Advances in Informatics, Management and Technology in Healthcare
J. Mantas et al. (Eds.)

doi:10.3233/SHTI220717

Automatic Wound Type Classification with Convolutional Neural Networks

Leila MALIHI[e], Jens HÜSERS[a,1], Mats L. RICHTER[e], Maurice MOELLEKEN[c], Mareike PRZYSUCHA[a], Dorothee BUSCH[d], Jan HEGGEMANN[b], Guido HAFER[b], Stefan WIEMEYER[b], Gunther HEIDEMANN[e], Joachim DISSEMOND[c], Cornelia ERFURT-BERGE[d] and Ursula HÜBNER[a]

[a] *Health Informatics Research Group, Osnabrück University of AS, Germany*
[b] *Christian Hospital Melle, Niels Stensen Hospitals, Germany*
[c] *Department of Dermatology, Venerology and Allergology, University Hospital of Essen, Germany*
[d] *Department of Dermatology, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nürnberg, Germany*
[e] *Institute of Cognitive Science, Osnabrück University, Germany*

Abstract. Chronic wounds are ulcerations of the skin that fail to heal because of an underlying condition such as diabetes mellitus or venous insufficiency. The timely identification of this condition is crucial for healing. However, this identification requires expert knowledge unavailable in some care situations. Here, artificial intelligence technology may support clinicians. In this study, we explore the performance of a deep convolutional neural network to classify diabetic foot and venous leg ulcers using wound images. We trained a convolutional neural network on 863 cropped wound images. Using a hold-out test set with 80 images, the model yielded an F1-score of 0.85 on the cropped and 0.70 on the full images. This study shows promising results. However, the model must be extended in terms of wound images and wound types for application in clinical practice.

Keywords: Clinical Decision Support System, Health Information Technology, Diabetic Foot Ulcer, Image Classification, Wound Care, Transfer Learning, Convolutional Neural Networks

1. Introduction

Chronic wounds are ulcerations of the skin and subcutaneous tissue that fail to heal in the physiological time span. Causing immobility, necrosis, and infections, chronic wounds pose a severe health threat to patients since they are associated with high morbidity and mortality. Alas, their prevalence rises across societies. For example, in the U.K., between 2013 and 2018, prevalence increased by 66% [1].

The timely classification of the underlying clinical condition that causes the stalled wound healing is key for preventing morbidity and mortality. For example, venous leg ulcers are driven by venous insufficiency, and diabetic foot ulcers are driven by diabetes. When this underlying condition is known, causal therapy can be initiated, amplifying the odds of healing. This challenging task requires domain knowledge. However, when this

[1] Corresponding Author, Jens Hüsers, Osnabrück University of AS, Health Informatics Research Group, PO Box 1940, 49009 Osnabrück, Germany; E-Mail: j.huesers@hs-osnabrueck.de.

knowledge on wound diagnostics is unavailable, e.g., in routine care situations in non-specialised centres, technology such as artificial intelligence may support clinicians in wound classification. Artificial intelligence may accomplish this classification task based on deep neural networks that use wound images as a data source. Similar has been done for similar medical problems like classifying skin cancer [2] or wound infections [3].

This study hypothesises that artificial intelligence technology can classify a wound according to its underlying condition. We test this hypothesis on the two most prevalent wound types, i.e., diabetic foot ulcer and venous leg ulcer [1].

2. Methods

For this study, we compiled a dataset of images showing either diabetic foot ulcers or venous leg ulcers. The images originated from two wound care centres that specialised in either wound type. Selected from the wound documentation, the images showed wounds without cremes, gels, and dressing remains. The wounds on the images had different healing stages and provided diverse wound characteristics, among them signs of maceration, infection, or necrosis. Additionally, we did not standardise the camera angle or distance between the camera and wound. We excluded images that did not show the complete wound. We split the images into a training, validation, and test set. The test set was then extended using available openly available images that confirm to the inclusion criteria. All images were labelled by a clinician with experience in wound care using bounding boxes to locate the images. Then, the images were cropped using those bounding boxes. As a result, the final image contained only the wound without any background or anatomical landmarks.

For image classification, we used the xception model as the deep convolutional neural network with pre-trained weights [4]. In a previous study, this model was successfully deployed for a similar classification task on skin lesions and was found to be efficient and reliable [5]. For training, we used pre-trained weights based on the imagenet dataset. Furthermore, we froze the weights on the first two layers to transfer the information on the feature extraction of the pre-trained weights. The final model had 21.2 Mio parameters, of which 55 thousand were frozen and thus not trainable.

For model training, we rescaled all images to 224 to 224 pixels. Furthermore, we applied data augmentation by randomly shifting and shearing the image and manipulating the image's brightness to increase the model's external validity. We used the Python (3.9) version of the open-source software library TensorFlow (2.8).

For model evaluation, we calculated precision (sensitivity), specificity, recall (positive predictive value), accuracy, and the F1-score. We evaluated this model twice: for the test sets of the cropped and the uncropped images. In the evaluation, the diabetic foot served as the reference category.

3. Results

The dataset used for training contained 909 wound images. The dataset contained 448 images of diabetic foot ulcers and 461 of venous leg ulcers. The images' average raw height and width were 2705 and 3374 pixels, respectively. After cropping, the number of images in the dataset increased to 990 because some of them depicted more than one wound. The training set contained 863 images (87%), the validation set 47 images (5%), and the hold-out test set 80 images(8%). The latter contained 31 openly available wound images. The evaluation of the 80 test images yielded higher performance metrics on the cropped images, e.g., the F1-score for the cropped images was 0.85 vs. 0.70 of the uncropped images. On the cropped images, the model showed a high recall (sensitivity) of 0.94; thus, when an image shows a diabetic foot ulcer, it is detected in 94 per cent of all cases. When an image shows a venous leg ulcer, it is detected in 75 per cent of all cases (specificity). Evaluated on the full images, the model's performance decreases to 0.65 (recall) and 0.70 (specificity) (Table 1).

Table 1. Evaluation metrics for the test dataset with additional external available images. Metrics in brackets represent the results on the test set alone (without the addional external images)

Test-Dataset	Accuracy	Recall (Sensitivity)	Specificity	Precision (PPV)	F1-Score
Cropped Images	0.83 (0.90)	0.94 (1.00)	0.74 (0.79)	0.75 (0.79)	0.85 (0.91)
Full Images	0.67 (0.85)	0.65 (0.84)	0.70 (0.81)	0.65 (0.87)	0.70 (0.83)

4. Discussion

This study tested the hypothesis that artificial intelligence technology driven by deep convolutional neural networks and using wound images can classify the wound type. The model yielded promising preformance, especially for the cropped images, as the F1-score (0.85) and the recall is high (0.94). The model's performance was poorer for the full uncropped images most likely because it was trained on the cropped images. Higher performance on the full uncropped images would allow to skip the extra step of cropping done by a clinician. However, the results suggest that this step is beneficial.

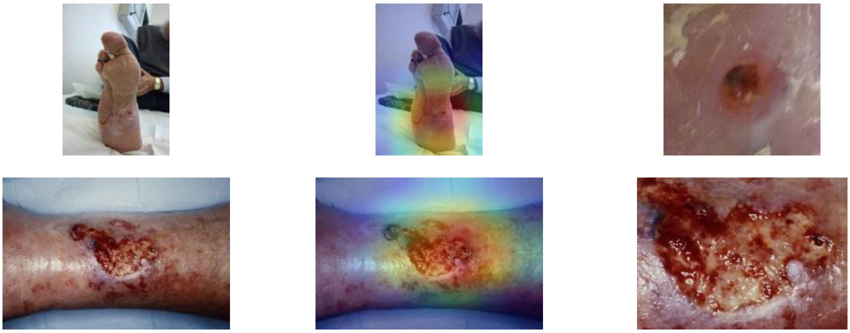

Figure 1. *Left*: Full Image *Middle*: Full Image with GradCam *Right*: Cropped Images; *Upper Row*: DFU *Lower Row*: Venous Leg Ulcer

To evaluate whether to further pursue the goal of wound classification in full uncropped images, it is critical to understand if the model is able to focus on the salient features. Deep neural networks like the xception model are highly complex and are considered black boxes, and understanding them is hard. One possible approach gaining insights is to use gradient class activation mapping (Grad-CAM) [6]. Grad-CAM is an algorithm visualizing the salient features in the last convolutional layer that mainly drives the final classification (Figure 1). This approach may not only provide insights but also further facilitate a clinician's understanding of the models' inference and thus builds trust in the application. Another approach are object detectors which are convolutional neural networks that can simultaneously locate and classify objects. A study on the detection of diabetic foot ulcers shows promising results. Thus, this approach should be investigated futher. Yet, in this study, we decided to evaluate classification models trained only on the cropped wound itself to test the ability to infer the wound type using just the wound itself, thereby reducing leaky information in the images, such as facility-specific background entities and anatomical landmarks.

This study has limitations. First, we used a relatively small dataset for a model that is considered data intensive. So, adding more data may improve the model's performance by reducing epistemic uncertainty. However, it remains unclear if this strategy is successful, as the aleatoric uncertainty must be considered high because when physicians visually inspect the wound when making a diagnosis they consider further information such as the wound's body site and further biomarkers. Another limitation is that we selected just two wound types in this study. However, a system used in clinical practice supporting clinicians must consider all wound types, including rare wound types such as the pyoderma gangraenosum. Thus, a system for actual clinical use should be the goal.

This study presents a deep convolutional neural network trained to classify two types of wound images and extends the toolbox of automatic wound classification, e.g. tools are available that detect wound characteristics like macerations in wound images [7]. The performance metrics show promising results. However, the model must be extended in terms of wound images and wound types for application in clinical practice.

Funding

This study was funded by the German Federal Ministry of Education and Research (BMBF) (grant: 16SV8616).

References

[1] Guest JF, Fuller GW, Vowden P. Cohort study evaluating the burden of wounds to the UK's National Health Service in 2017/2018: update from 2012/2013. BMJ Open. 2020;10:e045253.

[2] Haggenmüller S, Maron RC, Hekler A, Utikal JS, Barata C, Barnhill RL, Beltraminelli H, Berking C, Betz-Stablein B, Blum A, Braun SA. Skin cancer classification via convolutional neural networks: systematic review of studies involving human experts. European Journal of Cancer. 2021 Oct 1;156:202-16.

[3] Cassidy B, Kendrick C, Reeves ND, Pappachan JM, O'Shea C, Armstrong DG, Yap MH. Diabetic foot ulcer grand challenge 2021: evaluation and summary. InDiabetic Foot Ulcers Grand Challenge 2021 Sep 27 (pp. 90-105). Springer, Cham.

[4] Chollet F. Xception: Deep learning with depthwise separable convolutions. Proc IEEE Conf Comput Vis Pattern Recognit. 2017. p. 1251–1258.

[5] Moataz L, Salama GI, Abd Elazeem MH. Skin Cancer Diseases Classification using Deep Convolutional Neural Network with Transfer Learning Model. J Phys Conf Ser. 2021;2128:012013.

[6] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. InProceedings of the IEEE international conference on computer vision 2017 (pp. 618-626).

[7] Hüsers J, Hafer G, Heggemann J, Wiemeyer S, John SM, Hübner U. Development and Evaluation of a Bayesian Risk Stratification Method for Major Amputations in Patients with Diabetic Foot Ulcers. InInformatics and Technology in Clinical Care and Public Health 2022 (pp. 212-215). IOS Press.