



Regular article

Skin temperature of the foot: Reliability of infrared image analysis based in the angiosome concept



Adérito Seixas^{a,b,*}, Joana Azevedo^a, Iuri Pimenta^a, Kurt Ammer^c, Rui Carvalho^d, João Paulo Vilas-Boas^e, Joaquim Mendes^f, Ricardo Vardasca^{c,f}

^a Escola Superior de Saúde, Universidade Fernando Pessoa, Porto, Portugal

^b LABIOMEPE, INEGI-LAETA, Faculdade de Desporto, Universidade do Porto, Porto, Portugal

^c Medical Imaging Research Unit, University of South Wales, Pontypridd, United Kingdom

^d Hospital Santo António, Centro Hospitalar do Porto, E.P.E., Porto, Portugal

^e LABIOMEPE, CIFI2D, Faculty of Sports, University of Porto, Porto, Portugal

^f LABIOMEPE, INEGI-LAETA, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal

ARTICLE INFO

Keywords:

Angiosomes

Foot

Interrater reliability

Intrarater reliability

Skin temperature

ABSTRACT

Objective: Studies reporting the reliability of image analysis when assessing skin temperature of the foot are scarce. The aim of this study was to assess the interrater and intrarater reliability of the analysis of foot skin temperature based on the angiosome concept and the association between skin temperature differences and the differences in size of the ROIs.

Methodology: Thermograms from 26 feet were analysed by two independent assessors and each assessor analysed the same images on different occasions. Mean temperature values of each of the six ROIs were extracted for analysis. Relative reliability was assessed by Intraclass Correlation Coefficient (ICC) measures and absolute reliability was assessed using Bland and Altman agreement measures and standard error of measurement (SEM). The Spearman correlation coefficient was used to assess the association between the skin temperature differences and the differences in size of the ROIs in the interrater and intrarater analysis.

Results: The ICC values evidenced excellent interrater and intrarater reliability with the 95% confidence intervals (CI) ranging between 0.962 and 1.000 and the SEM ranged between 0.00 °C and 0.36 °C. The mean absolute difference (bias) between the measurements ranged between 0.002 °C and 0.117 °C and small to moderate associations between the differences in skin temperature and the difference in the number of pixels were identified.

Conclusion: The excellent interrater and intrarater reliability measures suggest that the methodology of analysis was reliable and may be used in research and clinical settings. Although statistical significant associations between the skin temperature differences and the differences in size of the ROIs were found, the magnitude of the skin temperature differences between assessments and between assessors (0.02–0.17 °C) is not clinically relevant.

1. Introduction

The human body is a very efficient thermal system and the skin is the interface between the body and the environment. The skin, being a highly efficient radiator, radiates energy similarly to a black-body and is a key factor in thermoregulation [1,2]. Thermal imaging, unlike other imaging modalities, provides real-time physiological information on skin temperature distribution through the recording of thermograms [3].

Several factors can affect the accuracy, precision and responsiveness

of thermal imaging measurements, such as the surface being imaged, the camera system, patient preparation, patient position, environment conditions, image processing, image analysis, image exchange and image presentation [4]. Temperature measurements from thermal images are based on the definition of regions of interest (ROIs) but the shape, size and placements of the ROIs are not defined similarly by researchers, even when analysing the same body region. This is perfectly acceptable as the definition of the ROIs is dependant of the goal of the study, but the reproducibility of the adopted methodology should be tested and reported. Diagnostic accuracy of thermal imaging is

* Corresponding author at: Universidade Fernando Pessoa, Praça 9 de Abril 349, 4249-004 Porto, Portugal.
E-mail address: aderito@ufp.edu.pt (A. Seixas).

influenced by the shape and size of ROIs [e.g. 5]. Large variations in the size of ROIs leads to significant differences in temperature readings [6] and individual errors of temperature measurements, even in small ROIs over small finger joints, may be as high as 2.35 °C in newly trained thermographers [7].

Skin temperature is an important factor associated with foot complications. An increase in skin temperature, when compared with the contralateral limb, can predict neuropathic foot ulceration [8]. The International Working Group on the Diabetic Foot has suggested that monitoring skin temperature can reduce the incidence of recurrent plantar ulcers [9]. An increase in foot skin temperature has been observed in diabetic patients with neuropathy [10] and infection [11], and lower values of skin temperature have been documented in patients with diabetes and peripheral artery disease [12], suggesting that foot skin temperature assessment has a role in clinical practice, increasing the management quality of patients at risk to develop foot complications.

Research has been published on the topic of skin temperature analysis of the foot, both with automatic and manual analysis. An automated analysis [e.g. 13,14,15] is not available in most clinical setting realities and manual analysis is often performed [12,16–18]. As previously discussed, the placement of ROIs in thermograms contributes to uncertainty in measurements, however, the number of studies assessing the reliability of the analysis of foot temperature is very limited. Recently Silva, Castro, Carvalho, Chaves, Ruela and Iunes [18] have investigated inter and intrarater reliability of the analysis of thermograms but important information regarding the analysis is missing. The procedure for temperature extraction was not described and it is not clear if the measurements were based in a single pixel or in an area of pixels because this information was not reported.

An angiosome is a three-dimensional territory supplied by specific arteries and drained by specific veins [19]. The foot and ankle are supplied mainly by three arteries, the anterior tibial artery (ATA), the posterior tibial artery (PTA) and the peroneal artery (PA). The dorsum of the foot is supplied by the ATA, the lateral border of the ankle and heel are supplied by the PA and the remaining parts of the sole of the foot are supplied by three branches of the PTA. The medial calcaneal artery (MCA) supplies the medial aspect of the heel, the medial plantar artery (MPA) supplies the instep and the lateral plantar artery (LPA) supplies the lateral midfoot and forefoot [20–23]. The use of the angiosome concept to define the regions of interest is becoming increasingly popular but, to date, the reliability of the placement of ROIs based in this concept has not been assessed. Therefore, the aim of this study was to assess the interrater and intrarater reliability of the analysis of

foot skin temperature based on the angiosome concept and the association between skin temperature differences and the differences in size of the ROIs.

2. Methodology

This study was conducted with data from a larger prospective cohort study (ClinicalTrials.gov Identifier: NCT03254095), approved by the ethical committee of Centro Hospitalar do Porto.

2.1. Sample size requirements

Minimum sample size was calculated with the formula by Bonett [24]. Considering two raters, a confidence interval width of 0.2, a desired intraclass correlation coefficient of 0.9 and a significance level of 0.05 the minimum sample size was 20 feet.

2.2. Skin temperature assessment

Skin temperature measurements were performed in the morning, away from airflow and infrared radiation sources, after a 10-minute acclimation period in a room with controlled ambient temperature (23.3 ± 0.6 °C) and relative humidity ($54.4 \pm 5.5\%$). An infrared camera (FLIR Systems, E60, Wilsonville, OR, USA), which has a sensor array size of 320x240, noise equivalent temperature difference (NETD) of 50 mK at 30 °C and $\pm 2\%$ of repeatability of the overall reading with emissivity set to 0.98 was used to acquire images of the plantar and dorsal aspects of the feet. The camera was always turned on at least 40 min before the first assessment of the day to allow sensor stabilization. The camera was positioned perpendicular to the feet at 1 m distance for image acquisition. The patients were seated, with knees extended and legs supported by a chair while acquiring plantar thermograms, and when acquiring dorsal thermograms patients were seated with knees flexed. All thermograms were acquired by the same researcher.

2.3. Reliability assessment

Thermograms from 13 diabetic foot patients (26 feet) were analysed with FLIR ResearchIR Max software (FLIR Systems, version 4.30.0.69) by two blind assessors. Both assessors had less than one year of experience analysing thermal images but were previously instructed about the angiosome concept and respective ROIs corresponding to the ATA, PA, MCA, LPA, MPA. The assessors were asked to evaluate all the

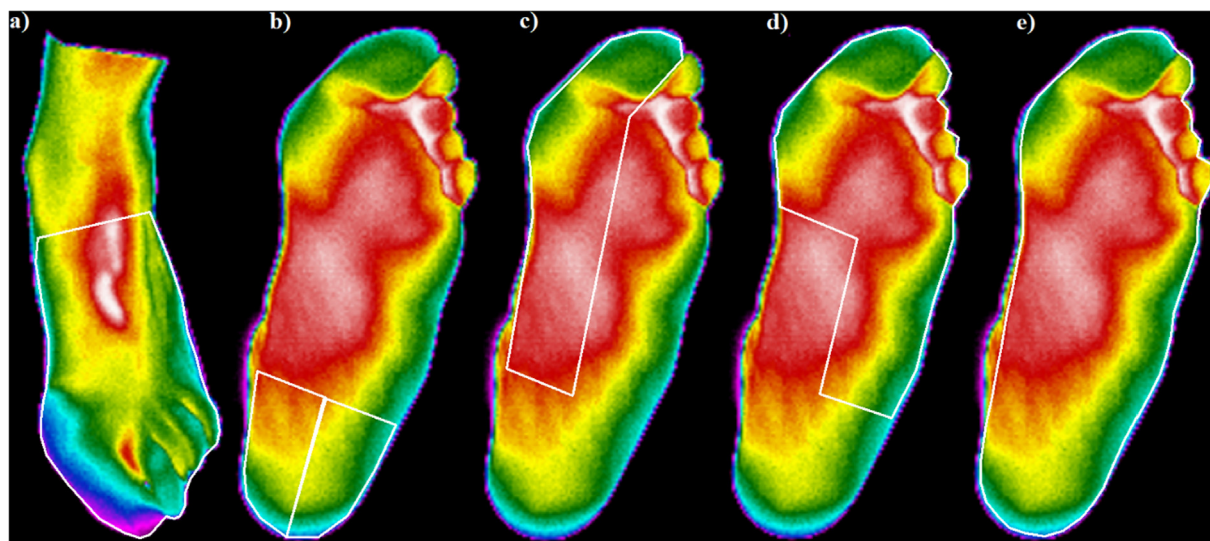


Fig. 1. Regions of interest corresponding – from left to right – to (a) the ATA, (b) the MCA and PA, (c) the MPA, (d) the LPA and (e) the entire plantar surface.

thermal images independently and were asked to manually draw 6 ROIs: ATA, PA, MCA, LPA, MPA and the sole of the foot (Fig. 1) and to extract the mean temperature values from each ROI. The ROIs were drawn according to Nagase et al. [22] for the PA, MCA, LPA and MPA, and according to Alexandrescu and Triffaux [23] for the ATA.

To assess interrater reliability the extracted temperature values from the different ROIs from both assessors were compared and to assess intrarater reliability, three weeks after the first analysis, the same images were analysed by both assessors. All analyses were performed independently in different computers and no information was exchanged.

The analyses were performed in a blind way. For interrater reliability the assessors were not aware that the same images were assessed by two different persons and for the intrarater reliability the images were randomized and were given to assessors as if they were from different patients. The randomization scheme was generated by using a web platform (<http://www.randomization.com>).

Relative reliability was assessed by Intraclass Correlation Coefficient (ICC) measures and absolute reliability was assessed using Bland and Altman agreement measures and standard error of measurement (SEM).

2.4. Data analysis

Data analysis was performed using Statistical Package for the Social Sciences (SPSS Statistics, IBM, version 25) and GraphPad Prism 7 (Graphpad Prism Software Inc., San Diego, USA). The inter and intrarater reliability was assessed calculating the ICC and respective 95% confidence intervals. ICC values were interpreted according to [25], values below 0.5 represent poor reliability, values between 0.5 and 0.75 represent moderate reliability, values between 0.75 and 0.90 represent good reliability and values above 0.90 represent excellent reliability. When assessing interrater reliability, a two-way random-effects, absolute agreement, single measurement model was used and when assessing intrarater reliability, a two-way mixed-effects, absolute agreement, single measurement model was used. The interrater and intrarater agreement was also assessed using the technique described by Bland and Altman [26], assessing agreement and limits of agreement (average \pm 2 standard deviation of differences). The SEM, which was calculated as $SEM = \text{standard deviation (SD)} \times \sqrt{(1 - ICC)}$ [25], was computed. The association between the differences in skin temperature and the difference in the number of pixels between the first and the second assessment and between the two raters [27] was assessed with the Spearman correlation coefficient. In all statistical analysis a two-sided p -value of < 0.05 identified significant differences and 95% confidence intervals (CI) were also reported.

3. Results

3.1. Reliability of infrared imaging analysis

Thermal images from 26 feet were analysed. The results for interrater and intrarater reliability are presented in Table 1.

The results suggest excellent interrater and intrarater reliability in all ROIs. In the interrater reliability analysis, the lower bound of the 95% CI ranged between 0.962 and 1.000 and the SEM ranged between 0.00 and 0.36 °C. In the intrarater reliability analysis, the lower bound of the 95% CI ranged between 0.985 and 1.000 and the SEM ranged between 0.09 and 0.29 °C. All ICC values from the interrater and intrarater reliability analysis were statistically significant ($p < 0.001$).

The graphical representation of the agreement between temperature measurements of assessors 1 and 2 can be seen in Fig. 2. For all the plots except the plots for the MCA and MPA (Fig. 2(b) and (d), respectively), 95% or more of the differences between assessors were within the limits of agreement (LOA). Only 92.3% of the differences in the MCA and 91.1% in the MPA were within the LOA.

The agreement plots between the temperature measurements of the first and second assessments can be seen in Fig. 3. For all the plots except for the LPA, represented in Fig. 2(e), 95% or more of the differences between assessors were within the limits of agreement (LOA). Only 91.1% of the differences in the LPA were within the LOA.

The mean difference (bias) between the measurements of both assessors was low, with a maximum value of -0.117 °C (LOA: -0.499 to 0.264) in the ATA and a minimum value of 0.017 °C (LOA: -0.155 to 0.190) in the MPA. The bias was lower for the measurements between assessments, with a maximum value of 0.037 °C (LOA: -0.098 to 0.171) in the LPA and a minimum value of -0.002 °C (LOA: -0.150 to 0.146) in the MPA.

3.2. Differences in ROIs size between assessments and between assessors

Considering the measurements of both assessors (56 measurements), between assessments the number of pixels varied 4.88% in the ATA, 8.93% in the MCA, 8.51% in the PA, 6.08% in the MPA, 2.65% in the LPA and 1.96% in the sole of the foot. Considering the measurements of both assessments (56 assessments), between assessors the number of pixels varied 7.29% in the ATA, 22.22% in the MCA, 21.55% in the PA, 6.87% in the MPA, 8.46 in the LPA and 1.69% in the sole of the foot. The results for each assessor and for each assessment are presented in Table 2.

3.3. Association between skin temperature differences and size of ROIs

The results for the Spearman correlation analysis are presented in Table 3. When assessing the association between the differences in skin temperature and the differences in the number of pixels between assessments, a low positive association was found in the PA and sole of the foot and a moderate positive association was found in the MPA. Between assessors, low positive associations were found in the MCA, PA and LPA.

4. Discussion

Numerous sources of bias when using thermal imaging have been identified, such as the imaged surface, the camera system, patient preparation, patient position, environment conditions, image processing, image exchange, image presentation and finally, image analysis [4]. This study demonstrated that the assessment of skin temperature of the foot according to the angiosome concept is a reliable approach. The 95% CI have indicated excellent intrarater and interrater reliability for all regions of interest and the bias between assessors and between assessments was low. It has also been demonstrated that the variation in the size of the ROIs, when considering both assessors (maximum variation of 8.93% in the MCA) and both assessments (maximum variation of 22.22% in the MCA) was small. Small to moderate associations between the differences in skin temperature and the differences in the number of pixels were found.

Considering the number of published articles using thermal imaging, few reported the reliability of image analysis [e.g. 28,29]. Recently, Silva, Castro, Carvalho, Chaves, Ruela and Lunes [18] reported excellent interrater and intrarater reliability of foot skin temperature analysis but given the lack of information regarding the ROIs, the skin temperature value that was extracted to be analysed (e.g. mean, minimum, maximum) and the model of ICC that was used in the analysis, caution in the interpretation of the results is recommended. Moreover, considering the information provided, the authors used point measurements which may be meaningless considering the extension of the foot and the ICC values may have been overestimated because the definition selected by the authors for ICC calculation was “consistency” and not “absolute agreement” as recommended [30]. We defined the ROIs based in the angiosomes, an established and useful concept used by clinicians [20] and calculated ICC values using a two-way random-

Table 1

Skin temperature values (mean ± standard deviation) from both assessors and both assessments in each ROI. Interrater and intrarater reliability of infrared imaging analysis (ICC, 95% CI, SEM).

	ROIs	Skin Temperature (°C)		Interrater Reliability			Intrarater Reliability			
		Assessor 1	Assessor 2	ICC	95% CI	SEM	ICC	95% CI	SEM	
Assessment 1	ATA	28.8 ± 1.6	28.9 ± 1.6	0.991	0.962 – 0.997	0.30	Rater 1	0.998	0.996 – 0.999	0.14
	MCA	27.5 ± 1.8	27.6 ± 1.7	0.992	0.981 – 0.996	0.31		0.997	0.992 – 0.998	0.19
	PA	27.3 ± 1.8	27.3 ± 1.7	0.989	0.976 – 0.995	0.36		0.993	0.985 – 0.997	0.29
	MPA	28.0 ± 1.4	28.0 ± 1.4	0.998	0.997 – 0.999	0.13		0.998	0.996 – 0.999	0.13
	LPA	27.6 ± 1.6	27.5 ± 1.6	0.997	0.994 – 0.999	0.18		0.999	0.998 – 1.000	0.10
Assessment 2	Sole	27.7 ± 1.5	27.7 ± 1.4	0.999	0.998 – 1.000	0.09	0.999	0.998 – 1.000	0.09	
	ATA	28.8 ± 1.6	28.9 ± 1.6	0.989	0.970 – 0.995	0.33	Rater 2	0.997	0.993 – 0.999	0.17
	MCA	27.5 ± 1.8	27.5 ± 1.7	0.997	0.993 – 0.999	0.19		0.999	0.996 – 0.999	0.11
	PA	27.3 ± 1.7	27.3 ± 1.7	0.996	0.992 – 0.998	0.22		0.998	0.995 – 0.999	0.15
	MPA	28.1 ± 1.4	28.0 ± 1.4	0.998	0.995 – 0.999	0.13		0.999	0.998 – 1.000	0.09
LPA	27.5 ± 1.6	27.5 ± 1.6	0.997	0.992 – 0.999	0.18	0.999		0.994 – 1.000	0.10	
	Sole	27.7 ± 1.4	27.7 ± 1.4	1.000	0.999 – 1.000	0.00	0.999	0.998 – 1.000	0.09	

effects, absolute agreement, single measurement model for interrater reliability, and a two-way mixed-effects, absolute agreement, single measurement model for intrarater reliability [30].

Other authors have also reported interrater and intrarater reliability of image analysis in other body regions. Rodrigues-Bigaton, Dibai Filho, de Souza Costa, Packer and de Castro [31] reported ICC values ranging between 0.841 and 0.874 in the region of the temporomandibular joint, however, the authors have used point measurements and only reported relative reliability measures (ICC) without confidence intervals. Costa, Dibai Filho, Packer and Rodrigues-Bigaton [32] have reported excellent intrarater and interrater reliability of image analysis of masticatory and upper trapezius muscles, only using relative reliability measures (ICC). The authors analysed the images using two techniques, muscle length (using a line ROI) and muscle central portion (point measurement) and based in the 95% CI the ICC values ranged between 0.992 and 0.999 when using the line ROI and between 0.778 and 0.999 using the point measurement. Another study [33] assessed the interrater and intrarater reliability of three methodologies of analysing skin temperature over the upper trapezius muscle (point, line and area) and found that the reliability when using the mean was higher than when using minimum or maximum values. The ICC values for mean temperature were similar

for the three methodologies, with ICC values ranging between 0.915 and 0.996. The ICC values were slightly better when using the point methodology. It is important to note that both Costa, Dibai Filho, Packer and Rodrigues-Bigaton [32] and Dibai-Filho, Guirro, Ferreira, Brandino, Vaz and Guirro [33] used styrofoam markers to delineate the ROIs, which may have contributed to increase the reliability of the analysis. One study [29] assessed the interrater and intrarater reliability of analysing thermal images of the breasts and reported ICC values ranging between 0.948 and 0.999. Interestingly, ICC values using the minimum temperature value were slightly higher than the ICC values when using the mean. However, when using the maximum value, the reliability was lower but still within the cut-off to be considered excellent reliability. These authors also reported drawing a horizontal line between the armpits as a reference to delineate the ROIs, which may have contributed the high reliability of the analysis. Finally, Ammer [34] assessed the intrarater reliability of analysing three ROIs in the upper limb (elbow, upper arm and forearm) and over the ankle, the dorsal and the plantar foot. The ICC values ranged between -0.42 and 0.99 in the upper limb, with higher reliability in the analysis of the forearm, and between -0.59 and 0.97 in the lower limb, with higher values in the right and left dorsal foot. This study is of interest because

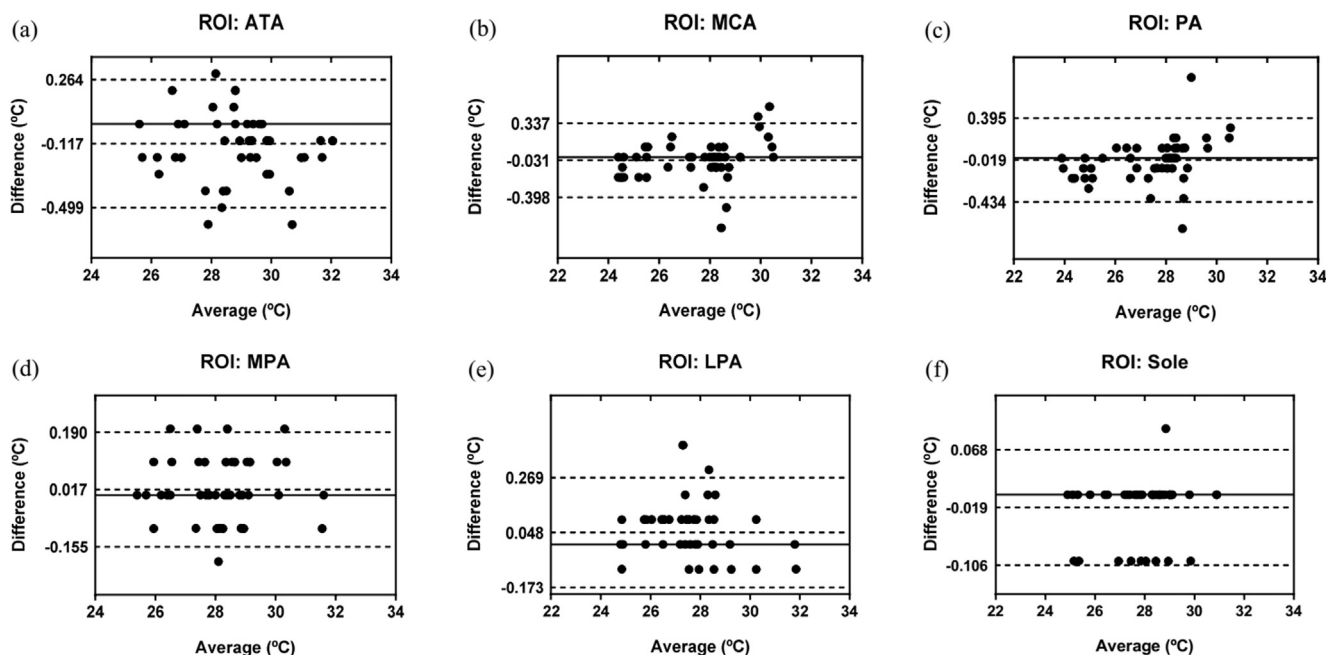


Fig. 2. Bland-Altman plots (bias and LOA) of the difference of temperature measurements measured by assessors 1 and 2, against the mean of the pair in each ROI.

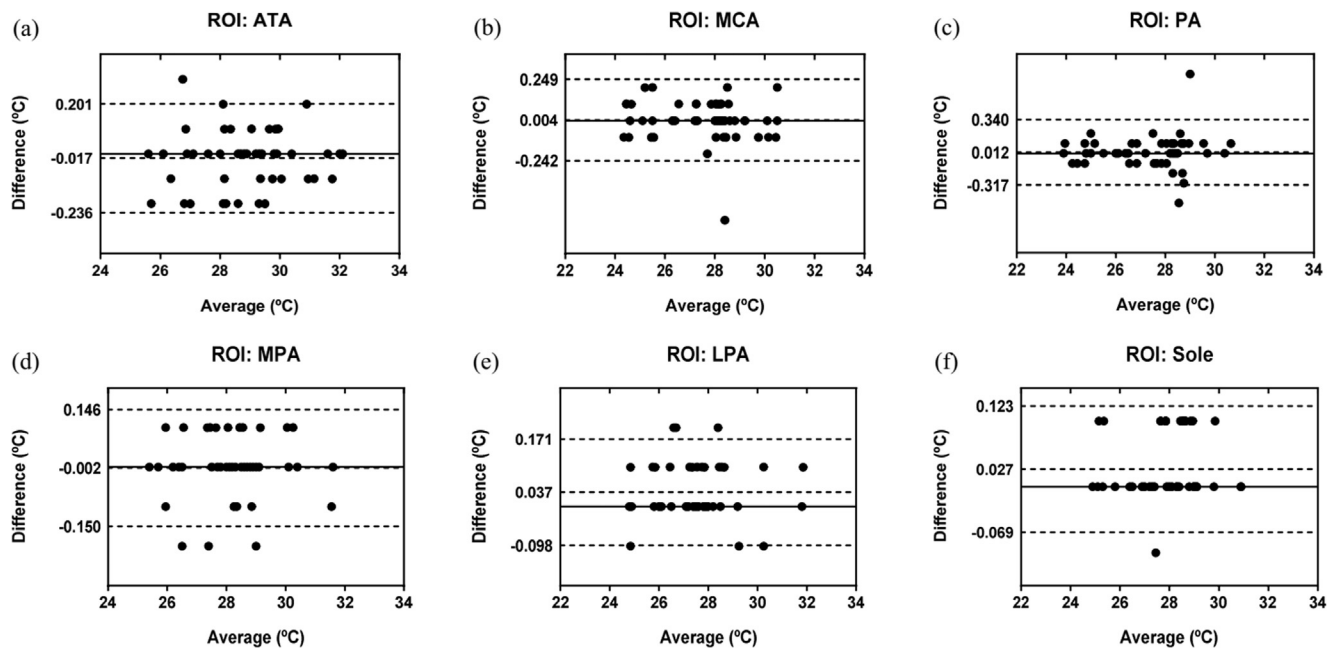


Fig. 3. Bland-Altman plots (bias and LOA) of the difference of temperature measurements from both assessments, against the mean of the pair in each ROI.

it assessed similar ROIs to our study (dorsal and plantar foot). In the right and left dorsal foot the ICC values were 0.73 (95% CI: -0.03 to 0.97) and 0.76 (95% CI: -0.19 to 0.97) and in the right and left plantar foot the ICC values were 0.26 (95% CI: -0.59 to 0.88) and 0.70 (95% CI: -0.25 to 0.96). In our study the intrarater ICC values for the ATA, in the dorsal foot, for the first and second assessor were 0.998 (95% CI: 0.996–0.999) and 0.997 (95% CI: 0.993–0.999) and for the plantar foot, for the first and second assessor, were 0.999 (95% CI: 0.998–1.000) and 0.999 (95% CI: 0.998–1.000). These differences might be related to the fact that Ammer included five newly trained assessors and only one image of each ROI was assessed.

Only three studies reported absolute reliability measures, i.e. SEM, minimum detectable change (MDC) or Bland and Altman’s analysis. Tan, Sanjay and Morgan [28] reported the 95% LOA for interrater and intrarater reliability of six ocular skin temperature indices, and the results demonstrate that the LOA were wider in the intrarater analysis. However, the results obtained in this study suggest that the bias and 95% LOA were higher in the interrater reliability analysis. Silva et al. (2018), Dibai-Filho, Guirro, Ferreira, Brandino, Vaz and Guirro [33] and de Jesus Guirro, Vaz, das Neves, Dibai-Filho, Carrara and de Oliveira Guirro [29] reported the SEM and MDC. The MDC is a measure

Table 3

Association between differences (mean ± standard deviation) in skin temperature and number of pixels in intrarater and interrater analysis. Spearman correlation coefficient (r_s) with respective 95% CI. Note: * $p < 0.05$; ** $p < 0.01$.

ROIs	Skin Temperature	Number of Pixels	r_s (95% CI)
Difference Between Assessments			
ATA	0.08 ± 0.08	321 ± 281	0.211 (-0.084 to 0.520)
MCA	0.08 ± 0.10	78 ± 75	0.235 (-0.236 to 0.610)
PA	0.10 ± 0.13	76 ± 75	0.353 (-0.069 to 0.697) [*]
MPA	0.04 ± 0.06	203 ± 191	0.617 (0.344–0.801) ^{**}
LPA	0.05 ± 0.06	147 ± 108	-0.004 (-0.279 to 0.253)
Sole	0.03 ± 0.05	176 ± 121	0.300 (0.021–0.553) [*]
Difference Between Assessors			
ATA	0.17 ± 0.15	483 ± 330	-0.095 (-0.339 to 0.185)
MCA	0.12 ± 0.15	197 ± 113	0.354 (-0.043 to 0.637) [*]
PA	0.14 ± 0.16	196 ± 113	0.391 (0.028–0.649) ^{**}
MPA	0.06 ± 0.07	219 ± 164	0.271 (-0.027 to 0.494)
LPA	0.08 ± 0.09	474 ± 244	0.456 (0.196–0.628) ^{**}
Sole	0.02 ± 0.04	153 ± 105	-0.017 (-0.284 to 0.262)

Table 2

Number of pixels in each ROI (mean ± standard deviation) from both assessments and both assessors and mean variation of the number of pixels (% , mean ± standard deviation) in each ROI.

	ROIs	Number of Pixels		Intrarater Difference (%)		Interrater Difference (%)	
		Assessment 1	Assessment 2			Assessment 1	Assessment 2
Rater 1	ATA	6547 ± 1018	6595 ± 926	4.30 ± 3.96		7.43 ± 5.23	
	MCA	763 ± 125	836 ± 139	11.88 ± 11.15		28.69 ± 12.90	
	PA	768 ± 96	846 ± 134	11.69 ± 10.44		27.53 ± 10.65	
	MPA	3377 ± 606	3354 ± 591	8.64 ± 6.33		8.09 ± 5.54	
	LPA	5769 ± 808	5780 ± 819	2.50 ± 1.58		9.11 ± 3.88	
	Sole	8987 ± 1262	9107 ± 1301	1.57 ± 1.28		1.81 ± 1.22	
Rater 2	ATA	6654 ± 1045	6762 ± 1147	5.45 ± 4.73		7.16 ± 4.80	
	MCA	1016 ± 138	972 ± 141	5.99 ± 4.03		15.76 ± 8.40	
	PA	1019 ± 156	985 ± 141	5.34 ± 4.29		15.57 ± 9.59	
	MPA	3211 ± 591	3275 ± 560	3.52 ± 3.20		5.64 ± 5.17	
	LPA	5265 ± 705	5340 ± 724	2.79 ± 2.20		7.81 ± 4.23	
	Sole	9039 ± 1306	9198 ± 1282	2.35 ± 1.27		1.56 ± 1.08	

used in metrology that is considered the minimal amount of change that is not likely to be related to chance variation when repeated measurements are performed. Instruments that demonstrate good stability across repeated measurements will have small MDC [35]. Reporting the MDC when the reliability of the analysis of the same images is assessed is not meaningful and this measure of absolute reliability was not considered in the present study. In the study of Silva *et al.* (2018) the SEM in the foot varied between 0.00 °C and 0.23 °C in the intrarater reliability analysis and between 0.08 °C and 0.73 °C in the interrater analysis. Our results were similar with SEM values ranging between 0.09 °C and 0.20 °C in the intrarater reliability analysis and between 0.00 °C and 0.36 °C in the interrater reliability analysis. Dibai-Filho, Guirro, Ferreira, Brandino, Vaz and Guirro [33] compared the reliability of three methodologies to analyse the skin temperature in the upper trapezius, and considering the analysis of the mean values, the SEM ranged between 0.13 °C and 0.34 °C in the intrarater reliability analysis and between 0.43 °C and 0.48 °C in the interrater reliability analysis. When using the minimum and maximum values the SEM values were higher and ranged between 0.28 °C and 1.57 °C in the intrarater reliability analysis and between 0.50 °C and 1.22 °C in the interrater reliability analysis. Moreover, although ICC values were similar between the methodologies of assessment, the SEM was globally higher when the analysis was based in the assessment of a point, compared to the analysis based in the measurement of a line and an area of pixels. In the study of de Jesus Guirro, Vaz, das Neves, Dibai-Filho, Carrara and de Oliveira Guirro [29] the SEM ranged between 0.04 °C and 0.15 °C in the intrarater analysis and between 0.06 °C and 0.28 °C in the interrater reliability analysis. SEM values were globally lower when using the minimum temperature value in the analysis, and higher when the maximum temperature values were analysed.

It appears that extracting the mean value of an area of pixels and using strategies helping to delineate the ROIs may improve relative and absolute reliability measures. However, the literature reporting the analysis of the size of the ROIs and its effect in the reliability estimates is scarce. Some authors used only point temperature assessments to calculate reliability measures [18,31], which makes this analysis inviable and only the study of de Jesus Guirro, Vaz, das Neves, Dibai-Filho, Carrara and de Oliveira Guirro [29] mentioned that the size of the ROIs had a maximum of 5% variation. In the present study, considering the measurements of both assessors, the difference in the size of the ROIs between the repeated assessments has not exceeded 8.93% in the MCA, which is a small difference. Differences were higher when considering the measures of both assessments, between the two assessors, reaching 22.22% also in the MCA. The lowest variation in the size of the ROIs was found in the foot sole with 1.96% and 1.69% in the intrarater and interrater comparisons, respectively. Comparison of these results with previously published literature was not possible but, considering the work of Ammer [6] which state that significant temperature differences occur when the number of pixels within a ROI varies by 100% or more, the differences in size of the ROIs are well below that value. The MCA was the ROI where the variations in size were higher, but the intrarater and interrater reliability was always excellent and the lowest ICC values were not in the MCA. Nonetheless, small to moderate association between the differences in skin temperature and the differences in the number of pixels were identified. However, considering the mean temperature differences between assessments (0.03–0.10 °C) and the mean temperature difference between assessors (0.02–0.17 °C) we are confident that these differences are not clinically relevant.

The Bland-Altman plots in Fig. 2(f), 3(d), (e) and (f) reveals that the difference between measurements (Y-axis) only assumed a limited number of values (3 or 4). In the interrater analysis, this only happens in Fig. 2(f) (Sole), while in other graphs the values are more scattered. In the intrarater analysis, this happens in Fig. 3(d), (e) and (f) (MPA, LPA and Sole). Looking carefully to Tables 1, 2 and 3 it is possible to note that these are the ROIs with smaller mean temperature difference between measurements, smaller percentage difference between the

number of pixels and higher agreement, therefore the differences between measurements assume a limited number of values. These findings are opposite, for instance, to the interrater analysis of the ATA, in Fig. 2(a), where the agreement is lower, with wider 95% CI and larger mean skin temperature differences.

Finally, every study has limitations. Although we believe that this manuscript may contribute to the topic of reliability in thermal imaging analysis, there are some aspects that must be pointed. Images were only assessed by two raters, with similar experience. Adding more assessors, with different levels of experience, could improve the knowledge about this topic. Moreover, the angiosome concept was used to define the ROIs, and although clinically important, this concept is not the only, or the more common, theoretical basis of defining ROIs to analyse foot skin temperature. The reliability analysis of foot skin temperature assessment with ROIs based on the most common ulceration sites would also be of interest.

5. Conclusion

This study assessed the interrater and intrarater reliability of the analysis of foot skin temperature based on the angiosome concept. The methodology of analysis was reliable as evidenced by the excellent interrater and intrarater reliability measures, supporting its use in clinical and research settings. There is an association between the difference in skin temperature and the difference in the number of pixels, that although statistically significant is not clinically relevant.

Conflict of interest

No potential conflict of interest was reported by the authors.

Acknowledgment

The authors acknowledge Fundação para a Ciência e a Tecnologia through grant LAETA - UID/EMS/50022/2013.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

- [1] E. Ring, Thermal imaging today and its relevance to diabetes, *J. Diabetes Sci. Technol.* 4 (2010) 857–862.
- [2] J.D. Hardy, The radiation of heat from the human body: III the human skin as a black-body radiator, *J. Clin. Invest.* 13 (1934) 615–629.
- [3] E. Ring, K. Ammer, Infrared thermal imaging in medicine, *Physiological Meas.* 33 (2012) R33.
- [4] K. Ammer, E. Ring, Standard procedures for infrared imaging in medicine, in: N. Diakides, E. Bronzino (Eds.), *Medical Infrared Imaging: Principles and Practices*, CRC Press, Boca Raton, FL, 2013pp. 32.31–32.14.
- [5] K. Ammer, The sensitivity of infrared imaging for diagnosing Raynaud's phenomenon and for Thoracic Outlet Syndrome is dependent on the method of temperature extraction from thermal image, *Thermology Int.* 18 (2008) 81–88.
- [6] K. Ammer, Temperature readings from thermal images are less dependent on the number of pixels of the measurement area than on variation of room temperature, *Thermology Int.* 15 (2005) 131–133.
- [7] K. Ammer, Repeatability of the standard view both dorsal hands. Results from a training course on medical infrared imaging, *Thermology Int.* 14 (2004) 99–102.
- [8] V.J. Houghton, V.M. Bower, D.C. Chant, Is an increase in skin temperature predictive of neuropathic foot ulceration in people with diabetes? A systematic review and meta-analysis, *J. Foot Ankle Res.* 6 (2013) 31.
- [9] J. van Netten, P.E. Price, L. Lavery, M. Monteiro-Soares, A. Rasmussen, Y. Jubiz, S. Bus, Prevention of foot ulcers in the at-risk patient with diabetes: a systematic review, *Diabetes/Metab. Res. Rev.* 32 (2016) 84–98.
- [10] S. Bagavathiappan, J. Philip, T. Jayakumar, B. Raj, P.N.S. Rao, M. Varalakshmi, V. Mohan, Correlation between plantar foot temperature and diabetic neuropathy: a case study by using an infrared thermal imaging technique, *J. Diabetes Sci. Technol.* 4 (2010) 1386–1392.
- [11] J. Harding, D. Wertheim, R. Williams, J. Melhuish, D. Banerjee, K. Harding, Infrared imaging in diabetic foot ulceration, *Proceedings of the 20th Annual International*

- Conference of the IEEE Engineering in Medicine and Biology Society, 1998, IEEE, 1998, pp. 916–918.
- [12] A. Seixas, K. Ammer, R. Carvalho, J.P. Vilas-Boas, R. Vardasca, J. Mendes, Skin temperature in diabetic foot patients: a study focusing on the angiosome concept, in: J.M.R.S. Tavares, R.M. Natal Jorge (Eds.), *ViplIMAGE 2017: Proceedings of the VI ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing Porto, Portugal, October 18–20, 2017*, Springer International Publishing, Cham, 2018, pp. 1035–1040.
- [13] D. Hernandez-Contreras, H. Peregrina-Barreto, J. Rangel-Magdaleno, J. Gonzalez-Bernal, L. Altamirano-Robles, A quantitative index for classification of plantar thermal changes in the diabetic foot, *Infrared Phys. Technol.* 81 (2017) 242–249.
- [14] J.J. van Netten, M. Prijs, J.G. van Baal, C. Liu, F. van der Heijden, S.A. Bus, Diagnostic values for skin temperature assessment to detect diabetes-related foot complications, *Diabetes Technol. Ther.* 16 (2014) 714–721.
- [15] J.S. Wrobel, P. Ammanath, T. Le, C. Luring, J. Wensman, G.S. Grewal, B. Najafi, R. Pop-Busui, A novel shear reduction insole effect on the thermal response to walking stress, balance, and gait, *J. Diabetes Sci. Technol.* 8 (2014) 1151–1156.
- [16] P.-C. Sun, S.-H.E. Jao, C.-K. Cheng, Assessing foot temperature using infrared thermography, *Foot & Ankle Int.* 26 (2005) 847–853.
- [17] A. Gatt, O. Falzon, K. Cassar, C. Ellul, K.P. Camilleri, J. Gauci, S. Mizzi, A. Mizzi, C. Sturgeon, L. Camilleri, Establishing differences in thermographic patterns between the various complications in diabetic foot disease, *Int. J. Endocrinol.* 2018 (2018).
- [18] N.C. Silva, H.A. Castro, L.C. Carvalho, É.C. Chaves, L.O. Ruela, D.H. Iunes, Reliability of infrared thermography images in the analysis of the plantar surface temperature in diabetes mellitus, *J. Chiropractic Med.* 17 (2018) 30–35.
- [19] G.I. Taylor, J.H. Palmer, The vascular territories (angiosomes) of the body: experimental study and clinical applications, *Br. J. Plast. Surg.* 40 (1987) 113–141.
- [20] C.E. Attinger, K.K. Evans, E. Bulan, P. Blume, P. Cooper, Angiosomes of the foot and ankle and clinical implications for limb salvage: reconstruction, incisions, and revascularization, *Plast. Reconstr. Surg.* 117 (2006) 261S–293S.
- [21] Y. Kagaya, N. Ohura, H. Suga, H. Eto, A. Takushima, K. Harii, 'Real angiosome'-assessment from peripheral tissue perfusion using tissue oxygen saturation foot-mapping in patients with critical limb ischemia, *Eur. J. Vasc. Endovasc. Surg.* 47 (2014) 433–441.
- [22] T. Nagase, H. Sanada, K. Takehara, M. Oe, S. Iizaka, Y. Ohashi, M. Oba, T. Kadowaki, G. Nakagami, Variations of plantar thermographic patterns in normal controls and non-ulcer diabetic patients: novel classification using angiosome concept, *J. Plast., Reconstr. Aesthetic Surg.* 64 (2011) 860–866.
- [23] V.-A. Alexandrescu, F. Triffaux, Ischemic Ulcer Healing: Does Appropriate Flow Reconstruction Stand for All That We Need? Wound Healing-New insights into Ancient Challenges, *InTech*, 2016, pp. 247–277.
- [24] D.G. Bonett, Sample size requirements for estimating intraclass correlations with desired precision, *Statistics Med.* 21 (2002) 1331–1335.
- [25] L.G. Portney, M.P. Watkins, *Foundations of Clinical Research: Applications to Practice*, 3rd ed., F. A. Davis Company, Philadelphia, 2015.
- [26] J.M. Bland, D. Altman, Statistical methods for assessing agreement between two methods of clinical measurement, *Lancet* 327 (1986) 307–310.
- [27] T.J. Cole, D.G. Altman, Statistics Notes: What is a percentage difference? *BMJ* 358 (2017) j3663.
- [28] L.L. Tan, S. Sanjay, P.B. Morgan, Repeatability of infrared ocular thermography in assessing healthy and dry eyes, *Contact Lens Anterior Eye* 39 (2016) 284–292.
- [29] R.R. de Jesus Guirro, M.M.O.L.L. Vaz, L.M.S. das Neves, A.V. Dibai-Filho, H.H.A. Carrara, E.C. de Oliveira Guirro, Accuracy and reliability of infrared thermography in assessment of the breasts of women affected by cancer, *J. Med. Syst.* 41 (2017) 87.
- [30] T.K. Koo, M.Y. Li, A guideline of selecting and reporting intraclass correlation coefficients for reliability research, *J. Chiropractic Med.* 15 (2016) 155–163.
- [31] D. Rodrigues-Bigaton, A.V. Dibai Filho, A.C. de Souza Costa, A.C. Packer, E.M. de Castro, Accuracy and reliability of infrared thermography in the diagnosis of arthralgia in women with temporomandibular disorder, *J. Manipulative Physiol. Ther.* 36 (2013) 253–258.
- [32] A.C. Costa, A.V. Dibai Filho, A.C. Packer, D. Rodrigues-Bigaton, Intra and inter-rater reliability of infrared image analysis of masticatory and upper trapezius muscles in women with and without temporomandibular disorder, *Braz. J. Phys. Ther.* 17 (2013) 24–31.
- [33] A.V. Dibai-Filho, E.C. Guirro, V.T. Ferreira, H.E. Brandino, M.M. Vaz, R.R. Guirro, Reliability of different methodologies of infrared image analysis of myofascial trigger points in the upper trapezius muscle, *Braz. J. Phys. Ther.* 19 (2015) 122–128.
- [34] K. Ammer, Need for standardisation of measurements in thermal imaging, *Thermography Lasers Med.* (2003) 13–17.
- [35] S.M. Haley, M.A. Fragala-Pinkham, Interpreting change scores of tests and measures used in physical therapy, *Phys. Ther.* 86 (2006) 735–743.