



Article

Segmentation Agreement and AI-Based Feature Extraction of Cutaneous Infrared Images of the Obese Abdomen after Caesarean Section: Results from a Single Training Session

Charmaine Childs ^{1,*}, Harriet Nwaizu ¹, Oana Voloaca ¹ and Alex Shenfield ²

¹ Centre for Applied Health and Social Care Research, Health Research Institute, Sheffield Hallam University, Sheffield S10 2BP, UK

² Department of Engineering and Mathematics, Sheffield Hallam University, Sheffield S1 1WB, UK; a.shenfield@shu.ac.uk

* Correspondence: c.childs@shu.ac.uk

Abstract: Background: Infrared thermography in women undergoing caesarean section has promise to identify a surgical site infection prodrome characterised by changes in cutaneous perfusion with concomitant influences on temperature distribution across the abdomen. This study was designed to compare abdominal and wound regions of interest (ROI) and feature extraction agreement between two independent users after a single training session. **Methods:** Image analysis performed manually in MATLAB with each reviewer 'blind' to results of the other. Image ROIs were annotated via pixel-level segmentation creating pixel masks at four time-points during the first 30 days after surgery. **Results:** A total of 366 matched image pairs (732 wound and abdomen labels in total) were obtained. Distribution of mask agreement using Jacquard similarity co-efficient ranged from 0.35 to 1. Good segmentation agreement (coefficient ≥ 0.7) (for mask size and shape) was observed for abdomen, but poor for wound (coefficient < 0.7). From feature extraction, wound cold spots were observed most in those who later developed wound infections. **Conclusions:** Reviewer performance, with respect to the input (image) data in the first stage of algorithm development, reveals a lack of correspondence (agreement) of the ROI indicating the need for further work to refine the characteristics of output labels (masks) before an unsupervised algorithm works effectively to learn patterns and features of the wound.

Keywords: thermography; infrared; obese; abdomen; segmentation; region of interest; image features; caesarean section; machine learning



Citation: Childs, C.; Nwaizu, H.; Voloaca, O.; Shenfield, A. Segmentation Agreement and AI-Based Feature Extraction of Cutaneous Infrared Images of the Obese Abdomen after Caesarean Section: Results from a Single Training Session. *Appl. Sci.* **2023**, *13*, 3992. <https://doi.org/10.3390/app13063992>

Academic Editor: Igor Pušnik

Received: 21 February 2023

Revised: 17 March 2023

Accepted: 17 March 2023

Published: 21 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Over the last 40 to 50 years, the rate of obesity has tripled [1]. As a global health concern, the prevalence of obesity has risen across race, age, and gender. Most pronounced in women of reproductive age [2], obesity during pregnancy is a major obstetric complication especially in view of the increased risk of a caesarean section birth. Furthermore, with high rates of surgical site infection (SSI) affecting 28–50% of obese and morbidly obese women, respectively [3,4], the interplay between obesity, caesarean section, and wound infection has become a major healthcare problem for post-operative recovery. The human and economic consequences of SSI are driving national and international organisations to tackle the problem of rising SSI prevalence across all surgical specialities [5,6].

In our recent studies of obese women giving birth by caesarean section, rather than waiting until infective complications present, we have focused on finding new ways to explore a prodromal phase of incipient wound tissue breakdown at the surgical site using long wave infrared thermography (IRT) to map the distribution of heat across the abdomen and wound site [3,7]. As changes in cutaneous temperature are predominately influenced by local skin blood flow, areas of high (and lower) temperature have been used to assess

skin perfusion [8,9]. Although thermography is recognised for its diagnostic potential in identifying locations of abnormal chemical and blood vessel activity across many medical pathologies [10], there is a gap in knowledge of the relationship between skin temperature and skin perfusion in the wider context of surgical wound complications [11].

However, with new approaches to wound assessment using IRT in women undergoing caesarean section, we have shown that wound abdomen temperature differences (WATD) predict wound complications, notably surgical site infection (SSI) as early as at day 2 and at day 7 after surgery [3], time-points that precede overt signs of infection, typically around the second week after surgery. Furthermore, low skin temperature in the region of the wound, coupled with discrete ‘cold spots’ along the incision, have good predictive capability to indicate later SSI [3,12,13]. Thus, utilising the postoperative thermal signature to identify skin ‘cold’ spots, commensurate with low cutaneous temperature and corresponding reduction in blood flow, is a potential clinical tool which may help shed light on the underlying mechanisms leading to infective and non-infective wound complications.

However, identifying one (or more) surface anatomy regions of interest (ROI) from the heat map is both labour-intensive and subjective, as is the refining and ‘fine-tuning’ of pixel level features. Problems emerge due to ‘crowding’ of objects, whereby pixel features are not easily assigned to the ‘correct’ bounding box or cluster group [14]. When overlap occurs between ROIs (such as that shown in Figure 1) [15], feature extraction from a specific ROI can be difficult and lead to misclassifications. To improve upon the reliability of our existing, manual, image post-processing and wound outcome modelling undertaken by a single, experienced researcher (CC) and reported previously [3], our overarching objective was to use the acquired images to refine and improve scope in image analysis using AI-based segmentation.

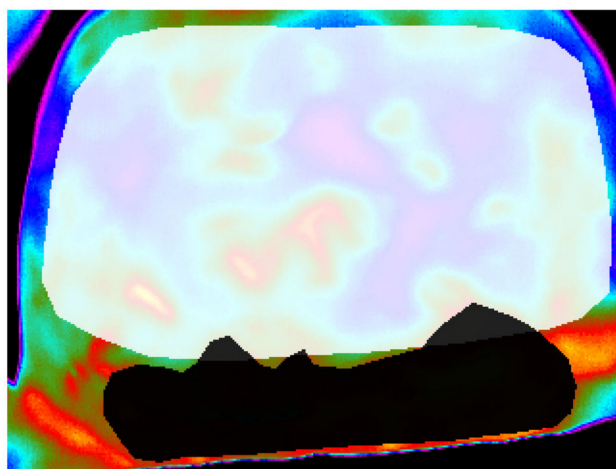


Figure 1. An illustration of overlapping ROI masks.

The aim of this study therefore was twofold: (a) to compare instance segmentation and feature extraction agreement between two new and independent users after a single training session and (b) to determine the occurrence of thermal cold regions and ‘spots’ across abdomen and wound, respectively.

2. Materials and Methods

Abdominal skin and incisional wound thermal maps previously obtained from 50 obese postpartum women who underwent caesarean section birth were analysed. The international guidelines for conducting clinical thermographic imaging were followed (<https://iactthermography.org/standards/medical-infrared-imaging/>, accessed on 15 March 2023). Thermal camera T450sc uncooled microbolometer (FLIR, Taby, Sweden) specifications were the following: detector 320×240 pixel size, Field of View (FOV) $25.0^\circ \times 19^\circ$, thermal sensitivity NETD 30 mK, spectral range 7.5–13 microns, with accuracy of $\pm 1^\circ\text{C}$

or $\pm 1\%$ of reading for limited temperature range, which in this study was physiological temperatures in the range of 30–40 °C

Before imaging commenced, emissivity was set to 0.98 and focused at a distance of 1 m from the abdominal/wound surface. Acclimation with ambient conditions was achieved by exposure of the abdomen for 15 min before imaging. This was particularly of importance for those women for whom an abdominal panniculus created a large skin fold.

Postoperative images were obtained at four nominal time-points after surgery, providing a total of 732 images (366 pairs for reference of abdomen and wound' sites). Images were obtained on day 2 (in hospital) and after discharge to the community with follow-up at days 7, 15, and day 30. Wound outcome assessment (healed, slow healing, infection) was determined at day 30, using a validated SSI wound assessment criteria [12] with confirmation from the General Practitioner regarding need for wound treatment, antibiotic prescribing and, from hospital records, the need for surgical readmission for wound management and or surgical revision.

2.1. Selection of Regions of Interest (ROIs)

Camera orientation and position, at approximately 1 metre from the body surface, focused on the umbilicus centrally and anatomical boundaries; costal margin (superiorly) to iliac crest (inferiorly).

A second IRT image was usually required to identify ROI 2 at the site of Pfannenstiel wound incision [16] where the presence of a large abdominal panniculus obscured the wound view. For wound site, on lifting of the abdominal panniculus, surface anatomy was broadly identified between iliac and pubic crest. Inferiorly, a cotton sheet was placed at the level of the pubic symphysis. On occasions (and in the absence of a large panniculus), ROI 1 and ROI 2 were visible from one single image. Areas excluded from the field of view (FOV) include skin folds of the breasts and lateral aspects of the abdomen.

2.2. Segmentation and Image Annotation

The aim of the segmentation procedure was manual identification and labelling of the abdomen and wound areas for determining the usefulness and feasibility of training an artificial intelligence-based segmentation algorithm and thus automating the identification of both abdomen (ROI 1) and wound (ROI 2) regions.

One graduate engineer and one doctoral student in biomedical sciences were appointed to review the images and to undertake image analysis. Neither had previous experience of the surgical specialty nor wound assessment. One training event consisting of two separate, online meetings were undertaken one week apart. Although both researchers attended the same session, they were not acquainted with each other and did not communicate directly during the study. Both researchers received the same information regarding the segmentation process, the details required to identify the abdomen and, separately, wound ROI. Both received digital images which matched each thermal image.

The manual segmentation procedure was performed in MATLAB (R2020a: Mathworks). The image labeller application was used to annotate each ROI label via pixel-level segmentation and creation of pixel masks for the series of 50 abdominal (Figure 1) and wound sites (Figure 2). The segmentation process was performed by two independent researchers, 'blinded' to the segmentation process of the other, and was broadly achieved in two stages to

- (i). manually segment abdomen from wound to produce two separate image masks, and
- (ii). undertake per-pixel cluster and grouping of observable objects; in this study, 'objects' are the appearance and number of areas of 'cold spots', areas appearing on abdomen and wound maps as darker regions (commensurate with lowest temperature values (°C) in the extracted image).

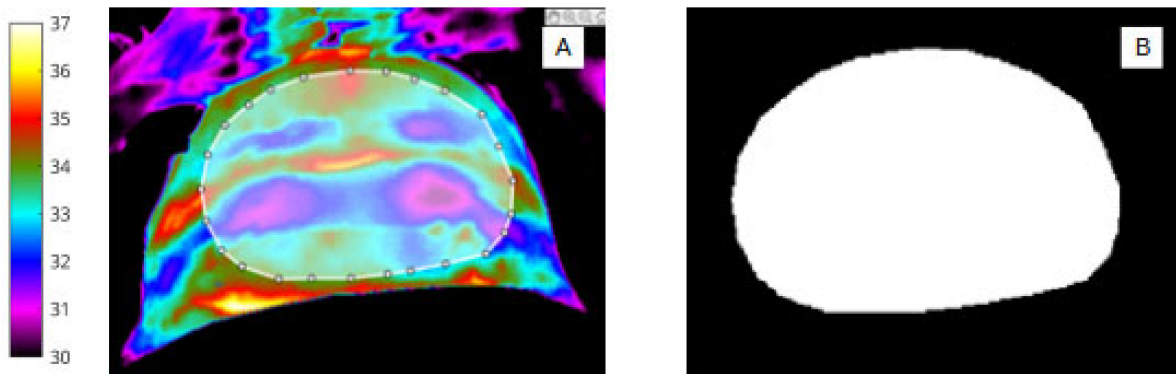


Figure 2. (A) Infrared image showing segmentation of the abdomen (ROI 1) and white bounding box. (B) Binary mask of abdomen (ROI 1).

The overarching objective was to produce a fully annotated image data set and ground truth masks reflecting both the ROIs themselves and the high and low skin perfusion loci across both ROIs.

2.3. Analysis

Distribution of agreement of mask pairs was calculated using Jaccard similarity coefficient (see Equation). In this case, the Jaccard Similarity coefficient measures the intersection over the union of the mask pairs (that is, it provides the ratio of the overlapping masks vs. the separate).

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}. \quad (1)$$

2.4. Abdomen and Wound Segmentation

For abdomen (Figure 2A,B) and wound segmentation (Figure 3A,B), the respective regions were identified in infrared and the ROI was selected and cropped. With respect to the surgical site, ROI 2, determining the extent of the area, was often difficult to identify in infrared due to lack of clear anatomical landmarks. To help discern wound area from surrounding healthy skin, reference was made to the corresponding digital image.

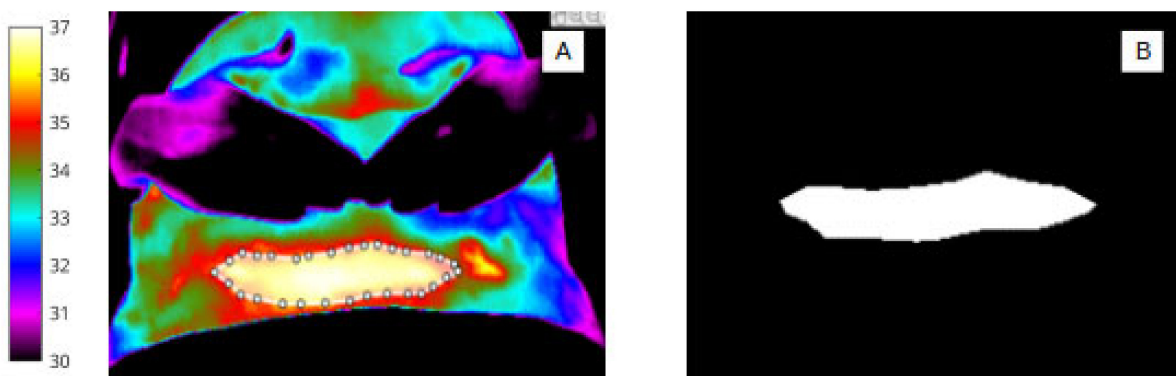


Figure 3. (A) Pixel-based wound segmentation (ROI 2). (B) Binary mask of wound region (ROI 2).

2.5. Segmentation of ‘Object’ Clusters

To identify pixel clusters of low temperature, shown qualitatively by the (rainbow) colour palette on the thermal image (Figure 4), automated minima regions were selected using MATLAB (by choosing continuous regions within the ROIs that are 0.5 degrees or more cooler than the surrounding areas) to produce pixel masks within the image bounding box (abdomen and wound).

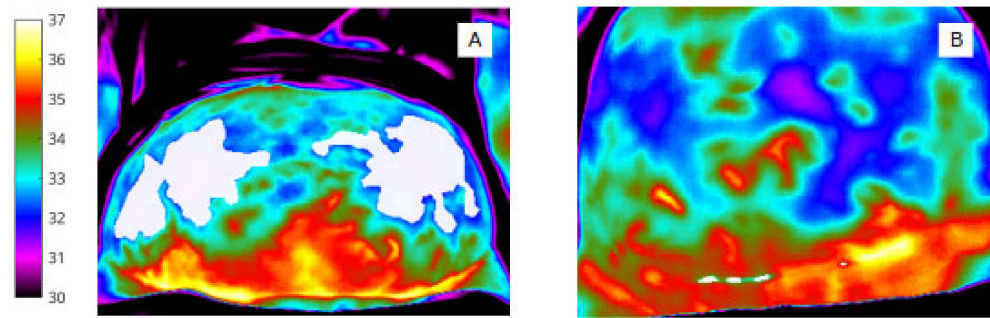


Figure 4. ‘Object’ labels identify extent and site of abdominal skin ‘cold regions’ (A) and wound ‘cold spot’ clusters, (B) (shown in white for clarity in this image).

2.6. Impact of Agreement Outcomes on Wound Complication Prognosis

Data extracted from thermal maps obtained in 50 women recruited during the original study [3] were used to calculate temperature over abdominal and wound regions as well as the temperature difference between wound and abdomen sites.

Using logistic regression, temperature difference (wound minus abdomen temperature) at day 2 was shown to have good prognostic performance; 74% accurate in surgical site infection (SSI) prediction with improved accuracy when analysed at day 7 (79% accurate). However, at day 2 and day 7, there were 12 and 9 false positives, respectively, for SSI.

These data were collected by one individual only throughout the study. For the current report, differences in the masks obtained by two independent reviewers were compared for the extracted temperature data from the masks obtained.

Ethical approval was obtained from Derby Research Ethics Committee 14/EM/0031, for the original study with an amendment to protocol (January 2021) to include additional data and image analysis.

2.7. Segmentation Agreement

Of 366 matched image pairs for both wound and abdomen sites (732 labels in total), the distribution of agreement ranged from 0.35 to 1. Whilst overall the average agreement was 0.74 (standard deviation 0.141), the average agreement for the abdomen was much better (0.85, with standard deviation 0.070) than the average agreement for the wound site (0.63, with standard deviation 0.107).

Good segmentation agreement (Jacquard similarity coefficient ≥ 0.7) of mask size and shape was observed for abdomen in all but four image label pairs (Figure 5), indicating that manual segmentation agreement of less than 70% between two observers occurred in just 1.1% of abdomen pairs only. Examples of the best agreement between reviewers in this study are shown in Figure 6 and agreement in the range of 50–70% in Figure 7. We can see from Figure 7 that the four cases of worst agreement in the segmentation of the abdomen ROI occurred mainly due to confounding factors (such as the inclusion of hands for patient 29 at day 15, disjoint temperature areas for patient 10 at day 16, and a colostomy bag in the abdomen ROI image for patient 42 at day 7).

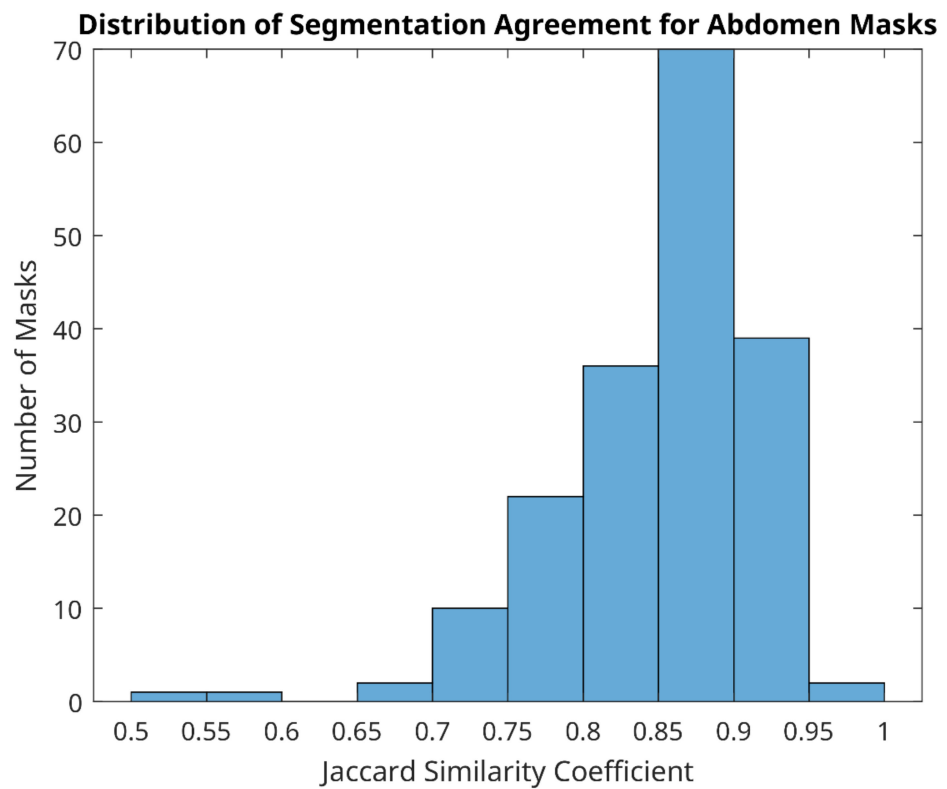


Figure 5. Distribution of segmentation agreement for abdomen masks between two reviewers.

Patient 34, Day 02, Agreement 0.960



Patient 46, Day 07, Agreement 0.969

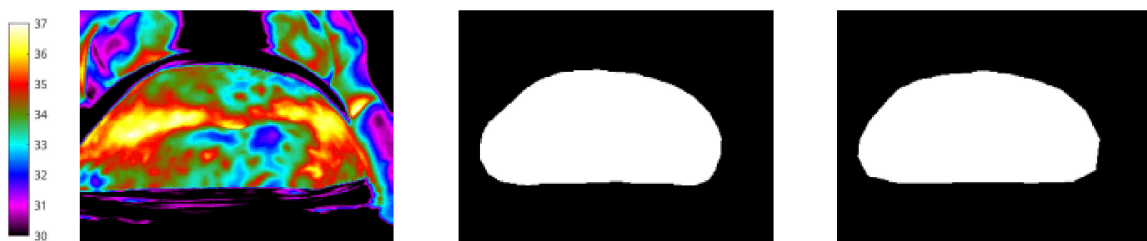


Figure 6. Best agreement (over 95%) reveals the potential for near identical mapping of mask outline at abdomen site. Left column shows temperature scale and thermal image of abdomen of two women; middle column shows segmentation masks produced by reviewer 1, and end column presents segmentation mask produced by reviewer 2.

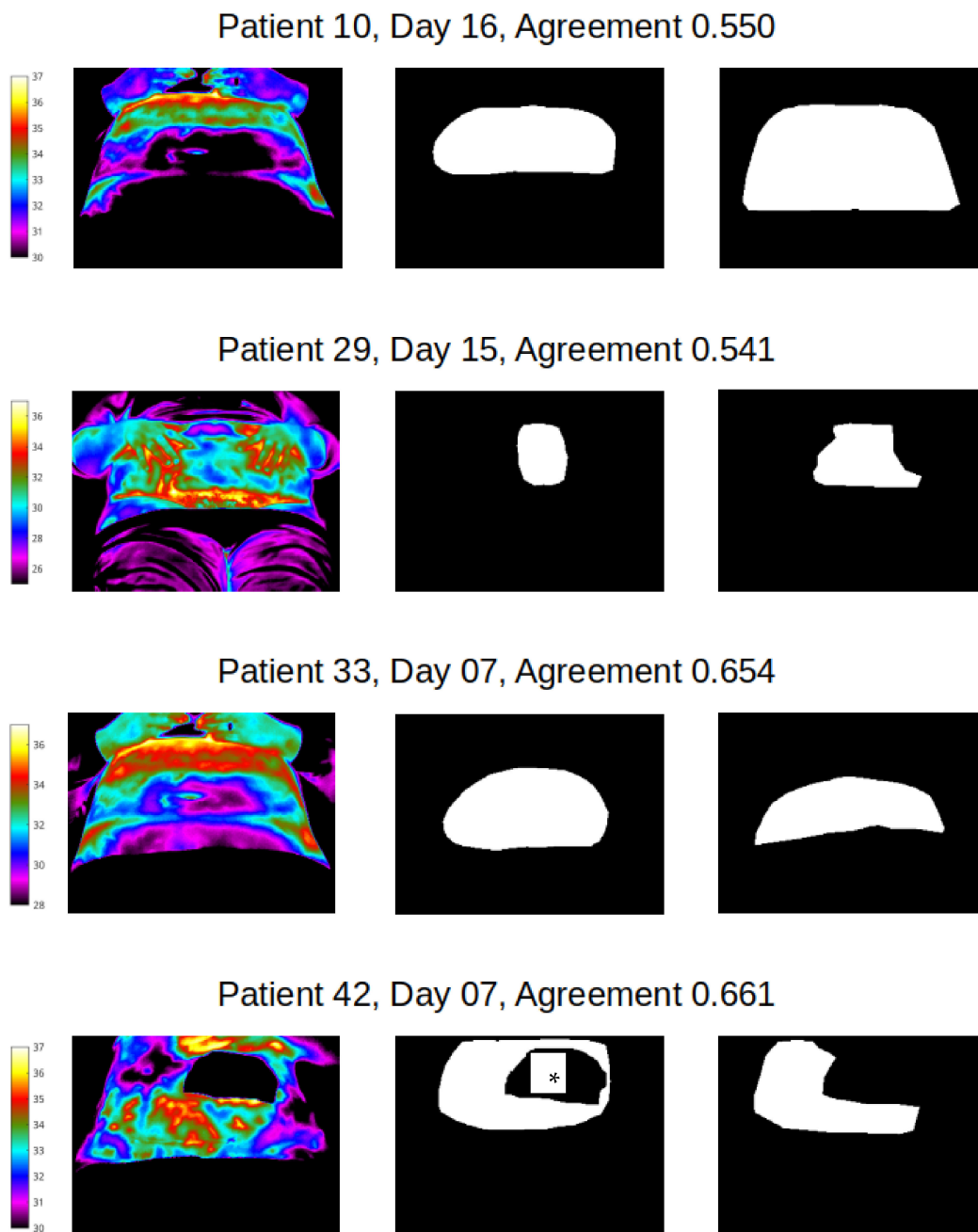


Figure 7. Thermal image (with temperature scale, left) and agreement co-efficient for abdomen pairs (reviewer 1 vs reviewer 2), all with coefficients of 50–70% * image shows position of colostomy bag.

However, distribution of agreement between reviewer 1 and 2 was mainly poor for wound site (Figure 8), with 128/183 (70%) of mask pairs with similarity coefficient < 0.7 and with 21/183 mask pairs (11.5%) having less than 50% agreement. Good agreement (Jaccard similarity coefficient ≥ 0.7) was observed in just 30% of masks for the wound site. Seven examples for wound agreement varying from poor (A) to good (B) are provided in Figure 9. The main cause of lack of agreement in the segmentation of the wound site ROI appears to be differences in opinion between the two reviewers in regard to how closely to follow temperature profiles, with reviewer 2 choosing to follow the temperature contours more closely during the segmentation process and reviewer 1 choosing to focus on the broader shape of the wound site ROI. This is an area that can be addressed through additional training.

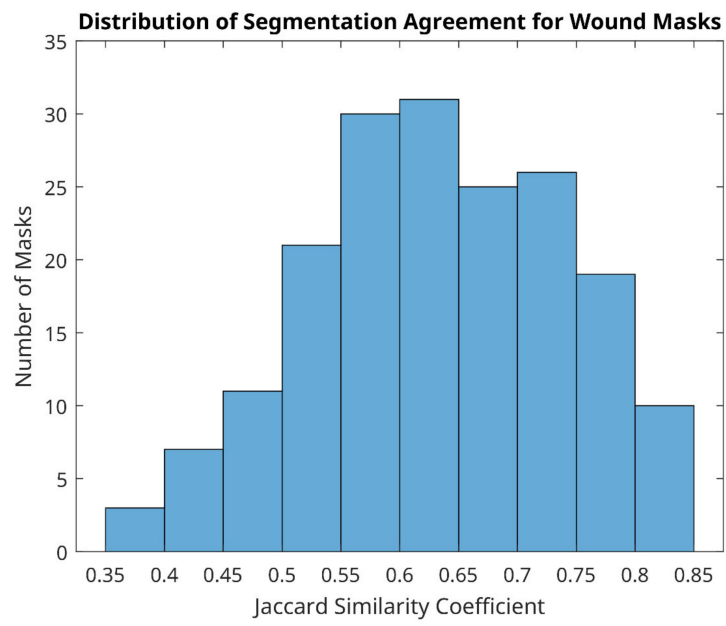
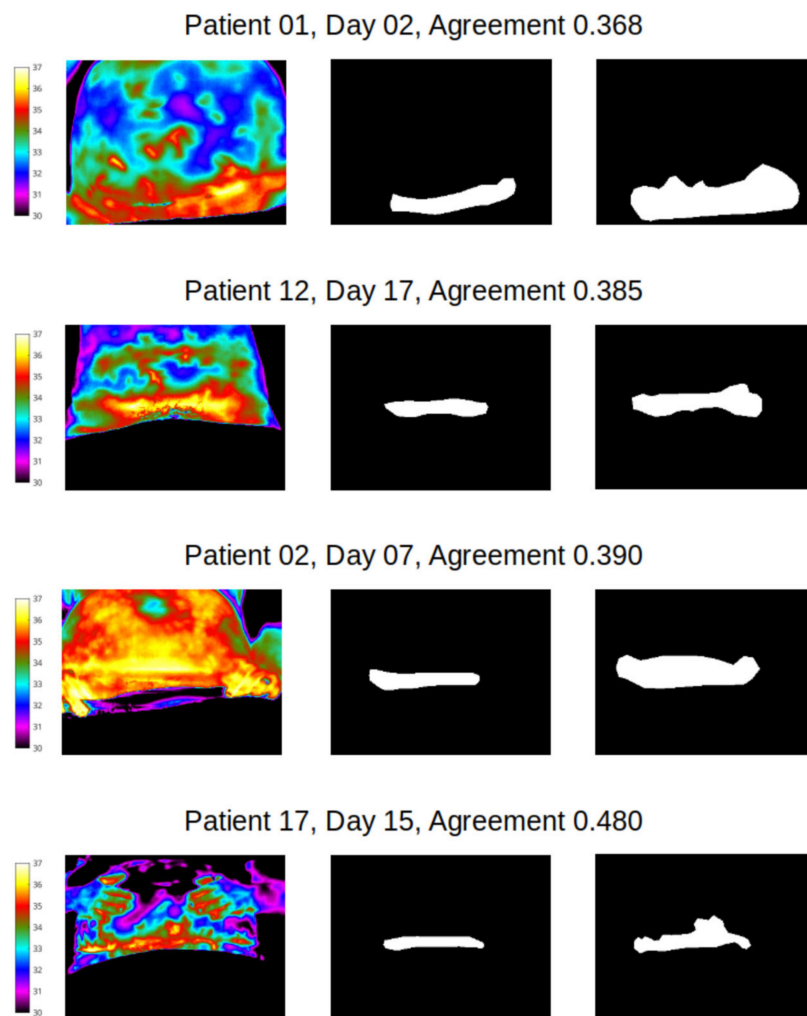


Figure 8. Distribution of segmentation agreement for wound masks between two reviewers.



(A)

Figure 9. Cont.

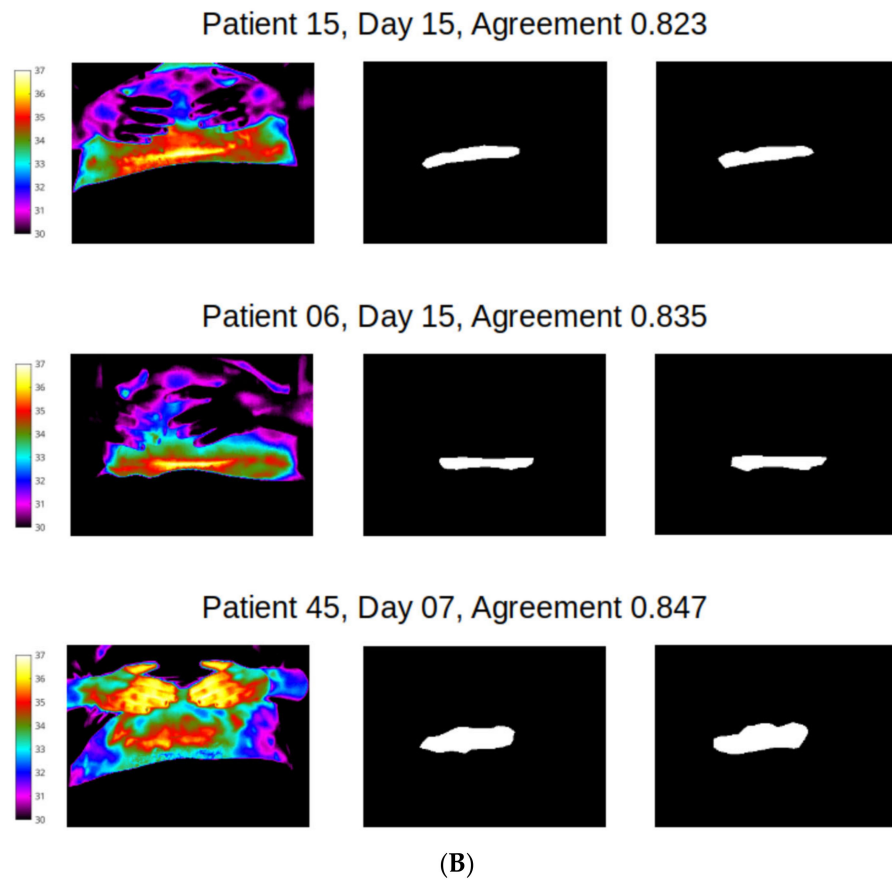


Figure 9. Differences in agreement: (A) poor (less than 50%); (B) agreement between observers (in the range 80–85%).

Results from the original study [3] using logistic regression between subjectively drawn ROIs for wound and abdomen were calculated at days 2, 7, 15 and 30 after surgery. For each 1 °C widening of temperature between wound and abdomen sites (wound minus abdomen temperature difference, WATD), there was a two-fold raised odds of SSI at day 2 (odds ratio (OR) 2.25, 95% CI 1.07–5.15; $p = 0.034$) with 70.1% of cases correctly classified. Improved model performance was noted at later times after surgery such that by day 7; 78.7% of cases were correctly classified (OR 2.45; 95% CI, 1.13–5.29 $p = 0.023$).

By comparison, results from the present study revealed that temperature extracted from abdomen and wound masks provided by two reviewers showed variable impact on model performance. For example, at day 2, temperature differences between the two ROIs (WATD) extracted from masks created by reviewer 1 showed 76% accuracy at day 2 (a difference of 5% compared with the original result, 70.9%). For reviewer 2, day 2 results obtained from the masks for abdomen and wound were 74% accurate to predict SSI outcome, a difference of 3% with respect to the original prediction for development of SSI.

2.8. Feature Extraction

(a) ROI features

The key temperature features extracted from masked areas were: mean, maximum, and minimum values of abdomen and wound ROIs as well as the average temperature difference between the two ROIs (WATD) (Table 1). Poor agreement between the two reviewers particularly affected the mean of wound ROIs.

Table 1. Difference in ROI features for abdomen and wound ROIs between reviewer 1 and reviewer 2 for data extracted at day 2 and day 7.

Feature	Mean Difference between Reviewers (°C)	Standard Deviation (°C)
Day 2 Abdomen Mean	0.04	0.04
Day 2 Wound Mean	0.12	0.10
Day 2 ROI Temperature Difference	0.11	0.09
Day 7 Abdomen Mean	0.07	0.07
Day 7 Wound Mean	0.10	0.08
Day 7 ROI Temperature Difference	0.12	0.109

(b) Cold spot features

One hypothesis was that women who proceeded to develop SSI after caesarean section would show the appearance of cold spots within the wound (as an early signature or prodrome of SSI). To test this hypothesis, the combined masks of both reviewers were used to determine the wound region of interest, and then the regional minima (i.e., spots within the ROI that were notably cooler than the surrounding area) were calculated using MATLAB. The number and total area (in pixels) of each of these cold spots was then calculated for every patient at each time point. Table 2 shows the mean and standard deviation of these features.

Table 2. Number of cold spot features and wound area (pixels) at follow-up points; day 2, 7 and 15 after surgery (n = 50 women).

Feature	Non-SSI		SSI	
	Mean	Std	Mean	Std
Day 2 Number of cold spots	2.8	3.5	4.0	2.9
Day 2 Area of cold spots (pixels)	48.2	71.6	75.3	51.6
Day 7 Number of cold spots	3.7	4.3	3.4	3.5
Day 7 Area of cold spots (pixels)	85.0	147.2	45.4	50.8
Day 15 Number of cold spots	2.4	2.5	3.3	4.0
Day 15 Area of cold spots (pixels)	53.2	65.5	86.4	99.1

std, standard deviation. Thermal camera, focal plane array, uncooled microbolometer, pixel resolution 320 × 240 (76,800 pixels); pixel size 1.39 mm at 1 m distance from target.

3. Discussion

In this study cohort of obese women, we have used caesarean section as a model of acute incisional surgery. The main objective, to refine and improve scope for image analysis using AI-based segmentation, has been achieved, in that it is now possible to understand the extent to which the wound infection outcomes observed in the original study are altered by a different set of reviewers following a brief training overview. Manual segmentation of paired regions produced boundary masks, each with a set of per-pixel temperature data. Whilst the abdomen covered a larger area of interest, agreement between reviewers was good with only four masks having less than 70% agreement. By contrast, segmentation of the wound required attention to jagged edges and uneven margins, which clearly exaggerated subjectivity of opinion as to where the boundary between wound and healthy skin lay. This led to the majority of wound pair masks showing poor agreement; 30% of mask pairs only had a similarity coefficient of ≥ 0.7 . Examining the thermal images and corresponding masks for the least agreement in segmentation of the wound ROI in Figure 8a, a key factor appears to be the ability of the reviewer to closely follow the temperature

contours when undertaking segmentation of the region of interest. This could potentially be somewhat mitigated by providing more guidance in respect to the segmentation approach that the reviewers should use. However, this does highlight the subjective nature of this segmentation process.

Whether a longer training session would have improved the result such that reviewers would have produced masks that were more perfectly aligned is unclear, but what we do see (as shown in Table 1) is that differences in the segmentation process and resultant mask shape and size leads to differences in the temperature values derived from each mask. When comparing both novice reviewers' wound and abdomen segmentation results to produce a delta value (wound minus abdomen temperature, WATD) against the original segmentation by an experienced clinician [3], the confirmed SSI events differed (at day 2) by 5% and 3% for reviewer1 and reviewer2, respectively.

Given that the average wound ROI temperature difference between women in the original study [3] (who later developed an SSI, compared with those who did not and who went on to heal without complications) was 0.2 °C at day 2 and 0.1 °C at day 7, better agreement of the wound area masks is needed between the reviewers to provide confidence in the ability to use these temperature features from the wound in a predictive capacity. We can see that the mean temperature difference between reviewers for the abdomen ROI, where there was a much more consistent (and more regular) mask shape, had better agreement than that for the wound ROI, where the mask shape was irregular and inconsistent. This indicates that for the larger (abdominal) ROI, the related temperature features are more robust to segmentation inconsistencies occurring between the reviewers than for irregular-shaped areas such as the wound.

These ROI temperature differences, and the knock-on effect this has on the consistency and performance of the logistic regression classification performance, shows the need for an independent and objective approach to ROI segmentation. Although we believe that this may partly be addressable by additional training, it is difficult to clearly identify the regions of interest in thermal images even for an experienced clinician without a true colour image overlay.

In this study, a relatively large image sample size was used in this medical machine learning segmentation application. Smaller sample sizes have been used by others. For example, Ronnenberger, et al. [17] used 30 images for training of their UNet model. It is recognised, however, that there are limitations with respect to the number of reviewers involved. Nevertheless, their performance with respect to the input (image) data in the first stage of algorithm development reveals a lack of correspondence (agreement) of the ROI indicating the need for further work to refine the characteristics of output labels (masks) before an unsupervised algorithm works effectively to learn patterns and features of the wound, i.e., those wounds destined to heal without complications and those that fail to heal optimally.

Nevertheless, it was clear from Table 2 that women who proceeded to develop an SSI had a larger number of cold spots within the wound at day 2 (by 1.07 on average) and that these cold spots covered a large area of the wound (23.92 pixels greater on average). Interestingly, the difference in number of cold spots was much lesser at day 7 (3.29 for women who developed an SSI vs. 2.79 for women who did not) and the number of cold spots was nearly the same at day 15 (2.69 for patients who developed an SSI vs. 2.39 for patients who did not). However, whilst the descriptive statistics in Table 2 indicated that the number and extent of cold spots within the wound might have predictive capability, univariate logistic regression analysis of each of these features was inconclusive for this sample size studied. An example is included here, in Figure 10.

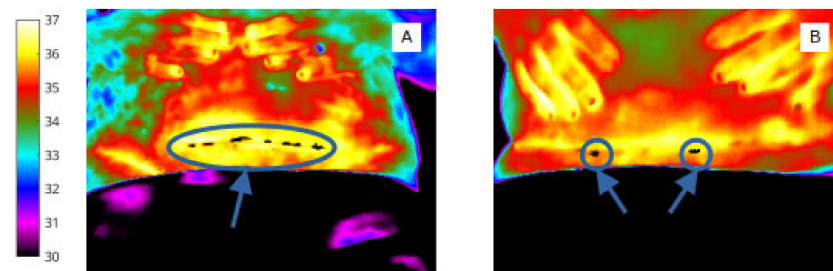


Figure 10. Cold spots in the wound site at day 2 of a woman who proceeded to develop SSI after caesarean section (A) and a woman with an uncomplicated wound recovery period (B).

The current technological challenge for wound imaging automation is in developing independent methods to view and describe visual features. Clinicians see a mix of features over time. However, if clues or signatures as to the likelihood of downstream complications can be provided as an aid to diagnosis and prompt treatment, there is benefit in progressing with wound imaging automation. From a clinical viewpoint, however, there are obstacles, especially as the long-held traditions of simply looking at the wound with an ‘experienced eye’ is both comfortable and accepted practice worldwide; one which has functioned for centuries. As a motivation to be fulfilled, it became clear that an independent method for predicting wound complications, specifically SSI, was needed once we had observed that visual inspection and opinion of the wound by experienced clinicians showed minimal agreement even amongst the most experienced clinicians. This lack of agreement in the visual assessment of a wound, whether by verbally reporting what is seen at the time of inspection or by use of digital photography [18,19], is concerning and resonates with current concerns across the UK National Health Service (NHS) of the unwarranted variations in medical practice, now a national programme for improvement (Getting it right first time (GIRFT) [20].

With reference to wound care, standards and treatments remain plagued by myths and inconsistency [21], resulting in escalating treatment costs, particularly those complex chronic wounds where identification and assessment of slow or stalled healing continue to be a management challenge. It is a testament to the issues driving the need for improvement that wound care recently became a topic of debate in the House of Lords.

To supersede the conventional ‘wisdom’ of what clinicians recognise as a problem and then to translate this learning to specific wound features via AI automation requires a detail of the many layers of judgement that is made almost unconsciously many times a day. This backdrop to wound care nationally, and the recognition that new diagnostic technologies are needed in the setting of wound care [22] is a recommendation aligning well to the need to make progress in wound assessment.

It is now clear from a number of reports that rapid developments in computer processing will revolutionise automation of wound tracking and assessment to improve the ability to predict wound healing outcomes. However, although automated documentation of wound metrics such as depth, size and volume indicate wound healing progress [23] with devices available commercially (<https://www.researchandmarkets.com/reports/5578241/digital-wound-measurement-devices-market-by#product--summary>, accessed on 15 March 2020), AI-based systems for pattern recognition and feature extraction for wound healing prognosis are still in their infancy [24].

Automation and wound image recognition starts with the region of interest; including aspects that should be included and those that should be omitted in the analysis. Focus is then on ‘object’ definitions that warrant classifying as predictors of complications. Turning to the wound as an image obtained during the course of routine clinical care, there remains a need for the clinician to interpret the image and to identify ‘objects’ that can then be classified to a specific pathological category, understood by the clinician and relevant to the patients presenting condition. This is the case irrespective of how the digital wound picture or scan was derived. For example, in the event of imaging using computed tomography

(CT) or Colour Doppler [25]). In this study, we have focused on infrared thermography of acute incisional wounds where automation has potential for predicting healing outcomes providing that the segmentation process is reliable.

As a starting point, image data were already available as infrared maps of the abdomen and wound site through laborious image post processing undertaken by a single investigator with many years of experience of both wound assessment and thermal imaging. By introducing two novice reviewers to review the images, the problem inherent in technology development has been highlighted; accurate segmentation and image boundaries need to be established at the earliest stages of image automation. These are based on the way people recognize objects and features that ultimately provide a route to reliable, independent wound healing risk stratification and ultimately diagnostic or prognostic signatures of later complications.

4. Conclusions

From a large medical image dataset acquired during a prospective thermal imaging study of women with post-operative wounds, methods were developed to refine the original, manual, segmentation process, by introducing granular, pixel-level image labelling. Undertaken by two independent reviewers with no previous experience of wound imaging, reviewer performance, with respect to the input (image) data in the first stage of algorithm development, reveals that the more regular (ellipse)-shaped abdomen, with its 'smooth' outlines, produces a better agreement between raters than the irregular-shaped wound region; in this study, it was the incisional wound area. The differences observed between reviewers with respect to segmentation, whilst small in terms of temperature values, had an impact on predicted wound outcome. The differences observed between the raters was equivalent to the differences observed in the original study between event (SSI) and no event (uncomplicated wound healing).

With regard to feature extraction from within the wound area, thermal cold spots representing low temperature features were present; the pixel area of cold spots was greatest at day 2 for those who later developed SSI.

In summary, segmentation performed by different reviewers has the potential to introduce errors, the magnitude of which could impact primary outcome measures; in this study, it was wound complications (infection) versus no infection. This work indicates the need for further refinement of the characteristics of output labels (masks) before an unsupervised algorithm works effectively to learn patterns and features of the wound.

Author Contributions: Study conception and design (C.C. and A.S.), original study recruitment and consent (C.C.), protocol development (A.S. and C.C.), data collection (C.C.), image analysis (O.V., H.N. and A.S.), preparation and writing of the manuscript (C.C. and A.S.). All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by a grant from Grow MedTech/Research England.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by ethics review board of Derby Research Ethics Committee. 14/EM/0031.

Informed Consent Statement: Informed consent was obtained from all subjects for analysis of image data.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to intellectual property restrictions.

Acknowledgments: We thank all the women who willingly provided their time to participate in the original study. Without their contribution, this work would not have been possible.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. World Health Organisation. Obesity and Overweight. 9 June 2021. Available online: <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight> (accessed on 19 March 2023).
2. Ogunwole, S.M.; Zera, C.A.; Stanford, F.C. Obesity Management in Women of Reproductive Age. *JAMA* **2021**, *325*, 433. [CrossRef] [PubMed]
3. Childs, C.; Wright, N.; Willmott, J.; Davies, M.; Kilner, K.; Ousey, K.; Soltani, H.; Madhuvrata, P.; Stephenson, J. The surgical wound in infrared: Thermographic profiles and early stage test-accuracy to predict surgical site infection in obese women during the first 30 days after caesarean section. *Antimicrob. Resist. Infect. Control.* **2019**, *8*, 7. [CrossRef] [PubMed]
4. Yeeles, H.; Trinick, S.; Childs, C.; Soltani, H.; Farrell, T. Postpartum Infection in Morbidly Obese Women after Caesarean Section: Does Early Prophylactic Oral Antibiotic Use Make a Difference? *Open J. Obstet. Gynecol.* **2014**, *4*, 547–549. [CrossRef]
5. Molnlycke. A State of the Nation Report on Surgical Site Infections in the UK. 2020. Available online: <https://www.molnlycke.co.uk/news-events/news/new-state-of-the-nation-report-calls-for-urgent-action-to-drive-down-surgical-site-infection-rates/> (accessed on 19 March 2023).
6. Piednoir, E.; Robert-Yap, J.; Baillet, P.; Lermite, E.; Christou, N. The Socioeconomic Impact of Surgical Site Infections. *Front. Public Health* **2021**, *9*, 712461. [CrossRef] [PubMed]
7. Childs, C.; Siraj, M.; Fair, F.; Selvan, A.; Soltani, H.; Wilmott, J.; Farrell, T. Thermal territories of the abdomen after caesarean section birth: Infrared thermography and analysis. *J. Wound Care* **2016**, *25*, 499–512. [CrossRef] [PubMed]
8. Hennessy, O.; Potter, S.M. Use of infrared thermography for the assessment of free flap perforators in autologous breast reconstruction: A systematic review. *JPRAS Open* **2020**, *23*, 60–70. [CrossRef] [PubMed]
9. Nergård, S.; Mercer, J.B.; de Weerd, L. Impact on Abdominal Skin Perfusion following Abdominoplasty. *Plast. Reconstr. Surg. Glob. Open* **2021**, *9*, e3343. [CrossRef] [PubMed]
10. Ng, E.Y.K.; Etehadtavakol, M. (Eds.) *Application of Infrared to Biomedical Sciences*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 5–17.
11. Childs, C.; Soltani, H. Abdominal Cutaneous Thermography and Perfusion Mapping after Caesarean Section: A Scoping Review. *Int. J. Environ. Res. Public Health* **2020**, *17*, 8693. [CrossRef] [PubMed]
12. Siah, C.; Childs, C. Thermographic mapping of the abdomen in healthy subjects and patients after enterostoma. *J. Wound Care* **2015**, *24*, 112–120. [CrossRef] [PubMed]
13. Siah, C.R.; Childs, C.; Chia, C.K.; Cheng, K.F.K. An observational study of temperature and thermal images of surgical wounds for detecting delayed wound healing within four days after surgery. *J. Clin. Nurs.* **2019**, *28*, 2285–2295. [CrossRef] [PubMed]
14. Selvan, A.N.; Childs, C. Infrared Thermal Mapping, Analysis and Interpretation in Biomedicine. In *Application of Infrared to Biomedical Sciences*; Ng, E.Y.K., Etehadtavakol, M., Eds.; Springer Nature: Singapore, 2017; pp. 377–394.
15. Chen, L.; Strauch, M.; Merhof, D. Instance Segmentation of Biomedical Images with an Object-Aware Embedding Learned with Local Constraints. *arXiv* **2019**, arXiv:2004.09821.
16. Abuelghar, W.M.; El-Bishry, G.; Emam, L.H. Caesarean deliveries by Pfannenstiel versus Joel-Cohen incision: A randomised controlled trial. *J. Turk. Gynecol. Assoc.* **2013**, *14*, 194–200.
17. Ronneberger, O.; Fischer, P.; Brox, T. *U-Net: Convolutional Networks for Biomedical Image Segmentation*; Springer International Publishing: Berlin/Heidelberg, Germany, 2015.
18. Carrión, H.; Jafari, M.; Bagood, M.D.; Yang, H.-Y.; Isseroff, R.R.; Gomez, M. Wound Healing Assessment Using Digital Photography: A Review. *Br. J. Healthc. Med. Res.* **2016**, *4*, 1.
19. Carrión, H.; Jafari, M.; Bagood, M.D.; Yang, H.-Y.; Isseroff, R.R.; Gomez, M. Automatic wound detection and size estimation using deep learning algorithms. *PLoS Comput. Biol.* **2022**, *18*, e1009852. [CrossRef] [PubMed]
20. Carter, P.R. *Operational Productivity and Performance in English NHS Acute Hospitals: Unwarranted Variations*; An Independent Report for the Department of Health; Department of Health: London, UK, 2016.
21. Flanagan, M. *Barriers to the Implementation of Best Practice in Wound Care*; Wounds: London, UK, 2005; p. 1.
22. O'Neill, J. Tackling Drug-Resistant Infections Globally: Final Report and Recommendations. In *Review on Antimicrobial Resistance*; Wellcome Trust and HM Government: London, UK, 2016.
23. Flanagan, M. Wound measurement: Can it help us to monitor progression to healing? *J. Wound Care* **2003**, *12*, 189–194. [CrossRef] [PubMed]
24. Anisuzzaman, D.M.; Wang, C.; Rostami, B.; Gopalakrishnan, S.; Niezgod, J.; Yu, Z. Image-Based Artificial Intelligence in Wound Assessment: A Systematic Review. *Adv. Wound Care* **2022**, *11*, 687–709. [CrossRef]
25. Evans, D.H.; Jensen, J.A.; Nielsen, M.B. Ultrasonic colour Doppler imaging. *Interface Focus* **2011**, *1*, 490–502. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.