

799 **Fig. 1** A PRISMA Flowchart summarizing the process of acquiring the final included studies.

800 **Fig. 2** Geographical distribution of the countries where included studies were conducted.

801 **Fig. 3** Frequency of diseases addressed by the included studies.

802 **Fig. 4** Comparison of the total sample sizes across all included studies.

803 **Fig. 5** A summary of the whole process of NFC and AI analysis.

804 **S.1** Shows the risk of bias and quality assessment of all included studies using QUADAS 2.

805 REFERENCES

- 806 1. Cracowski JL, Roustit M. Human Skin Microcirculation. *Compr Physiol*. 2020;10(3):1105-54.
- 807 2. Haggerty A, Nirmalan M. Capillary dynamics, interstitial fluid and the lymphatic system. *Anaesthesia & Intensive Care Medicine*. 2022;23(2):130-7.
- 808 3. Chovatiya R, Medzhitov R. Stress, inflammation, and defense of homeostasis. *Mol Cell*.
- 809 2014;54(2):281-8.
- 810 4. Goydin AP, Shutova SV, Fabrikantov OL. Evaluation of the diagnostic capabilities of nailfold
- 811 capillaroscopy in diabetic retinopathy. *Vestnik oftalmologii*. 2023;139(1):16-26.
- 812 5. Shah R, Petch J, Nelson W, Roth K, Noseworthy MD, Ghassemi M, et al. Nailfold capillaroscopy
- 813 and deep learning in diabetes. *J Diabetes*. 2023;15(2):145-51.
- 814 6. Zisa D, Bloostein A, Jannat-Khah D, Cutolo M, Smith V, Fernandez D. Nailfold
- 815 Videocapillaroscopic Abnormalities Correlate with Disease Activity Measures and Cutaneous Damage in
- 816 Patients with Idiopathic Inflammatory Myopathies. *Arthritis and Rheumatology*. 2022;74:3710-1.
- 817 7. Etehad Tavakol M, Fatemi A, Karbalaie A, Emrani Z, Erlandsson BE. Nailfold Capillaroscopy in
- 818 Rheumatic Diseases: Which Parameters Should Be Evaluated? *Biomed Res Int*. 2015;2015:974530.
- 819 8. Cutolo M, Matucci Cerinic M. Nailfold capillaroscopy and classification criteria for systemic
- 820 sclerosis. *Clinical and Experimental Rheumatology*. 2007;25(5):663-5.
- 821 9. Ingegnoli F, Gualtierotti R. A systematic overview on the use and relevance of capillaroscopy in
- 822 systemic sclerosis. *Expert Rev Clin Immunol*. 2013;9(11):1091-7.
- 823 10. Marasco E, Duesing C, Keymel S, Zanframundo G, Codullo V, Sander O, et al. ANALYSIS OF
- 824 NAILFOLD CAPILLAROSCOPY FINDINGS AND CLINICAL FEATURES OF PATIENTS WITH
- 825 SYSTEMIC LUPUS ERYTHEMATOSUS AND PULMONARY ARTERIAL HYPERTENSION.
- 826 *Annals of the Rheumatic Diseases*. 2023;82:1504-5.
- 827 11. Cutolo M, Melsens K, Wijnant S, Ingegnoli F, Thevissen K, De Keyser F, et al. Nailfold
- 828 capillaroscopy in systemic lupus erythematosus: A systematic review and critical appraisal.
- 829 *Autoimmunity Reviews*. 2018;17(4):344-52.
- 830 12. Zeni S, Ingegnoli F. [Raynaud's phenomenon]. *Reumatismo*. 2004;56(2):77-81.
- 831 13. Matucci-Cerinic C, Nagaraja V, Prignano F, Kahaleh B, Bellando-Randone S. The role of the
- 832 dermatologist in Raynaud's phenomenon: a clinical challenge. *J Eur Acad Dermatol Venereol*.
- 833 2018;32(7):1120-7.
- 834 14. Smith V, Ickinger C, Hysa E, Snow M, Frech T, Sulli A, et al. Nailfold capillaroscopy. *Best*
- 835 *Practice & Research Clinical Rheumatology*. 2023;37(1):101849.
- 836 15. Gracia Tello B, Ramos Ibañez E, Fanlo Mateo P, Sáez Cómet L, Martínez Robles E, Ríos Blanco
- 837 JJ, et al. The challenge of comprehensive nailfold videocapillaroscopy practice: a further contribution.
- 838 *Clin Exp Rheumatol*. 2022;40(10):1926-32.
- 839 16. Pauling JD. Could nailfold videocapillaroscopy usher in a new era of preventative disease-
- 840 modifying therapeutic intervention in systemic sclerosis? *Rheumatology (United Kingdom)*.
- 841 2017;56(7):1053-5.
- 842 17. Amelia Chiara T, Vanessa S, Carmen P, Marianna M, Sabrina P, Caterina C, et al. Quantitative
- 843 Alterations of Capillary Diameter Have a Predictive Value for Development of the Capillaroscopic
- 844 Systemic Sclerosis Pattern. *The Journal of Rheumatology*. 2016;43(3):599.
- 845

- 846 18. Kassani PH, Ehwerhemuepha L, Martin-King C, Kassab R, Gibbs E, Morgan G, et al. Artificial
847 intelligence for nailfold capillaroscopy analyses – a proof of concept application in juvenile
848 dermatomyositis. *Pediatric Research*. 2024;95(4):981-7.
- 849 19. Cutolo M, Trombetta AC, Melsens K, Pizzorni C, Sulli A, Ruaro B, et al. Automated assessment
850 of absolute nailfold capillary number on videocapillaroscopic images: Proof of principle and validation in
851 systemic sclerosis. *Microcirculation*. 2018;25(4):e12447.
- 852 20. Mazedo C, Silva S, Aguiar R, Barcelos A, Bastos JM. Correlation between arterial stiffness and
853 nailfold capillary microscopic abnormalities in systemic sclerosis: Results from a single centre cross-
854 sectional study. *Artery Research*. 2021;27:S22.
- 855 21. Nevskaya T, Baron M, Pope J. Criterion-related validity of european scleroderma study group
856 activity index in an early scleroderma cohort. *Journal of Rheumatology*. 2017;44(6):878.
- 857 22. Ornowska S, Chojnowski M, Felis-Giemza A, Dudek Ł, Olesińska M. Microvascular damage – a
858 marker of specific organ involvement in mixed connective tissue disease? *Reumatologia*. 2021;59(2):115-
859 20.
- 860 23. Herrick AL, Berks M, Taylor CJ. Quantitative nailfold capillaroscopy-update and possible next
861 steps. *Rheumatology (Oxford)*. 2021;60(5):2054-65.
- 862 24. Cutolo M, Gotelli E, Smith V. Reading nailfold capillaroscopic images in systemic sclerosis:
863 manual and/or automated detection? *Rheumatology (United Kingdom)*. 2023;62(7):2335-7.
- 864 25. Smith V, Herrick AL, Ingegnoli F, Damjanov N, De Angelis R, Denton CP, et al. Standardisation
865 of nailfold capillaroscopy for the assessment of patients with Raynaud's phenomenon and systemic
866 sclerosis. *Autoimmunity Reviews*. 2020;19(3):102458.
- 867 26. Lenharo M. An AI revolution is brewing in medicine. What will it look like? *Nature*.
868 2023;622(7984):686-8.
- 869 27. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med*. 2022;28(1):31-
870 8.
- 871 28. Yin H, Wu Z, Huang A, Luo J, Liang J, Lin J, et al. Automated nailfold capillary density
872 measurement method based on improved YOLOv5. *Microvasc Res*. 2023;150:104593.
- 873 29. Tello BCG, Ibañez ER, Comet LS, Del Castillo AG, Aznar CPS, O'Callaghan AS, et al. External
874 clinical validation of automated software to identify structural abnormalities and microhaemorrhages in
875 nailfold videocapillaroscopy images. *Clinical and Experimental Rheumatology*. 2023;41(8):1605-11.
- 876 30. Garaiman A, Nooralahzadeh F, Mihai C, Gonzalez NP, Gkikopoulos N, Becker MO, et al. Vision
877 transformer assisting rheumatologists in screening for capillaroscopy changes in systemic sclerosis: an
878 artificial intelligence model. *Rheumatology (Oxford)*. 2023;62(7):2492-500.
- 879 31. Bharathi PG, Berks M, Dinsdale G, Murray A, Manning J, Wilkinson S, et al. A deep learning
880 system for quantitative assessment of microvascular abnormalities in nailfold capillary images.
881 *Rheumatology (Oxford, England)*. 2023;62(6):2325-9.
- 882 32. Liu S, Li Y, Zhou J, Hu J, Chen N, Shang Y, et al. Segmenting nailfold capillaries using an
883 improved U-net network. *Microvasc Res*. 2020;130:104011.
- 884 33. Urwin SG, Griffiths B, Allen J. Quantification of differences between nailfold capillaroscopy
885 images with a scleroderma pattern and normal pattern using measures of geometric and algorithmic
886 complexity. *Physiol Meas*. 2017;38(2):N32-n41.
- 887 34. Cheng C, Lee CW, Daskalakis C. A Reproducible Computerized Method for Quantitation of
888 Capillary Density using Nailfold Capillaroscopy. *J Vis Exp*. 2015(105):e53088.
- 889 35. Berks M, Tresadern P, Dinsdale G, Murray A, Moore T, Herrick A, et al. An automated system
890 for detecting and measuring nailfold capillaries. *Med Image Comput Comput Assist Interv*. 2014;17(Pt
891 1):658-65.
- 892 36. Gronenschild EH, Muris DM, Schram MT, Karaca U, Stehouwer CD, Houben AJ. Semi-
893 automatic assessment of skin capillary density: proof of principle and validation. *Microvasc Res*.
894 2013;90:192-8.
- 895 37. Goffredo M, Schmid M, Conforto S, Amorosi B, D'Alessio T, Palma C. Quantitative color
896 analysis for capillaroscopy image segmentation. *Med Biol Eng Comput*. 2012;50(6):567-74.

- 897 38. Jones BF, Oral M, Morris CW, Ring EF. A proposed taxonomy for nailfold capillaries based on
898 their morphology. *IEEE Trans Med Imaging*. 2001;20(4):333-41.
- 899 39. Hu Q, Mahler F. New system for image analysis in nailfold capillaroscopy. *Microcirculation*.
900 1999;6(3):227-35.
- 901 40. Murray AK, Feng K, Moore TL, Allen PD, Taylor CJ, Herrick AL. Preliminary clinical
902 evaluation of semi-automated nailfold capillaroscopy in the assessment of patients with Raynaud's
903 phenomenon. *Microcirculation*. 2011;18(6):440-7.
- 904 41. Kassani PH, Ehwerhemuepha L, Martin-King C, Kassab R, Gibbs E, Morgan G, et al. Artificial
905 intelligence for nailfold capillaroscopy analyses – a proof of concept application in juvenile
906 dermatomyositis. *Pediatric Research*. 2023.
- 907 42. Liu PR, Lu L, Zhang JY, Huo TT, Liu SX, Ye ZW. Application of Artificial Intelligence in
908 Medicine: An Overview. *Curr Med Sci*. 2021;41(6):1105-15.
- 909 43. Poon AIF, Sung JJY. Opening the black box of AI-Medicine. *J Gastroenterol Hepatol*.
910 2021;36(3):581-4.
- 911 44. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA
912 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.
- 913 45. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an
914 updated list of essential items for reporting diagnostic accuracy studies. *Bmj*. 2015;351:h5527.
- 915 46. Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS:
916 a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC*
917 *Medical Research Methodology*. 2003;3(1):25.
- 918 47. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI
919 statement: updated guidance for reporting clinical prediction models that use regression or machine
920 learning methods. *BMJ*. 2024;385:e078378.
- 921 48. Sounderajah V, Ashrafian H, Rose S, Shah NH, Ghassemi M, Golub R, et al. A quality
922 assessment tool for artificial intelligence-centered diagnostic test accuracy studies: QUADAS-AI. *Nature*
923 *Medicine*. 2021;27(10):1663-5.
- 924 49. Sounderajah V, Ashrafian H, Golub RM, Shetty S, De Fauw J, Hooft L, et al. Developing a
925 reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI
926 protocol. *BMJ Open*. 2021;11(6):e047709.
- 927 50. Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, et al. Protocol for
928 development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic
929 and prognostic prediction model studies based on artificial intelligence. *BMJ Open*. 2021;11(7):e048008.
- 930 51. Jayakumar S, Sounderajah V, Normahani P, Harling L, Markar SR, Ashrafian H, et al. Quality
931 assessment standards in artificial intelligence diagnostic accuracy systematic reviews: a meta-research
932 study. *npj Digital Medicine*. 2022;5(1):11.
- 933 52. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: A
934 Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Annals of Internal Medicine*.
935 2011;155(8):529-36.
- 936 53. Hernandez-Boussard T, Bozkurt S, Ioannidis JPA, Shah NH. MINIMAR (MINimum Information
937 for Medical AI Reporting): Developing reporting standards for artificial intelligence in health care. *J Am*
938 *Med Inform Assoc*. 2020;27(12):2011-5.
- 939 54. Olczak J, Pavlopoulos J, Prijs J, Ijpma FFA, Doornberg JN, Lundström C, et al. Presenting
940 artificial intelligence, deep learning, and machine learning studies to clinicians and healthcare
941 stakeholders: an introductory reference with a guideline and a Clinical AI Research (CAIR) checklist
942 proposal. *Acta Orthop*. 2021;92(5):513-25.
- 943 55. Shah R, Petch J, Nelson W, Roth K, Noseworthy MD, Ghassemi M, et al. Nailfold capillaroscopy
944 and deep learning in diabetes. *Journal of Diabetes*. 2023;15(2):145-51.
- 945 56. Maciejewska M, Sikora M, Maciejewski C, Alda-Malicka R, Czuwara J, Rudnicka L. Raynaud's
946 Phenomenon with Focus on Systemic Sclerosis. *J Clin Med*. 2022;11(9).
- 947 57. Zhu M. Recall, precision and average precision. 2004.

- 948 58. Schonenberg-Meinema D, Bergkamp SC, Nassar-Sheikh Rashid A, Gruppen MP, Middelkamp-
949 Hup MA, Armbrust W, et al. Nailfold capillary scleroderma pattern may be associated with disease
950 damage in childhood-onset systemic lupus erythematosus: important lessons from longitudinal follow-up.
951 *Lupus Sci Med*. 2022;9(1).
- 952 59. Pawlik KK, Bohdziewicz A, Maciejewska M, Prado J, Czuwara J, Olszewska M, et al. evaluation
953 of cutaneous microcirculation in systemic sclerosis. An update. *Przegląd Dermatologiczny*.
954 2023;110(4):499-517.
- 955 60. Smith V, Vanhaecke A, Guerra M, Angelis RD, Deschepper E, Denton C, et al. Fast track
956 algorithm: How to differentiate a scleroderma pattern from a non-scleroderma pattern. *Annals of the*
957 *Rheumatic Diseases*. 2019;78:1224-5.
- 958 61. Schonenberg-Meinema D, Bergkamp S, Nassar-Sheikh Rashid A, Gruppen M, Hak AE, Hissink
959 Muller PCE, et al. Scleroderma pattern in nailfold capillaries of (childhood-onset) systemic lupus
960 erythematosus: Lessons from longitudinal follow-up. *Annals of the Rheumatic Diseases*. 2021;80(SUPPL
961 1):937.
- 962 62. Batko B, Maga P, Urbanski K, Ryszawa-Mrozek N, Schramm-Luc A, Koziej M, et al.
963 Microvascular dysfunction in ankylosing spondylitis is associated with disease activity and is improved
964 by anti-TNF treatment. *Sci Rep*. 2018;8(1):13205.
- 965 63. Keret S, Mazzawi J, Slobodin G, Rimar O, Rosner I, Rozenbaum M, et al. Nailfold video
966 capillaroscopy as a useful diagnostic tool in systemic vasculitis. *Microvasc Res*. 2022;143:104406.
- 967 64. Matsuda S, Kotani T, Wakura R, Suzuka T, Kuwabara H, Kiboshi T, et al. Examination of
968 nailfold videocapillaroscopy findings in ANCA-associated vasculitis. *Rheumatology (Oxford)*.
969 2023;62(2):747-57.
- 970 65. Screm G, Mondini L, Confalonieri P, Salton F, Trotta L, Barbieri M, et al. Nailfold
971 Capillaroscopy Analysis Can Add a New Perspective to Biomarker Research in Antineutrophil
972 Cytoplasmic Antibody-Associated Vasculitis. *Diagnostics [Internet]*. 2024; 14(3).
- 973 66. Lambova SN, Müller-Ladner U. Capillaroscopic pattern in inflammatory arthritis. *Microvasc Res*.
974 2012;83(3):318-22.
- 975 67. Hofstee HMA, Serné EH, Roberts C, Hesselstrand R, Scheja A, Moore TL, et al. A multicentre
976 study on the reliability of qualitative and quantitative nail-fold videocapillaroscopy assessment.
977 *Rheumatology*. 2011;51(4):749-55.
- 978 68. Smith RL, Sundberg J, Shamiyah E, Dyer A, Pachman LM. Skin involvement in juvenile
979 dermatomyositis is associated with loss of end row nailfold capillary loops. *J Rheumatol*.
980 2004;31(8):1644-9.
- 981 69. Ostrowski RA, Sullivan CL, Seshadri R, Morgan GA, Pachman LM. Association of normal
982 nailfold end row loop numbers with a shorter duration of untreated disease in children with juvenile
983 dermatomyositis. *Arthritis Rheum*. 2010;62(5):1533-8.
- 984 70. van den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, et al. 2013
985 classification criteria for systemic sclerosis: an American College of Rheumatology/European League
986 against Rheumatism collaborative initiative. *Arthritis Rheum*. 2013;65(11):2737-47.
- 987 71. Frank van den H, Dinesh K, Jaap F, Sindhu RJ, Murray B, Alan T, et al. 2013 classification
988 criteria for systemic sclerosis: an American college of rheumatology/European league against rheumatism
989 collaborative initiative. *Annals of the Rheumatic Diseases*. 2013;72(11):1747.
- 990 72. Smith V, Vanhaecke A, Herrick AL, Distler O, Guerra MG, Denton CP, et al. Fast track
991 algorithm: How to differentiate a "scleroderma pattern" from a "non-scleroderma pattern". *Autoimmun*
992 *Rev*. 2019;18(11):102394.
- 993 73. Cutolo M, Sulli A, Pizzorni C, Accardo S. Nailfold videocapillaroscopy assessment of
994 microvascular damage in systemic sclerosis. *The Journal of rheumatology*. 2000;27(1):155-60.
- 995 74. Pavlov-Dolijanović S, Damjanov N, Ostojić P, Susić G, Stojanović R, Gacić D, et al. The
996 prognostic value of nailfold capillary changes for the development of connective tissue disease in children
997 and adolescents with primary raynaud phenomenon: a follow-up study of 250 patients. *Pediatr Dermatol*.
998 2006;23(5):437-42.

- 999 75. Ingegnoli F, Herrick AL, Schioppo T, Bartoli F, Ughi N, Pauling JD, et al. Reporting items for
1000 capillaroscopy in clinical research on musculoskeletal diseases: a systematic review and international
1001 Delphi consensus. *Rheumatology (Oxford)*. 2021;60(3):1410-8.
- 1002 76. Smith V. When and how to perform capillaroscopy. *Atlas of capillaroscopy in rheumatic*
1003 *diseases*; Elsevier; 2010. p. 33-42.
- 1004 77. Lambova S, Hermann W, Müller-Ladner U. Capillaroscopic Pattern at the Toes of Systemic
1005 Sclerosis Patients: Does It "Tell" More Than Those of Fingers? *JCR: Journal of Clinical Rheumatology*.
1006 2011;17(6).
- 1007 78. Dinsdale G, Roberts C, Moore T, Manning J, Berks M, Allen J, et al. Nailfold capillaroscopy—
1008 how many fingers should be examined to detect abnormality? *Rheumatology*. 2019;58(2):284-8.
- 1009 79. Tama A, Mengko TR, Zakaria H, editors. Nailfold capillaroscopy image processing for
1010 morphological parameters measurement. 2015 4th International Conference on Instrumentation,
1011 Communications, Information Technology, and Biomedical Engineering (ICICI-BME); 2015 2-3 Nov.
1012 2015.
- 1013 80. Doshi NP, Schaefer G, Zhu SY. An Evaluation of Image Enhancement Techniques for Nailfold
1014 Capillary Skeletonisation. *Procedia Computer Science*. 2015;60:1613-21.
- 1015 81. Al-Gindy A, Al-Ahmad H, Qahwaji R, Tawfik A, editors. A novel blind image watermarking
1016 technique for colour RGB images in the DCT domain using green channel. 2008 Mosharaka International
1017 Conference on Communications, Computers and Applications; 2008 8-10 Aug. 2008.
- 1018 82. Li L-J. Semantic image understanding: from the web, in large scale, with real-world challenging
1019 data: Stanford University, USA; 2011.
- 1020 83. Rodriguez-Reyna TS, Bertolazzi C, Vargas-Guerrero A, Gutiérrez M, Hernández-Molina G,
1021 Audisio M, et al. Can nailfold videocapillaroscopy images be interpreted reliably by different observers?
1022 Results of an inter-reader and intra-reader exercise among rheumatologists with different experience in
1023 this field. *Clinical Rheumatology*. 2019;38(1):205-10.
- 1024 84. Sáez-Comet L, Fanlo-Mateo P, Gracia-Tello B, Antonio Todolí Parra J, Freire-Dapena M,
1025 Espinosa-Garriga G, et al. Nailfold capillaroscopy in the Spanish Group of Systemic Autoimmune
1026 Diseases (GEAS). Results of an electronic survey. *Medicina Clínica (English Edition)*. 2020;155(11):509-
1027 10.
- 1028 85. Gracia Tello BC, Ramos Ibañez E, Saez Comet L, Guillén Del Castillo A, Simeón Aznar CP,
1029 Selva-O'Callaghan A, et al. External clinical validation of automated software to identify structural
1030 abnormalities and microhaemorrhages in nailfold videocapillaroscopy images. *Clin Exp Rheumatol*.
1031 2023;41(8):1605-11.
- 1032 86. Ng S-A, Tan WH, Saffari SE, Low AHL. Evaluation of Nailfold Capillaroscopy Online Training
1033 Using the Fast Track Algorithm. *The Journal of rheumatology*. 2023;50(3):368-72.
- 1034 87. Sue-Ann N, Wen Hao T, Saffari SE, Low AHL. Evaluation of Nailfold Capillaroscopy Online
1035 Training Using the Fast Track Algorithm. *Journal of Rheumatology*. 2023;50(3):368-72.
- 1036 88. Neubauer-Geryk J, Hoffmann M, Wielicka M, Piec K, Kozera G, Brzeziński M, et al. Current
1037 methods for the assessment of skin microcirculation: Part 1. *Advances in Dermatology and*
1038 *Allergology/Postępy Dermatologii i Alergologii*. 2019;36(3):247-54.
- 1039 89. Cutolo M, Melsens K, Herrick AL, Foeldvari I, Deschepper E, De Keyser F, et al. Reliability of
1040 simple capillaroscopic definitions in describing capillary morphology in rheumatic diseases.
1041 *Rheumatology*. 2018;57(4):757-9.
- 1042 90. Karbalaie A, Emrani Z, Fatemi A, Etehadtavakol M, Erlandsson B-E. Practical issues in assessing
1043 nailfold capillaroscopic images: a summary. *Clinical rheumatology*. 2019;38(9):2343-54.
- 1044 91. Mannarino E, Pasqualini L, Fedeli F, Scricciolo V, Innocente S. Nailfold Capillaroscopy in the
1045 Screening and Diagnosis of Raynaud's Phenomenon. *Angiology*. 1994;45(1):37-42.
- 1046 92. Avouac J, Fransén J, Walker UA, Riccieri V, Smith V, Muller C, et al. Preliminary criteria for the
1047 very early diagnosis of systemic sclerosis: results of a Delphi Consensus Study from EULAR
1048 Scleroderma Trials and Research Group. *Ann Rheum Dis*. 2011;70(3):476-81.

- 1049 93. Avouac J, Lepri G, Smith V, Toniolo E, Hurabielle C, Vallet A, et al. Sequential nailfold
1050 videocapillaroscopy examinations have responsiveness to detect organ progression in systemic sclerosis.
1051 *Semin Arthritis Rheum.* 2017;47(1):86-94.
- 1052 94. Vanhaecke A, Cutolo M, Distler O, Riccieri V, Allanore Y, Denton CP, et al. Nailfold
1053 capillaroscopy in SSc: innocent bystander or promising biomarker for novel severe organ
1054 involvement/progression? *Rheumatology (Oxford).* 2022;61(11):4384-96.
- 1055 95. Emrani Z, Karbalaie A, Fatemi A, Etehadtavakol M, Erlandsson BE. Capillary density: An
1056 important parameter in nailfold capillaroscopy. *Microvasc Res.* 2017;109:7-18.
- 1057 96. Mihai C, Smith V, Dobrota R, Gheorghiu AM, Cutolo M, Distler O. The emerging application of
1058 semi-quantitative and quantitative capillaroscopy in systemic sclerosis. *Microvasc Res.* 2018;118:113-20.
- 1059 97. Vanessa S, Filip De K, Carmen P, Jens TVP, Saskia D, Alberto S, et al. Nailfold capillaroscopy
1060 for day-to-day clinical use: construction of a simple scoring modality as a clinical prognostic index for
1061 digital trophic lesions. *Annals of the Rheumatic Diseases.* 2011;70(1):180.
- 1062 98. Cutolo M, Herrick AL, Distler O, Becker MO, Beltran E, Carpentier P, et al. Nailfold
1063 Videocapillaroscopic Features and Other Clinical Risk Factors for Digital Ulcers in Systemic Sclerosis: A
1064 Multicenter, Prospective Cohort Study. *Arthritis Rheumatol.* 2016;68(10):2527-39.
- 1065 99. Karbalaie A, Abtahi F, Fatemi A, Etehadtavakol M, Emrani Z, Erlandsson B-E. Elliptical broken
1066 line method for calculating capillary density in nailfold capillaroscopy: Proposal and evaluation.
1067 *Microvascular Research.* 2017;113:1-8.
- 1068 100. Maldonado G, Chacko A, Lichtenberg R, Ionescu M, Rios C. Nailfold capillaroscopy in diabetes
1069 mellitus: a case of neo-angiogenesis after achieving normoglycemia. *Oxford Medical Case Reports.*
1070 2022;2022(9).
- 1071 101. Boulon C, Devos S, Mangin M, Decamps-Le Chevoir J, Senet P, Lazareth I, et al. Reproducibility
1072 of capillaroscopic classifications of systemic sclerosis: results from the SCLEROCAP study.
1073 *Rheumatology.* 2017;56(10):1713-20.
- 1074 102. Redisch W, Messina EJ, Hughes G, McEwen C. Capillaroscopic observations in rheumatic
1075 diseases. *Ann Rheum Dis.* 1970;29(3):244-53.
- 1076 103. Tello BG, Ibáñez ER, Mateo PF, Cómet LS, Robles EM, Blanco JJR, et al. The challenge of
1077 comprehensive nailfold videocapillaroscopy practice: a further contribution. *Clinical and Experimental*
1078 *Rheumatology.* 2022;40(10):1926-32.
- 1079 104. Cafaro G, Venerito V, Valentini V, Bursi R, Perricone C, Gerli R, et al. COMBINED
1080 NAILFOLD CAPILLAROSCOPY AND ULTRASONOGRAPHY OF THE NAIL-ENTHESIS
1081 COMPLEX TO DISCRIMINATE PSORIATIC DISEASE FROM RHEUMATOID ARTHRITIS
1082 PATIENTS AND HEALTHY SUBJECTS. *Annals of the Rheumatic Diseases.* 2023;82:218-9.
- 1083 105. Nam K, Mendoza FA, Wessner CE, Allawh TC, Forsberg F. Ultrasound quantitative assessment
1084 of ventral finger microvasculopathy in systemic sclerosis with Raynaud's phenomena: a comparative
1085 study. *RMD Open.* 2023;9(1).
- 1086 106. Lüders S, Friedrich S, Ohrndorf S, Glimm AM, Burmester GR, Riemekasten G, et al. Detection
1087 of severe digital vasculopathy in systemic sclerosis by colour Doppler sonography is associated with
1088 digital ulcers. *Rheumatology (Oxford).* 2017;56(11):1865-73.
- 1089 107. Yakimov BP, Gurfinkel YI, Davydov DA, Allenova AS, Budylin GS, Vasiliev VY, et al.
1090 Pericapillary Edema Assessment by Means of the Nailfold Capillaroscopy and Laser Scanning
1091 Microscopy. *Diagnostics (Basel).* 2020;10(12).
- 1092 108. Nitkunanantharajah S, Haedicke K, Moore TL, Manning JB, Dinsdale G, Berks M, et al. How
1093 well does deep learning differentiate between optoacoustic and optical nailfold capillaroscopy images
1094 from patients with systemic sclerosis versus those from healthy controls? *Rheumatology (United*
1095 *Kingdom).* 2020;59:ii79.
- 1096 109. Nitkunanantharajah S, Haedicke K, Moore TB, Manning JB, Dinsdale G, Berks M, et al. Three-
1097 dimensional optoacoustic imaging of nailfold capillaries in systemic sclerosis and its potential for disease
1098 differentiation using deep learning. *Sci Rep.* 2020;10(1):16444.

1099