Available online at www.sciencedirect.com
ScienceDirect
journal homepage: www.elsevier.com/locate/burns

Improving burn depth assessment for pediatric scalds by AI based on semantic segmentation of polarized light photography images

Marco Domenico Cirillo [a,b,c,*], *Robin Mirdell* [d,e,f], *Folke Sjöberg* [d,e,f], *Tuan D. Pham* [g,**]

[a] *Department of Biomedical Engineering, Linköping University, Linköping, Sweden*
[b] *Centre for Medical Image Science and Visualization, Linköping University, Linköping, Sweden*
[c] *Center for Medical Image Science and Visualization, Linköping University, Linköping, Sweden*
[d] *The Burn Centre, Linköping University Hospital, Linköping, Sweden*
[e] *Department of Plastic Surgery, Hand Surgery, and Burns, Linköping University, Linköping, Sweden*
[f] *Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden*
[g] *Center for Artificial Intelligence, Prince Mohammad Bin Fahd University, Khobar, Saudi Arabia*

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Artificial intelligence
Deep learning
Convolutional neural networks
U-Net
Semantic segmentation
Paediatric burns

ABSTRACT

This paper illustrates the efficacy of an artificial intelligence (AI) (a convolutional neural network, based on the U-Net), for the burn-depth assessment using semantic segmentation of polarized high-performance light camera images of burn wounds. The proposed method is evaluated for paediatric scald injuries to differentiate four burn wound depths: superficial partial-thickness (healing in 0–7 days), superficial to intermediate partial-thickness (healing in 8–13 days), intermediate to deep partial-thickness (healing in 14–20 days), deep partial-thickness (healing after 21 days) and full-thickness burns, based on observed healing time.
In total 100 burn images were acquired. Seventeen images contained all 4 burn depths and were used to train the network. Leave-one-out cross-validation reports were generated and an accuracy and dice coefficient average of almost 97% was then obtained. After that, the remaining 83 burn-wound images were evaluated using the different network during the cross-validation, achieving an accuracy and dice coefficient, both on average 92%.
This technique offers an interesting new automated alternative for clinical decision support to assess and localize burn-depths in 2D digital images. Further training and improvement of the underlying algorithm by e.g., more images, seems feasible and thus promising for the future.

1. Introduction

Burn wounds occur when the skin comes in contact with fire, hot water, electricity, or chemicals. Depending on temperature and contact duration with the skin, different burn depths develop. Burn depth may be classified into separate levels [1]: superficial partial-thickness (I), superficial to intermediate partial-thickness (II), intermediate to deep partial-thickness (III), deep partial and full-thickness burns (IV). Importantly,

* *Corresponding author at*: Department of Biomedical Engineering, Linköping University, Linköping, Sweden.
** *Corresponding author*.
E-mail addresses: marco.domenico.cirillo@liu.se (M.D. Cirillo), tpham@pmu.edu.sa (T.D. Pham).
https://doi.org/10.1016/j.burns.2021.01.011

burns of deep partial or full-thickness depth benefit from excision and skin grafting to heal appropriately. Patients less than 4 years old who get a burn due to hot water (Scald) represent the 30–40% of the patients arriving at a Burn Centre. Being an age defined and a homogenous group facing burns mainly on the trunk and arms they were chosen for this evaluation.

Burn depths are correctly classified by expert clinicians with an accuracy around 64–76% and around 50% by non-expert clinicians [2–7]. Today, one tool that has been used successfully as a decision support for clinicians are based on laser Doppler [8,9] and on its most recent development: laser speckle contrast imaging (LSCI) [10–12]. Such instruments have been advocated in order to improve burn depths assessment and they are used occasionally by clinicians as a decision support device [13]. These techniques provide perfusion images of the injured skin. Shortcomings include that they require training and knowledge to be fully operational and most importantly is that the image generating procedure is challenging and thus time consuming. This has led to the limited clinical use of the methodology. From an accuracy perspective, the technique also requires at least two consecutive measurements to be able to classify the burn depth with reliable accuracy [10].

For these reasons, an automatic, fast, objective, and accurate method is sought to evaluate such types of injuries and with the goal to help clinicians (decision support), decide if a patient will bemefit from surgical treatment of the burn wound or not.

1.1. *AI based burn depth assessment by semantic image segmentation*

Artificial intelligence based on Convolutional neural networks for semantic image segmentation as fully convolutional neural networks [14], SegNet [15], U-Net [16], etc. became very attractive models in medicine because they combine local and global image information after which a pixel-wise based classification is provided [17]. The only disadvantage is that these models require a demanding learning and training process at the beginning (by a large computer calculating capacity), but after that, they compute separate image segmentation in a few seconds. During the last years, the U-Net has become quite popular in the medical field, so that many modified U-Nets were created and applied in medical applications. For example: V-Net [18], to segment the prostate; DUnet [19], to segment retinal vessels; H-DenseUNet [20], for segment liver and tumours in it; Attention U-Net [21], to segment the pancreas; and No new-Net [22] (2nd place winner in BraTS 2018 challenge), to segment brain tumours. In this paper we used a modified U-Net with residuals to segment four different burn-depths (superficial partial-thickness (I), superficial to intermediate partial-thickness (II), intermediate to deep partial-thickness (III), deep partial and full-thickness (IV)) in images generated by a high-performance light camera with polarisation filters with the aim to provide automated and objective images to be used by the burn surgeon for the burn-wound assessment support.

2. Method

2.1. *Patient population*

Consecutively arriving children, in the age range 0–4 years, at the outpatient clinic at the Linköping Burn Centre were included. Laser Doppler and laser Doppler Speckle imaging data from this cohort has previously been presented in a series of publications [2,3,10,11,23,24]. In short, the patients were anesthetised rectally with ketamin [25] and the wound bed was properly cleaned prior to image capture. Image capturing was done in a climate controlled room with regular indoor lightning (no windows). For this study, based on a high-performance light camera, images were taken in parallel to the one presented in the previous publication [3].

2.2. *Data*

One hundred burn wound images were acquired from patients with age equal or less 4 years old using a TiVi700, which is a tissue viability imaging device (WheelsBridge AB, Sweden). TiVi700 is a high-performance digital camera equipped with polarisation filters and flashlights all around its lens to avoid the reflecting artefact due to room light and/or the camera flash and burn wound fluid.

An example of such data is given in Fig. 1a, which shows a burn wound image captured by the TiVi700; whereas Fig. 1b shows its ground-truth labelled manually by a burn clinician expert of the Linköping University Hospital Burn Centre. The ground-truths, as the one in Fig. 1b, were defined based on the wound's healing time: a superficial partial-thickness wound healed within 7 days, a superficial to intermediate partial-thickness healed between 8–13 days; an intermediate to deep partial-thickness healed within 14–20 days; and a deep partial or full-thickness, which did not heal within 20 days and underwent surgery. Importantly, surgery was always done after day 20, which gives ground-truth a high degree of reliability as all children were observed until day 20 and healing earlier than that was recorded by one clinician. These earlier healing events were divided into re-epithelialization within 7, 14 or 21 days, respectively.

The target of this project is achieving a segmentation result, as in Fig. 1b, from a burn wound image, as in Fig. 1a, using artificial intelligence, more specifically a convolutional neural network, similar to the U-Net proposed by Ronneberger et al. [16], but with different depth, loss function, optimizer, and applying the residuals theory on it.

Since each burn wound image has a really complex background rich of objects (i.e. healthy skin, blanket, medical tools, nurses' gloves, monitors, etc.), this is removed in order to let the CNN only focuses on segmenting the region of interest, the burn wound, and distinguish between the four different burn-depths.

Convolutional neural networks minimise the dice loss [18,26] to achieve a good segmentation result rather than the more generally used cross-entropy loss, because the former does not count in the true negatives (the background), which normally have the major number of pixels in the image. The higher the dice coefficient is the higher the accuracy is, but the contrary is not true. The dice loss is mathematically defined as

$$DL = 1 - D = 1 - 2\frac{\sum_{c=1}^{C} w_c \sum_{n}^{N} g_{cn} p_{cn}}{\sum_{c=1}^{C} w_c \sum_{n}^{N} g_{cn} + p_{cn}}, \quad (1)$$

where DL stands for the dice loss, D for dice coefficient, C for number of classes, N for number of pixels, w_c for the weight

Please cite this article in press as: M.D. Cirillo, et al., Improving burn depth assessment for pediatric scalds by AI based on semantic segmentation of polarized light photography images, Burns (2021), https://doi.org/10.1016/j.burns.2021.01.011

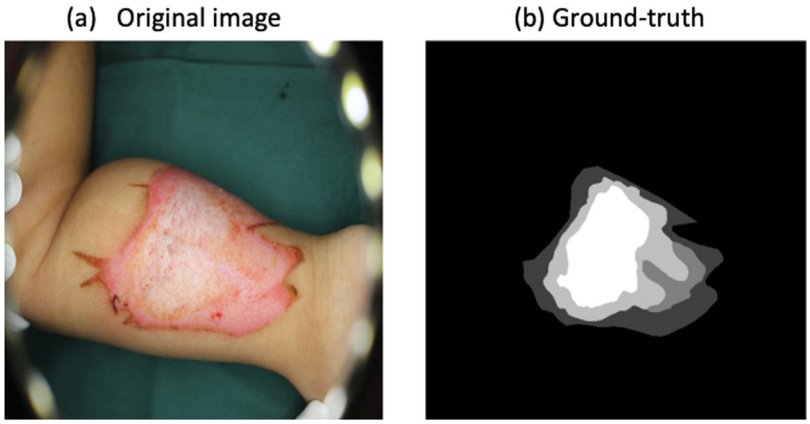


Fig. 1 – Original burn wound image (a) and its burn depth areas ground-truth (b) drew by a clinician specialist: white for deep partial and full-thickness depth; silver for intermediate to deep partial-thickness; grey for superficial to intermediate partial-thickness; dark grey for superficial partial-thickness; and, in the end, black for uninjured skin and the background.

assigned to class c, g_{cn} and p_{cn} for the n-th pixel which belongs to the ground-truth and to the network's prediction on c-th class respectively. When w_c is not a vector of ones, Eq. (1) represents the generalized dice-loss. Otherwise, w_c is defined as

$$w_c = \frac{N_s}{CN_c}, \tag{2}$$

where N_s is the number of pixels in the image and N_c the number of pixels that belong to the class c. In this way all the classes are balanced, because the network weighs each class according to the respective weight. If, for example, there are many pixels belonging to one class over the whole dataset, its weight will be low; vice-versa, if there are few pixels belonging to one class, its weight will be high according to Eq. (2). So, the network will pay more attention to learn a class represented by few pixels rather than a class by many pixels.

Since the burn image database has images representing all or some of the four burn depths, the segmentation step is applied only to the images that represent burns with all the burn depths (there are 17 in total) in order to enable the convolutional neural network to learn from a homogeneous dataset. A simplified diagram of the semantic segmentation is described in Fig. 2.

The accuracy (Acc), F_1 coefficient, intersection over union (IoU), precision (P) and sensitivity (S) are calculated for measuring the performance of the segmentation obtained from the second convolutional neural network using the ground truth. These metrics are calculated as:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$F_1 = \frac{2TP}{2TP + FP + FN} \tag{4}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{5}$$

$$P = \frac{TP}{TP + FP} \tag{6}$$

$$S = \frac{TP}{TP + FN} \tag{7}$$

where TP, TN, FP and FN represent true positive, true negative, false positive, and false negative, respectively. These values are calculated on the binary class images, so, for example there is a TP when both the ground-truth and model's prediction segmentation have value 1 for the same pixels. Fig. 3 illustrates the space of the defined metrics for an image segmentation.

The algorithm was written in Python 3.6, using the Keras library [27] functions on a super-computer with 512 GB RAM, 2 Intel(R) Xeon(R) CPU E5-2697 v4 @ 2.30 GHz, 18 cores each, and 3 Nvidia GTX 1080 8 GB.

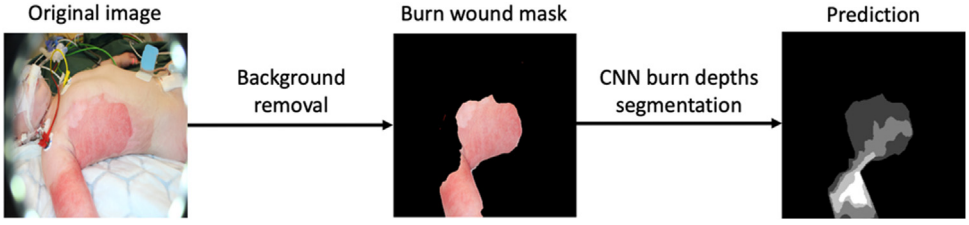


Fig. 2 – Diagram semantic segmentation process.

Please cite this article in press as: M.D. Cirillo, et al., Improving burn depth assessment for pediatric scalds by AI based on semantic segmentation of polarized light photography images, Burns (2021), https://doi.org/10.1016/j.burns.2021.01.011

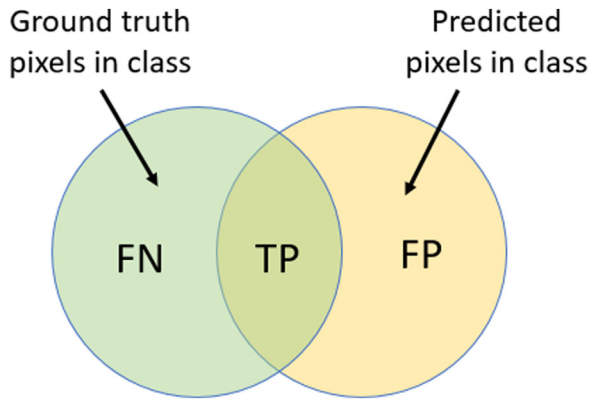


Fig. 3 – Illustration of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) between a binary ground-truth and its prediction.

This study was approved by the Regional Ethics Committee in Linköping and conducted in compliance with the "Ethical principles for medical research involving human subjects" of the Helsinki Declaration.

2.3. *Training of the algorithm*

Before starting the training process, since there were only 17 images available with all the burn depths present, data augmentation is strongly needed. In order to evaluate the convolutional neural network, leave-one-out cross-validation is computed, so 16 images are used for the training and validation set and just 1 for the testing set. On these 16 original images, rotations of 0°, 90°, 180° and 270° are applied and for each of these rotated images other new 40 images are created using the elastic deformation technique [28]. In the end, 3936 images are augmented from 16 original ones and then split 90–10% into training and validation set respectively for the second convolutional neural network training process.

From Table 1, it is possible to notice that the best network, the one with the highest dice coefficient, is the number 3. Minimizing the dice-loss, accuracy and dice coefficient converge at almost the same value and, after leave-one-out cross-validation, the system has average accuracy and dice coefficient of 96.81%. Moreover, the average weights for each class to balance the training process, calculated using Eq. (2) after image augmentation on each leave-one-out fold, are:

$$w = \begin{matrix} w_0\,(Background) \\ w_1\,(Superficial\,I) \\ w_2\,(Superficial\,Partial\,thickness\,II) \\ w_3\,(Deep\,Partial\,Thickness\,III) \\ w_4\,(Full\,Thickness\,IV) \end{matrix}\Big],$$

where w_0 is the weight which belongs to the background, whereas the others to the burn-depth classes I, II, III and IV respectively (see Eq. (2)). As wanted, the background weight has a small value and, on the other hand, the full and deep-thickness depth weight has a high value, whereas class II and III have similar weights, so probably the classification between them might be complicated.

Table 1 – Accuracy and dice coefficient values obtained after leave-one-out cross-validation.

Network	Accuracy	Dice coefficient
1	0.8814	0.8812
2	0.8042	0.8040
3	0.9977	0.9977
4	0.9968	0.9967
5	0.9972	0.9972
6	0.9930	0.9930
7	0.9568	0.9567
8	0.9937	0.9936
9	0.9906	0.9906
10	0.9911	0.9911
11	0.9976	0.9976
12	0.9852	0.9852
13	0.9865	0.9864
14	0.9867	0.9867
15	0.9840	0.9480
16	0.9898	0.9898
17	0.9619	0.9617
Average	0.9681 ± 0.0498	0.9681 ± 0.0498

Fig. 4, here below, shows four different semantic segmentation results, using the networks 3, 10, 12 and 16 of Table 1. on their respective test images. Each image illustrates the burn wound without the background, its ground-truth and the convolutional neural network's prediction. Moreover, it reports the accuracy, F_1 coefficient, intersection over union, precision and sensitivity metrics extracted from the ground-truth and the convolutional neural network's prediction for each class (see Eqs. (3)–(7)).

It is possible to conclude that Fig. 4 illustrates four good semantic segmentation results because the metrics reported have really high values. In Table 2 are reported the average of the same metrics over all the 17 burn-wound images for each class, and it is possible to notice that class II and class III are the ones with lower metrics values. This was expected since it happened also in [3] and also because burn expert clinicians have more difficulties to distinguish those classes. Nevertheless, they have high accuracy and suitable F_1 coefficient, precision and sensitivity to help the burn clinicians and surgeons to achieve a better diagnosis. There are no problems to distinguish class I and class IV since their metrics values have F_1 coefficient of 93.46% and 86.77%, intersection over union of 88.68% and 78.53%, precision of 93.35% and 83.96%, sensitivity of 93.86% and 92.80% respectively.

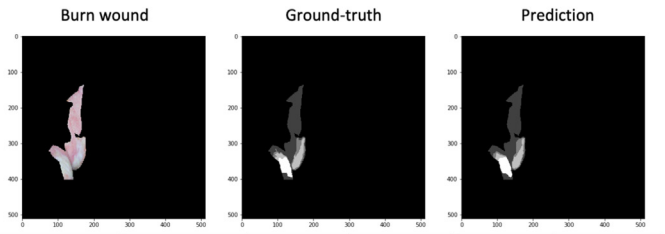
After having trained the algorithm on these 17 images the remaining 83 were examined.

3. Results

Since we did not have access to other than 83 burn-wound images which unfortunately did not contain all the burn-depth, the 17 convolutional neural networks created during the leave-one-out cross-validation needed to be used to evaluate the final set of images (n = 83). If a convolutional

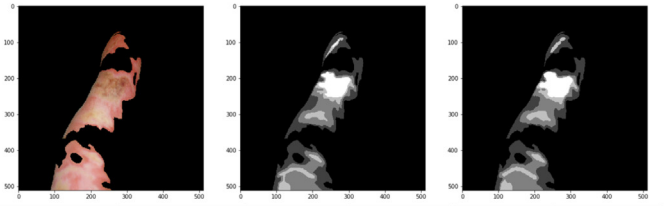
Please cite this article in press as: M.D. Cirillo, et al., Improving burn depth assessment for pediatric scalds by AI based on semantic segmentation of polarized light photography images, Burns (2021), https://doi.org/10.1016/j.burns.2021.01.011

Network 3



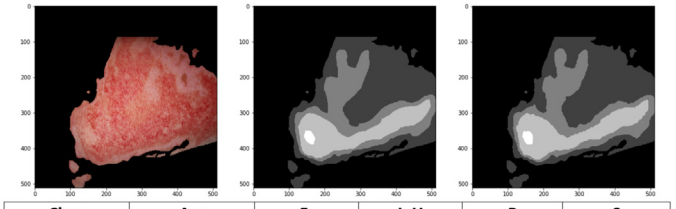
Class	Acc	F_1	IoU	P	S
I	0.9987	0.9773	0.9555	0.9694	0.9853
II	0.9981	0.7809	0.6405	0.7719	0.7900
III	0.9989	0.9032	0.8236	0.9434	0.8664
IV	0.9996	0.9586	0.9206	0.9671	0.9503

Network 10



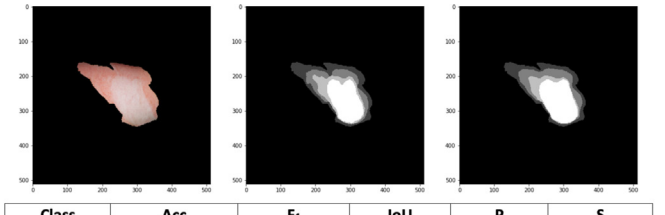
Class	Acc	F_1	IoU	P	S
I	0.9966	0.9719	0.9454	0.9748	0.9691
II	0.9934	0.9315	0.8717	0.9543	0.9097
III	0.9940	0.8497	0.7386	0.8172	0.8849
IV	0.9981	0.9387	0. 8848	0.9071	0.9725

Network 12



Class	Acc	F_1	IoU	P	S
I	0.9904	0.9781	0.9571	0.9762	0.9800
II	0.9858	0.9381	0.8834	0.9377	0.9385
III	0.9948	0.9682	0.9384	0.9742	0.9623
IV	0.9995	0.9175	0.8476	0.9102	0.9249

Network 16



Class	Acc	F_1	IoU	P	S
I	0.9956	0.9188	0.8499	0.9788	0.8658
II	0.9940	0.8480	0.7361	0.7903	0.9148
III	0.9939	0.7664	0.6212	0.8752	0.6816
IV	0.9961	0.9459	0.8974	0.9002	0.9965

Fig. 4 – Semantic segmentation results using the network 3, 10, 12 and 16 on the relative images. Each segmentation results show the burn wound image, the ground-truth and the convolutional neural network's prediction. Moreover, accuracy (Acc), F1 coefficient (F1), intersection over union (IoU), precision (P) and sensitivity (S) metrics are reported for each burn-depth class.

Please cite this article in press as: M.D. Cirillo, et al., Improving burn depth assessment for pediatric scalds by AI based on semantic segmentation of polarized light photography images, Burns (2021), https://doi.org/10.1016/j.burns.2021.01.011

Table 2 – Average of accuracy (Acc), F_1 coefficient (F_1), intersection over union (IoU), precision (P) and sensitivity (S) over all the 17 burn wound images for each burn-depth after leave-one-out cross-validation.

Class	Acc	F1	IoU	P	S
I	0.9925	0.9346	0.8868	0.9335	0.9389
II	0.9867	0.7890	0.6907	0.8423	0.7800
III	0.9763	0.7287	0.6177	0.7501	0.7464
IV	0.9806	0.8677	0.7853	0.8396	0.9280

neural network learnt how to distinguish four burn-depths in an image, it should be able to do that also in one image that does not present all of them. Accuracy and dice coefficient are reported in Table 3 for each network. From Table 3 it is possible to notice that all the networks report accuracies and dice coefficients above the 90%, with the 4-th the best one with approximately 93% for both of them.

4. Discussion

In this paper we used a modified U-Net with residuals to segment four different burn-depths (superficial partial-thickness (I), superficial to intermediate partial-thickness (II), intermediate to deep partial-thickness (III) and deep partial and full-thickness (IV)) in images generated by a high-performance light camera with polarisation filters with the aim to train the network to predict burn depth. After acquiring 100 burn images, seventeen images were used for training. Leave-one-out cross-validation reports were generated and an accuracy and dice coefficient average of almost 97% was then obtained. After that, the remaining 83 burn-wound images were evaluated using the different network during the cross-validation, achieving an accuracy and dice coefficient, both on average 92%. The F1 score, or dice score coefficient, is that metric typically used to evaluate image segmentation results because it does not consider in its equation (see Eqs. (1) and (4)) the true negatives, whereas it focuses more on the true positives and where the prediction in this clinical setting most often fails (false negative and false positives). In other words, it measures how good a predicted segmentation by the network overlaps with the "true" segmentation provided by the clinician specialist in this study made at day 20 after burn.

Table 3 – Accuracy and dice coefficient values for the remaining 83 burn wound images which do not show all the burn depths, but some of them.

Network	Accuracy	Dice coefficient
1	0.9202	0.9202
2	0.9160	0.9156
3	0.9173	0.9173
4	0.9306	0.9305
5	0.9218	0.9218
6	0.9070	0.9070
7	0.9257	0.9257
8	0.9079	0.9070
9	0.9143	0.9142
10	0.9191	0.9191
11	0.9207	0.9207
12	0.9207	0.9207
13	0.9251	0.9251
14	0.9147	0.9147
15	0.9239	0.9239
16	0.9218	0.9218
17	0.9152	0.9150
Average	0.9189 ± 0.0059	1.9188 ± 0.0060

4.1. *Related works*

Burn wounds assessment made by computer vision techniques are yet not so popular but there are some scientists who tried to investigate this field. Pinero et al. [6] identified 16 texture features for the burn image segmentation and classification. These features were then inspected by the sequential forward and backward selection methods via fuzzy-ARTMAP neural network. This method achieved an average accuracy of about 83% using 250 images, 49 × 49 pixels, divided in 5 burn appearance classes: blisters, bright red, pink-white, yellow-beige, and brown. Wantanajittikul et al. [29] used the Hilbert transform and texture analysis to extract feature vectors and then applied a support vector machine (SVM) classifier to classify burn depth. The best accuracy result for a 4-fold cross-validation was 90% using 5 images as the validation set and 34 images as the training set, and 75% correct classification on a blind test was then obtained. Acha et al. [30] applied multidimensional scaling (MDS) analysis and *k*-nearest neighbour classifier for burn-depth assessment. Using 20 images as a training set and 74 for testing, 66% accuracy was obtained for classifying burn wounds into three depths, and 84% accuracy was obtained for those that needed or did not need grafts. Serrano et al. [7] used a strict selection of texture features of burn wounds for the MDS analysis and SVM and obtained 80% accuracy in classifying those that needed grafts and those that did not. Chauhan et al. [31] used AI to classify body parts from 109 burn-wound images (30 portray burn wounds on the face, 35 on the hand, 23 on the back and 21 on the inner forearm) with size 350–450 × 300–400 pixels, achieving overall classification accuracy of 91% and 93% using a dependent and an independent convolutional neural network ResNet-50 respectively. We ourselves [3], also tried AI, similarly for the burn-depth classification. We collected 676 samples of size 224 × 224 pixels from 23 burn-wound images (almost 100 samples for each class: the four burn-depths plus the normal healthy skin and the background) and achieved an average, minimum, and maximum accuracy of 82, 72, and 88% respectively using the ResNet-101 after 10-fold cross-validation. Moreover, the average accuracy, sensitivity, and specificity were extracted for the four burn-depths: 91, 74, and 94%, respectively.

5. Study limitations

Constructing a training dataset, large volumes of study images are needed. Given the frequency of scalds, the collection of very large image databases for training purposes are not feasible and therefore the dataset used in this study may be

Please cite this article in press as: M.D. Cirillo, et al., Improving burn depth assessment for pediatric scalds by AI based on semantic segmentation of polarized light photography images, Burns (2021), https://doi.org/10.1016/j.burns.2021.01.011

claimed too small. This albeit the fact that almost two years' collection of patients were made. To improve this point, a specific image optimization technique was used (the elastic deformation technique [28]). By this measure the 16 initial training images were artificially expanded to 3936 images and thus improving the prediction metrics. Having more images for the training set is important for the further improvement of the technique.

Another study limitation is of course what is claimed "the final" healing result, and especially determining the day of total re-epithelialization used to train the prediction method. In this study we awaited the healing situation at day 20 to reduce the risk of a subjective effect on the outcome presented. However, this needs to be addressed further in coming studies.

6. Conclusion

In this paper, we wanted to extend the ambition beyond our previous publication [3], adding the local classification to the global one. As shown in the previous section, AI is a powerful tool that can be used to for the burn-depths assessment, achieving a global dice coefficient of 97% after leave-one-out cross-validation, and the average of the F_1 coefficients over all the 17 test images of 93%, 79%, 73% and 87% for superficial partial-thickness, superficial to intermediate partial-thickness, intermediate to deep partial-thickness, and deep partial and full-thickness burns respectively. These values are suitable for a better burn diagnosis since the expert clinicians on burns assess a burn wound with 75% accuracy compared to the 92% presented in this paper. Importantly it then needs to be stressed that the present paper is based on light photography images rather than laser Doppler based images. Nevertheless, the convolutional neural network performance and its metrics may surely increase with the availability of larger burn image databases. This obstacle might be overtaken with the use of Generative Adversarial Nets (GANs) [32–34] for the image augmentation on the training images. Such future improvements appear especially interesting given the accuracy and practical simplicity of the method presented.

Declarations of conflicts of interest

None.

REFERENCES

[1] Hettiaratchy S, Papini R. ABC of burns: initial management of a major burn: II—assessment and resuscitation. BMJ: Br Med J 2004;329:101.
[2] Cirillo MD, Mirdell R, Sjöberg F, Pham TD. Tensor decomposition for colour image segmentation of burn wounds. Sci Rep 2019;9(1):3291.
[3] Cirillo MD, Mirdell R, Sjöberg F, Pham TD. Time-independent prediction of burn depth using deep convolutional neural networks. J Burn Care Res 2019;40(6):857–63, doi:http://dx.doi.org/10.1093/jbcr/irz103.
[4] Jeschke MG. Burn care and treatment: a practical guide. Springer; 2013.
[5] Johnson RM, Richard R. Partial-thickness burns: identification and management. Adv Skin Wound Care 2003;16(4):178–87.
[6] Pinero BA, Serrano C, Acha JI, Roa LM. Segmentation and classification of burn images by color and texture information. J Biomed Opt 2005;10(3):034014.
[7] Serrano C, Boloix-Tortosa R, Gómez-Cía T, Acha B. Features identification for automatic burn classification. Burns 2015;41(8):1883–90.
[8] Wearn C, Lee KC, Hardwicke J, Allouni A, Bamford A, Nightingale P, et al. Prospective comparative evaluation study of Laser Doppler Imaging and thermal imaging in the assessment of burn depth. Burns 2018;44(1):124–33.
[9] Shin JY, Yi HS. Diagnostic accuracy of laser Doppler imaging in burn depth assessment: systematic review and meta-analysis. Burns 2016;42(7):1369–76.
[10] Mirdell R, Farnebo S, Sjöberg F, Tesselaar E. Accuracy of laser speckle contrast imaging in the assessment of pediatric scald wounds. Burns 2018;44(1):90–8.
[11] Mirdell R, Iredahl F, Sjöberg F, Farnebo S, Tesselaar E. Microvascular blood flow in scalds in children and its relation to duration of wound healing: a study using laser speckle contrast imaging. Burns 2016;42(3):648–54.
[12] Lindahl F, Tesselaar E, Sjöberg F. Assessing paediatric scald injuries using laser speckle contrast imaging. Burns 2013;39(4):662–6.
[13] Jaspers ME, van Haasterecht L, van Zuijlen PP, Mokkink LB. A systematic review on the quality of measurement techniques for the assessment of burn wound depth or healing potential. Burns 2019;45(2):261–81.
[14] Long J, Shelhamer E, Darrell T. Fully convolutional networks for sematic segmentation. Conference on Computer Vision and Pattern Recognition Proceedings 2015.
[15] Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell 2017;39:2481–95.
[16] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention 2015.
[17] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25(1):44.
[18] Milletari F, Navab N, Ahmadi S-A. V-net: fully convolutional neural networks for volumetric medical image segmentation. 2016 Fourth International Conference on 3D Vision (3DV) 2016.
[19] Jin Q, Meng Z, Pham TD, Chen Q, Wei L, Su R. DUNet: a deformable network for retinal vessel segmentation. Knowledge Based Syst 2019;178:149–62.
[20] Li X, Chen H, Qi XaDQ, Fu C-W, Heng P-A. H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes. IEEE Trans Med Imaging 2018;37(12):2663–74.
[21] Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, et al. Attention U-Net: learning where to look for the pancreas. arXiv 2018 preprint arXiv:1804.03999.
[22] Isensee F, Kickingereder PaWW, Bendszus M, Maier-Hein KH. No New-Net. International MICCAI Brainlesion Workshop. Springer; 2018. p. 234–44.
[23] Mirdell R, Farnebo S, Sjöberg F, Tesselaar E. Interobserver reliability of laser speckle contrast imaging in the assessment of burns. Burns 2019;45(6):1325–35.
[24] Elmasry M, Mirdell R, Tesselaar E, Farnebo S, Sjöberg F, Steinvall I. Laser speckle contrast imaging in children with scalds: its influence on timing of intervention, duration of healing and care, and costs. Burns 2019;45(4):798–804.
Grossmann B, Nilsson A, Sjöberg F, Nilsson L. Rectal ketamine during paediatric burn wound dressing

Please cite this article in press as: M.D. Cirillo, et al., Improving burn depth assessment for pediatric scalds by AI based on semantic segmentation of polarized light photography images, Burns (2021), https://doi.org/10.1016/j.burns.2021.01.011

[25] procedures: a randomised dose-finding study. Burns 2019;45 (5):1081–8.
[26] Sudre CH, Li W, Vercauteren T, Ourselin S, Cardoso MJ. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer; 2017. p. 240–8.
[27] Chollet F. Keras. [Online]. Available:. . https://keras.io.
[28] Simard PY, Steinkraus D, Platt JC. Best practices for convolutional neural networks applied to visual document analysis. 7th International Conference on Document Analysis and Recognition (ICDAR) 2003.
[29] Wantanajittikul K, Auephanwiriyakul S, Theera-Umpon N, Koanantakool T. Automatic segmentation and degree identification in burn color images. The 4th 2011 Biomedical Engineering International Conference 2012.
[30] Acha B, Serrano C, Fondón I, Gómez-Cía T. Burn depth analysis using multidimensional scaling applied to psychophysical experiment data. IEEE Trans Med Imaging 2013;32(6):1111–20.
[31] Chauhan J, Goswami R, Goyal P. Using deep learning to classify burnt body parts images for better burns diagnosis. Sipaim — Miccai biomedical workshop. Springer; 2018. p. 25–32.
[32] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. Advances in neural information processing systems. . p. 2672–80.
[33] Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: a review. Med Image Anal 2019;101552.
[34] Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J, Greenspan H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing 2018;321:321–31.

Please cite this article in press as: M.D. Cirillo, et al., Improving burn depth assessment for pediatric scalds by AI based on semantic segmentation of polarized light photography images, Burns (2021), https://doi.org/10.1016/j.burns.2021.01.011