

Time-Independent Prediction of Burn Depth using Deep Convolutional Neural Networks

Marco Domenico Cirillo, Robin Mirdell, Folke Sjöberg and Tuan Pham

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-157386>

N.B.: When citing this work, cite the original publication.

Cirillo, M. D., Mirdell, R., Sjöberg, F., Pham, T., (2019), Time-Independent Prediction of Burn Depth using Deep Convolutional Neural Networks, *Journal of Burn Care & Research*.

<https://doi.org/10.1093/jbcr/irz103>

Original publication available at:

<https://doi.org/10.1093/jbcr/irz103>

Copyright: Oxford University Press (OUP) (Policy B - Oxford Open Option B)

<http://www.oxfordjournals.org/>



Time-Independent Prediction of Burn Depth using Deep Convolutional Neural Networks

Abstract

We present in this paper the application of deep convolutional neural networks, which are a state-of-the-art artificial intelligence (AI) approach in machine learning, for automated time-independent prediction of burn depth. Colour images of four types of burn depth injured in first few days, including normal skin and background, acquired by a TiVi camera were trained and tested with four pre-trained deep convolutional neural networks: VGG-16, GoogleNet, ResNet-50, and ResNet-101. In the end, the best 10-fold cross-validation results obtained from ResNet- 101 with an average, minimum, and maximum accuracy are 81.66%, 72.06% and 88.06%, respectively; and the average accuracy, sensitivity and specificity for the four different types of burn depth are 90.54%, 74.35% and 94.25%, respectively. The accuracy was compared to the clinical diagnosis obtained after the wound had healed. Hence, application of AI is very promising for prediction of burn depth and therefore can be a useful tool to help in guiding clinical decision and initial treatment of burn wounds.

Keywords: Burn depth, time-independent prediction, deep convolutional neural network, artificial intelligence

1. Introduction

Burn depth is conventionally categorized into three degrees and one sub- category [1]: superficial burns (first degree), superficial partial-thickness burns (second degree), intermediate partial thickness burns (second degree), and full-thickness burns (third degree). Full and deep partial thickness burns require excision and split-thickness skin graft to heal in a satisfactory way. Additional degrees of burn depth are sometimes mentioned but there is no true consensus regarding these categories.

When calculating the percentage of total body surface area (%TBSA) burnt, only superficial partial-thickness and deeper burns are included in the area calculation, whereas first degree burns (with intact epidermis) are excluded. To provide the right clinical treatment to a burn patient, the %TBSA of the burn must be calculated as it dictates the early fluid resuscitation. The actual %TBSA is also useful for later surgical planning and for estimating mortality using the revised Baux score, which has proven to be a reliable predictor of both mortality and morbidity even though the %TBSA used is estimated through clinical means [2].

It is relatively easy to identify superficial and full-thickness burns. It is more challenging for burns of superficial partial-thickness and intermediate partial thickness, because, especially with burns of a higher %TBSA where the burn depth often is heterogenous. This makes it difficult to assess the depth of the burn in the first few days and correct evaluations provided by experience clinicians are only about 67% accurate [3].

Furthermore, given limited burn treatment facilities over a large geo-graphic environment, especially in middle- and low-income countries, the demand for developing automated methods for accurate and objective assessment of burn depth has been increasingly realized in burn research. The aim is not only to avoid unnecessary surgery, but also to identify correct patients and areas for surgery so optimal treatment can be undertaken without delay. The risk for infection caused by burn wounds increases with time and can complicate the infection if left without appropriate treatment. Therefore, the ability for early identification of burn depth is needed as it can avoid complications and improve the final outcome for the burn patient.

Several imaging systems have been used to evaluate burn depth during the first few days after injury. These systems measure the perfusion in the microcirculation that can be

used to determine burn depth. The most prominent imaging system for burn assessment to date is the laser Doppler imaging (LDI) [4], but only a few burn centres around the world have adopted this imaging system. Laser speckle contrast imaging (LSCI) is a newer imaging system, which also measures perfusion [5, 6, 7]. These devices, which are expensive and mainly available in burn centres, require several days to identify burn depth with reliable accuracy.

For all these reasons, it is highly desirable to develop cost-effective, faster, and objective methods for assessing burn depth so that the utilization of these methods is not limited to specialized burn clinics. Several automated techniques for burn wound analysis have been introduced in literature based on texture analysis of the burn images [8, 9, 10, 11, 12, 13, 14, 15, 16].

Pinero et al. [10] identified 16 texture features for the burn image segmentation and classification. These features are then inspected by the sequential forward and backward selection methods via fuzzy-ARTMAP neural network. This method achieved an average accuracy about 82.26% using 250 images, 49×49 pixels, divided in 5 burn appearance classes: blisters, bright red, pink-white, yellow-beige, and brown. Wantanajittikul et al. [12] 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 89.29% using 5 images as the validation set and 34 images as the training set, and 75.33% correct classification on a blind test was obtained. Acha et al. [14] 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, 66.2% accuracy was obtained for classifying burn wounds into three depths, and 83.8% accuracy was obtained for those that needed or did not need grafts. Serrano et al. [15] used a strict selection of

texture features of burn wounds for the MDS analysis and SVM, and obtained 79.73% accuracy was obtained in classifying those that needed grafts and those that did not. In this study, we applied deep convolutional neural networks (CNNs) for time-independent prediction of burn depth using colour images of burn wounds obtained from a cohort of paediatric patients. This paper is structured as follows. Section 2 describes the materials and methods. Section 3 reports burn-depth classification results and discussion on the use of four deep CNNs. Finally, Section 4 summarizes the research findings.

2. Materials and Method

2.1. Image acquisition

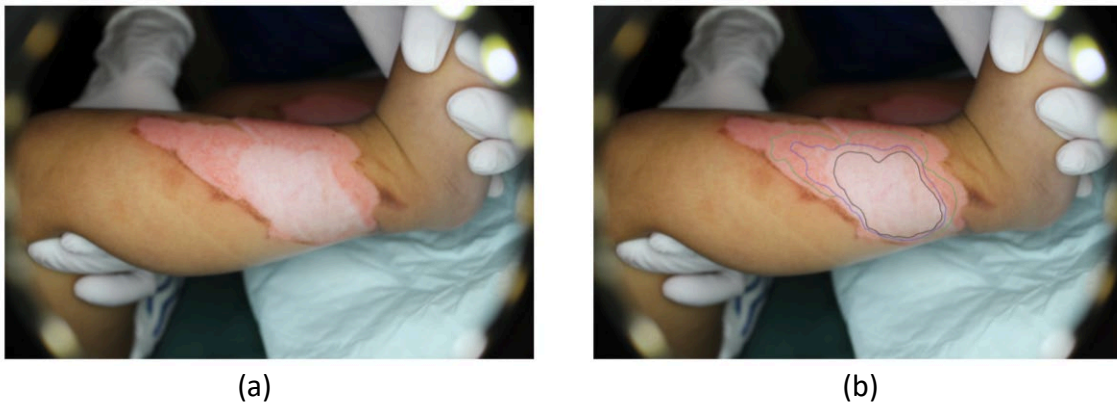


Figure 1: Original burn wound image (a) and the same labelled with different types of burn depth by a burn clinician (b): deep partial-thickness and full-thickness depth is marked by a black line, intermediate to deep partial-thickness by a blue line, superficial to intermediate partial-thickness by a green line; and the remaining unmarked burns is superficial partial-thickness.

In this study, 23 burn images in RGB colour space, each of size 3456×2304 pixels, were obtained from paediatric patients admitted to the Burn Centre of the Linköping University Hospital, Sweden. 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.

All images were taken using a TiVi700, which is a tissue viability imaging device (WheelsBridge AB, Sweden). TiVi700 is a high-performance digital camera equipped with

polarization filters and several lights all around its lens to avoid the reflecting artefact due to room light and/or the camera flash with burn moisture. Each image was manually labelled by a burn clinician of the Burn Centre to provide different areas of burn depth. Figure 1 shows a TiVi-captured image of a paediatric patient with a burn wound 19 hours old, containing different types of burn depth and located on the right leg. Figure 1b shows the location of different burn depth areas: deep partial-thickness and full-thickness depth are marked with a black line, intermediate to deep partial-thickness with a blue line, superficial to intermediate partial-thickness with a green line, and the remaining unmarked burns are the superficial partial-thickness type. Table 1 shows the numbers of burn-depth types acquired at various times from the 23 burn images.

Table 1: Burn-depth types acquired at various times from the 23 burn images, where superficial partial-thickness is defined as burn that heals within 7 days, superficial to intermediate partial-thickness heals between 8-13 days, intermediate to deep partial-thickness heals within 14-20 days, and deep or full thickness burn heals >> 21 days or underwent surgery.

Burn depth	< 2 days	2 - 4 days	5 - 7 days
Superficial partial-thickness	7	9	10
Superficial to intermediate partial-thickness	7	9	9
Intermediate to deep partial-thickness	6	8	9
Deep partial and full-thickness	2	4	3

2.2. Ground truth of burn depth

Different types of burn depths defined on the images were drawn based on the healing process and corresponds to the healing potential of the wound. In most patients the healing time could be established for different areas with a margin of error 2 days. Many of the patients also had perfusion images of the wounds, which were also used to evaluate healing time and to create a better understanding of heterogenous wounds. When perfusion images

were not available, the journal of the patient was used to establish the healing time or surgery of the most central area. All this information was put together to as objectively as possible to draw the lines corresponding to the observed healing time for each patient. All wound areas were classified into the following four categories: superficial partial-thickness healed within 7 days, superficial to intermediate partial-thickness healed between 8-13 days, intermediate to deep partial-thickness healed within 14-20 days, and deep or full thickness burn healed \gg 21 days or underwent surgery.

2.3. Burn image database

The original 23 images were used to extract 676 different regions of interest, each of size 224 x 224 pixels, where 119 samples belong to superficial, 120 to superficial partial-thickness, 108 to intermediate partial-thickness, 111 to deep partial-thickness and full-thickness, 111 to normal skin, and 107 to the background (bed blanket, nurse's gloves, scissors, etc).

Figure 2 shows an example of samples extracted from the original images that contain the four types of burn depth.

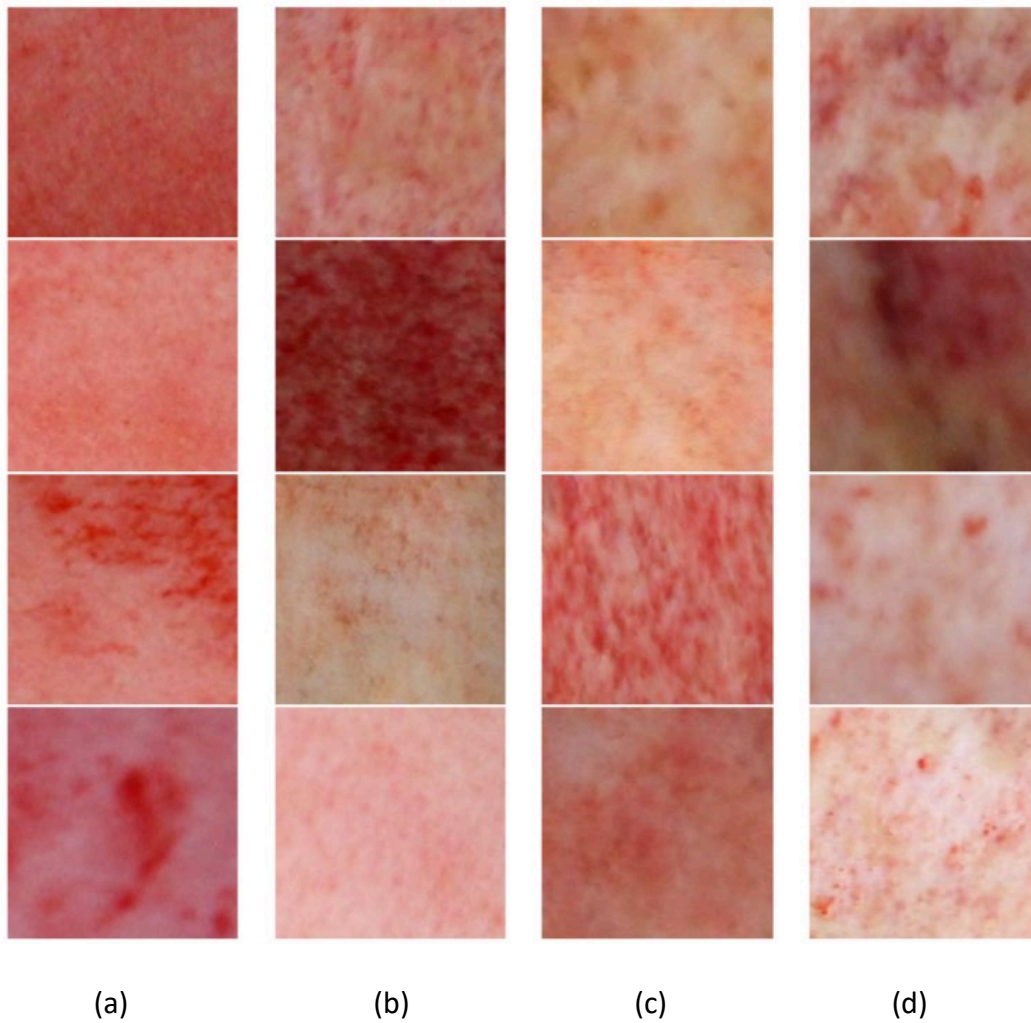


Figure 2: Example of extracted samples of four types of burn depth, each having 224×224 pixels in size: (a) superficial partial-thickness, (b) superficial partial-thickness, (c) intermediate partial-thickness, and (d) deep partial and full thickness.

2.4. Deep CNNs

Deep CNNs are the most advanced machine-learning approach in deep learning. CNNs are attractive because they can learn, recognize, classify or predict objects and patterns directly from raw data without the requirement for the selection of a feature extractor. However, a very huge database is needed for training such networks that can be used in practice with high precision. With a small data set, it is possible to resort to pre-trained deep CNN models that have already been trained with million natural images and use them as a starting point to learn a new task. This is called “transfer learning”, which means training again a pre-

trained CNN with a new image database of interest, keeping its weights and biases initialization and its architecture, changing only its last layers (fully connected, softmax and classification output layer) in order to classify the images with the number of labels under analysis, 6 in this case: superficial partial-thickness, superficial to intermediate partial-thickness, intermediate to deep partial thickness, deep partial and full-thickness, normal skin and background.

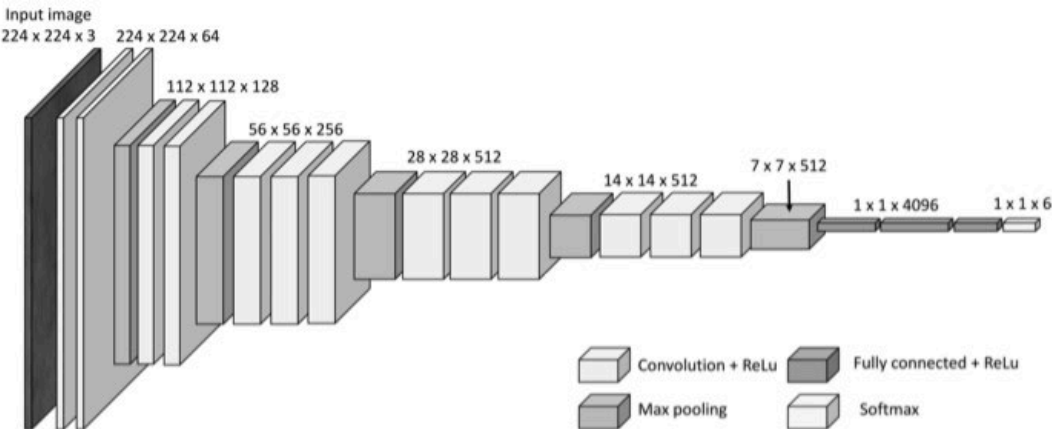


Figure 3: Graphical representation of pre-trained convolutional neural network VGG-16 further trained with the burn image dataset. An input burn sample $224 \times 224 \times 3$ is converted after several operations into a 3D array $1 \times 1 \times 6$, where 6 is the number of classes to be classified.

There exist several pre-trained convolutional neural networks in the public domain, such as AlexNet [17], GoogleNet [18], VGG-16, VGG-19 [19], ResNet-18, ResNet-50, resNet-101 [20], U-Net [21], inception-v3 [22], inception- v4 [23]. As an example, Figure 3 shows a graphical representation of the VGG-16 that has 41 layers to compute 2D convolutions, and at the end, a fully connected layer and the classification output. It also illustrates how a 3D array $224 \times 224 \times 3$ is converted at the end into a 3D array $1 \times 1 \times 6$, where 6 is the number of the classes to distinguish for this study as described in Section 2.3. Figure 4 shows the first 6 features learned by the 2nd, 4th, 7th, 9th, 12th, 16th, 19th, 23rd, 26th and 30th convolutional layer of the VGG-16 CNN being trained again with the new burn images. It is

can be observed that the deeper the layer is the more complicated and defined the CNN features are. These features appear very different from conventional texture features [8, 10, 12, 15].

To enhance the training of the pre-trained CNNs with the new data of burn, the data augmentation was applied in this study by rotating, reflecting and translating in different ways. The use of data augmentation can prevent the networks from overfitting and memorizing the exact details of the training images.

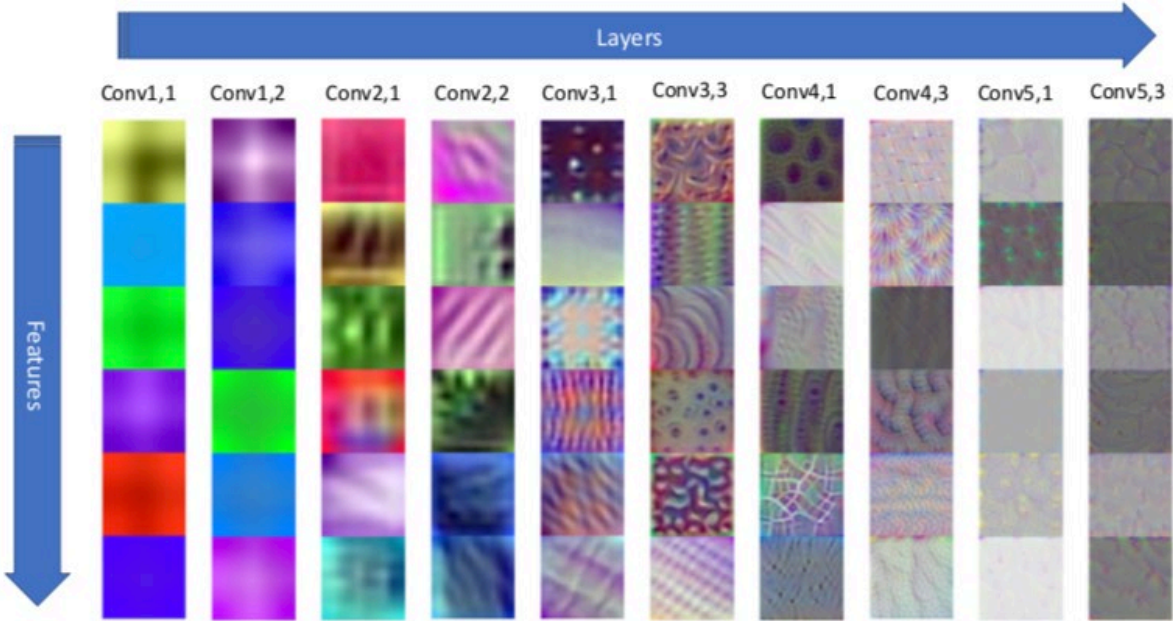


Figure 4: Visualization of the first 6 features learned by the 2nd, 4th, 7th, 9th, 12th, 16th, 19th, 23rd, 26th and 30th convolutional layer of the trained VGG-16 CNN.

3. Results and discussion

Table 2 describes the 6 classes and their sample sizes, where each image is of size 224 × 224 pixels, extracted for each class used for training and testing the pre-trained CNN models.

Table 2: Classes and samples of different types of burn depth.

Class	Description	Number of Samples
1	Superficial partial-thickness	119
2	Superficial to intermediate partial-thickness	120
3	Intermediate to deep partial-thickness	108
4	Deep partial and full-thickness	111
5	Normal, uninjured skin	111
6	Background	107

Each pre-trained CNN model was trained using mini batches of size = 64, epochs = 10, and learning rate = 0.001. The 10-fold cross-validation was applied to test the performance of each pre-trained CNN model. Table 3 shows the average, minimum, and maximum accuracy (ACC) rates of the four pre-trained CNN models.

The best CNN model for classifying the burn depth is the ResNet-101, whose average, minimum, and maximum ACC rates and MSE are 81.66%, 72.06%, 88.06%, and 58.75%, respectively. Both ResNet-50 and VGG-16 achieved accuracy rates better than the assessment by an experienced clinician at Burn Centers (about 64-76% [24]), whereas the accuracy obtained from the GoogleNet is comparable with that by the experienced clinician.

Table 3: Average (Avg), minimum (Min), and maximum (Max) accuracy (ACC) rates and mean-square errors obtained from 10-fold cross-validation of four pre-trained CNN models.

CNN	Avg ACC	Min ACC	Max ACC
Vgg-16	0.7753	0.7059	0.8657
GoogLeNet	0.7380	0.6567	0.7941
ResNet-50	0.7779	0.7313	0.8235
ResNet-101	0.8166	0.7206	0.8806

Table 4 shows the average of 10 different confusion matrices calculated with 10 runs of the 10-fold cross-validation for the ResNet-101. The diagonal cells represent the correct prediction, whereas the off-diagonal cells indicate misclassified samples. It can be noticed that the background (class 6) was almost perfectly detected, there is 0.2% mismatch with class 4. The normal skin (class 5) is also almost perfectly detected, returning the total percentage mismatch below the 0.73%. For the burn classes (classes 1 to 4), ResNet-101 predicted well for each class, with a maximum percentage of mismatch about 3.25% when the burn is intermediate partial-thickness but the network predicted superficial partial-thickness. ResNet-101, which achieved the highest accuracy (88.06%), predicted class 6 without any mis-classification, and similarly for class 5 with a mismatch 1.5% for class 1, whereas the maximum mismatch is 4.5% when the burn belongs to the class 2 but the network detected as class 1, and the others resulted in mismatches of 1.5%.

The confusion matrix showed in Table 4 can be changed into a binary one in order to identify the true positive (TP), true negative (TN), false positive (FP) and false negative (FN) for each burn depth. Based on these 4 binary matrices, other important values can be calculated, including accuracy (ACC), error rate (ER), sensitivity (SN), specificity (SP),

precision (P), and false positive rate (FPR). All these values are reported in Table 5 for each type of burn depth using the ResNet-101 model.

Table 4: Average confusion matrix of 10 different runs of 10-fold cross-validation using ResNet-101.

		Prediction					
Actual	Class	1	2	3	4	5	6
	1	8.8 13.01%	2 2.95%	0.4 0.59%	0.2 0.29%	0.5 0.73%	0 0%
	2	2.1 3.10%	8 11.83%	1.1 1.62%	0.8 1.18%	0 0%	0 0%
	3	0.3 0.44%	2.2 3.25%	7.4 10.94%	0.6 0.88%	0.3 0.44%	0 0%
	4	0.2 0.29%	0.5 0.73%	0.5 0.73%	9.8 14.49%	0.1 0.14%	0 0%
	5	0.1 0.14%	0.1 0.14%	0 0%	0.2 0.29%	10.7 15.82%	0 0%
	6	0 0%	0 0%	0 0%	0.2 0.29%	0 0%	10.5 15.53%

Table 5 shows the classification results obtained from ResNet-101, where average accuracy is 90.54%, 9.45% as error rate, 74.35% sensitivity, 75.19% precision, 5.74% false positive rate, and 94.25% specificity. This means that the ResNet-101 CNN can distinguish with high accuracy each burn depth.

Furthermore, Table 6 shows the results obtained from ResNet-101 with the data augmentation described earlier. The technique is applied three times with different type of operations: random rotation between $[-45^\circ, 45^\circ]$, random rotation between $[-10^\circ, 10^\circ]$, and random reflection on X and/or Y axis. Comparing these results with the ones obtained above by ResNet-101 without augmentation in Table 3, it is easy to conclude that the random rotations decrease the accuracy but still present a good accuracy between about 90% and 87%. On the other hand, data augmentation made by reflection on X and/or Y axis achieves an accuracy between 75% and 88%. This is probably caused by the fact that reflections do not require interpolations as rotations, except for rotations with 90° , 180° and 270° . Data augmentation with rotation and reflection can improve the accuracy.

Table 5: Accuracy (ACC), error rate (ER), sensitivity (SN), specificity (SP), precision (P), and false positive rate (FPR) for each burn class predicted by ResNet-101.

Class	ACC	ER	SN	SP	P	FPR
1	0.9049	0.0951	0.7395	0.9450	0.7652	0.0550
2	0.8625	0.1375	0.6667	0.9077	0.6250	0.0923
3	0.9109	0.0891	0.6852	0.9598	0.7872	0.040
4	0.9436	0.0564	0.8829	0.9578	0.8305	0.4222
Average	0.9054	0.0945	0.7435	0.9425	0.7519	0.0574

Table 6: Average (Av), minimum (Min), maximum (Max) accuracy obtained from 10-fold cross-validation with 10 epochs for each training for the resNet-101 CNN with and without data augmentation.

CNN	Avg ACC	Min ACC	Max ACC
ResNet-101 with random rotation between [-45°, 45°]	0.7766	0.6618	0.8971
ResNet-101 with random rotation between [-10°, 10°]	0.7868	0.7015	0.8676
ResNet-101 with random reflection on X and/or Y axis	0.8286	0.7500	0.8806
ResNet-101 with rotation 90° and random reflection on X and/or Y axis	0.8123	0.7059	0.8657

From the clinical point of view, it is much easier to assess the depth of the burn within 2-4 days after injury with an accuracy between 60-80%. However, assessment within the first 24 hours, the accuracy is about 50%. It was reported that burn sensitivity and specificity were 92.3% and 78.3%, respectively, between 0-24 hours, 100% and 90.4%, respectively, between 72-96 hours after injury, and 100% for both when combining the two measurements into a modified perfusion model [7]. This means that the method used in [7] can detect different types of burn depth only after 72-96 hours with the perfusion measured between 0-24 hours after injury. If this condition for measurement is not applicable, the combination cannot be done, and as a consequence, the classification cannot be obtained.

The burn image database used in this study is time-independent because it contains samples of different types of burn depth taken at different times and the pre-trained ResNet-101 CNN model was able to distinguish the four different types of burn depths in a few seconds with very high accuracy and specificity.

In this study, we evaluated all images from a healing potential perspective. Healing time was not based on initial clinical inspection, but rather on clinical inspection of reepithelialization often several days or weeks after injury. Clinical assessment of reepithelialization is generally regarded as a good method to calculate healing time. The person who did the marking of the images was often present during the dressing changes. In many cases, we also had several images of the healing process making a high degree of precision possible. Additionally, the person marking the images often had several LSCI-recordings to further increase the precision. For all cases we also used the patient's medical journal to check and compare our conclusions. All in all, we can be quite certain that our markings are close to the true healing time and should not be off by more than 2 days, seldom causing misclassification of wounds in material. Most of the wounds were scanned with an LSCI system, which is quite similar to LDI. LSCI was often used to better understand boundaries within heterogenous wounds.

As clinicians what is most relevant in the diagnosis of depth is the healing potential particularly within the 2-3 week window that predicts the risk of subsequent hypertrophic scarring. Hence an assessment of the AI used against healing potential and outcome would be most valuable. In Table 5, a breakdown can be seen of the accuracies for the different classes. For example, wounds healing between 2-3 weeks had a 91.09% accuracy and 94.36% for >21 days/surgery cases.

Since we have investigated wounds that are both just a few hours old and wounds that were up to 7 days old, several of the included wounds would have exhibited plenty of edema. The degree of this will of course affect the outcome, and it is one out of several other factors that keeps us from an accuracy closer to 100%. The most difficult aspect is likely the sub-clinically infected burn wound which could have a large effect on the healing

time and perhaps cause an overt burn wound infection. Regarding how the proposed classification system works in different parts of the body, we were quite liberal with the inclusion of burns on different body parts. As such, our system has similar accuracy in all body parts. In this study, the system was not trained to particularly target burns on different body parts. There are, however, a few body parts which can prove a bit different and which we did not have as many cases of. The first being soles of the feet and palms of the hands. The skin is a lot thicker in these parts and the microcirculation also has a different architecture. The second being the face, where we only had a limited amount of burns and some areas such as cheeks have a quite different vascular density compared to many other areas. The last area, which we have little information about, is the genital area as we were not able to include any such burns in our study.

4. Conclusion

Application of AI with state-of-the-art convolutional neural networks is very promising for prediction of burn depth and therefore can be a useful tool to help in guiding clinical decision and initial treatment of burn wounds. The ResNet-101 CNN was able to classify four types of burn depth in few seconds with an average accuracy of 91% and specificity of 94%. It would be of interest to implement more advanced networks such as inception-v3 [22], inception-v4 [23], but they require the use of a supercomputer for the training process. The natural step forward can be the implementation of the semantic segmentation [21, 25, 26] in order to have not only a global information of what a picture is showing but also the local information of the labels of interest. Semantic segmentation performs pixel-by-pixel classification and does not require a huge number of training images as for networks that perform global classification. As a result, burn clinicians will be informed not only the depth

of a burn wound but also its area. This would allow to measure the area of each burn depth in the acquired image for making surgical decision.

REFERENCES

- [1] S. Hettiaratchy, R. Papini, ABC of burns: Initial management of a major burn: II—assessment and resuscitation, *BMJ: British Medical Journal* 329 (2004) 101.

- [2] I. Steinvall, M. Elmasry, M. Fredrikson, F. Sjöberg, Standardised mortality ratio based on the sum of age and percentage total body surface area burned is an adequate quality indicator in burn care: An exploratory review, *Burns* 42 (2016) 28–40.

- [3] R. M. Johnson, R. Richard, Partial-thickness burns: identification and management, *Advances in Skin & Wound Care* 16 (2003) 178–187.

- [4] F. Kloppenberg, G. Beerthuisen, H. Ten Duis, Perfusion of burn wounds assessed by laser doppler imaging is related to burn depth and healing time, *Burns* 27 (2001) 359–363.

- [5] F. Lindahl, E. Tesselaar, F. Sjöberg, Assessing paediatric scald injuries using laser speckle contrast imaging, *Burns* 39 (2013) 662–666.

- [6] R. Mirdell, F. Iredahl, F. Sjöberg, S. Farnebo, E. Tesselaar, Microvascular blood flow in scalds in children and its relation to duration of wound healing: A study using laser speckle contrast imaging, *Burns* 42 (2016) 648–654.

- [7] R. Mirdell, S. Farnebo, F. Sjöberg, E. Tesselaar, Accuracy of laser speckle contrast imaging in the assessment of pediatric scald wounds, *Burns* 44 (2018) 90–98.
- [8] M. D. Cirillo, R. Mirdell, F. Sjöberg, T. D. Pham, Tensor decomposition for colour image segmentation of burn wounds, *Scientific Reports* 9 (2019) 329.
- [9] T. D. Pham, M. Karlsson, C. M. Andersson, R. Mirdell, F. Sjöberg, Automated vss-based burn scar assessment using combined texture and color features of digital images in error-correcting output coding, *Scientific reports* 7 (2017) 16744.
- [10] B. A. Pinero, C. Serrano, J. I. Acha, L. M. Roa, Segmentation and classification of burn images by color and texture information, *Journal of Biomedical Optics* 10 (2005) 034014.
- [11] H. Wannous, S. Treuillet, Y. Lucas, Robust tissue classification for reproducible wound assessment in telemedicine environments, *Journal of Electronic Imaging* 19 (2010) 023002.
- [12] K. Wantanajittikul, S. Auephanwiriyaikul, N. Theera-Umpon, T. Koanantakool, Automatic segmentation and degree identification in burn color images, in: *Biomedical Engineering International Conference (BMEiCON)*, IEEE, pp. 169–173.
- [13] R. Mukherjee, D. D. Manohar, D. K. Das, A. Achar, A. Mitra, C. Chakraborty, Automated tissue classification framework for reproducible chronic wound assessment, *BioMed research international* 2014 (2014).

- [14] B. Acha, C. Serrano, I. Fondón, T. Gómez-Cía, Burn depth analysis using multidimensional scaling applied to psychophysical experiment data, *IEEE Transactions on Medical Imaging* 32 (2013) 1111–1120.
- [15] C. Serrano, R. Boloix-Tortosa, T. Gómez-Cía, B. Acha, Features identification for automatic burn classification, *Burns* 41 (2015) 1883–1890.
- [16] J. Kawahara, A. BenTaieb, G. Hamarneh, Deep features to classify skin lesions, in: 13th International Symposium on Biomedical Imaging (ISBI), IEEE, pp. 1397–1400.
- [17] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: *Advances in Neural Information Processing Systems*, pp. 1097–1105.
- [18] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: *Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 1–9.
- [19] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556* (2014).
- [20] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 770–778.

- [21] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, pp. 234–241.
- [22] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, pp. 2818–2826.
- [23] C. Szegedy, S. Ioffe, V. Vanhoucke, A. A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in: Thirty- First AAAI Conference on Artificial Intelligence, AAAI, pp. 4278–4284.
- [24] M. S. Badea, C. Vertan, C. Florea, L. Florea, S. Badoiu, Automatic burn area identification in color images, in: International Conference on Communications (COMM), IEEE, pp. 65–68.
- [25] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Conference on Computer Vision and Pattern Recognition Proceedings, IEEE, pp. 3431–3440.
- [26] V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (2017) 2481–2495.