Challenges of Trustable AI and Added-Value on Health
B. Séroussi et al. (Eds.)

doi:10.3233/SHTI220397

An Image Based Object Recognition System for Wound Detection and Classification of Diabetic Foot and Venous Leg Ulcers

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

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

Abstract. Venous leg ulcers and diabetic foot ulcers are the most common chronic wounds. Their prevalence has been increasing significantly over the last years, consuming scarce care resources. This study aimed to explore the performance of detection and classification algorithms for these types of wounds in images. To this end, algorithms of the YoloV5 family of pre-trained models were applied to 885 images containing at least one of the two wound types. The YoloV5m6 model provided the highest precision (0.942) and a high recall value (0.837). Its mAP_0.5:0.95 was 0.642. While the latter value is comparable to the ones reported in the literature, precision and recall were considerably higher. In conclusion, our results on good wound detection and classification may reveal a path towards (semi-) automated entry of wound information in patient records. To strengthen the trust of clinicians, we are currently incorporating a dashboard where clinicians can check the validity of the predictions against their expertise.

Keywords. Diabetic Foot Ulcer, Venous Leg Ulcer, Wound Care, Artificial Intelligence, Image Classification, Clinical Decision Support System, Health Information Technology

[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.

1. Introduction

Chronic wounds are skin and tissue lesions that fail to heal due to different underlying conditions. They have become more frequent in recent years across the world. For example, the British National Health Service witnessed a rise in chronic wound cases from 1.05 million in the year 2014 to 1.58 million in 2018 [1]. In 2018, the most common chronic wound types in the UK were the venous leg ulcer and the diabetic foot ulcer. The increase was over proportionally for both wound types: during the five years, their numbers approximately doubled as the venous leg, and the diabetic foot cases increased by 101% and 93%, respectively. Accordingly, this rise will trigger the future need for wound care and larger consumption of health resources [1].

It is paramount to identify the wound type, i.e., the underlying condition that sustains skin and tissue damage, notably insufficient leg veins for venous leg ulcers [2,3] and diabetic pathophysiology for the diabetic foot ulcer. There are clinical situations where it is the wound image that primarily counts, particularly as more and more images of wounds taken at the point of care become available. Due to the importance of these images, several attempts have been made to support wound assessment and documentation, e.g. [4], and clinical decision making, e.g. [5], through machine learning.

The aim of this study, therefore, is to explore the ability and performance of detection and classification algorithms for diabetic foot ulcers and venous leg ulcers in wound images. This study is embedded in a project on prediction models in wound healing.

2. Methods

We compiled a dataset of wound images from medical records of two specialized wound care centres: the wound care centre of the Christian Hospital Melle and the Department of Dermatology, Venerology and Allergology of the University Hospital Essen, both situated in Germany. All images were part of the wound record, taken during encounters in routine care. The wound type and the images were derived from the medical records, building the ground truth for the classification. The selected images showed at least one diabetic foot ulcer or venous leg ulcer but not both types in one image. A clinician experienced in wound care annotated the type and the corresponding bounding box that located the wound in the image. Afterwards, the annotation was subsequently checked by a second clinician. The final dataset contained 435 diabetic foot ulcers and 450 venous leg ulcers, culminating in 885 images.

The images from both sites were not standardized regarding the angles and distances of the camera. The wounds on the images had different healing stages and provided diverse wound characteristics, among them signs of maceration, infection, or necrosis. All wounds in the images were free from dressing material, and cream or gel remains but may show rulers, information cards, the clinician's hand, or objects in the background. The predominant number of images represent patients with white skin colour.

We used the models from the YoloV5 family for single-shot detection of the two types of chronic wounds. We selected YoloV5 models as they recently showed the best performance in benchmarks for more general tasks [6]. For this task, we explored four YoloV5 models: YoloV5n6, YoloV5s, YoloV5m6, and YoloV5x6. All models used pre-trained weights based on the MS COCO dataset.

The wound images were randomly assigned to a training and test split with a 90% to 10% ratio. The colour channels were scaled from 24-bits to a range of the closed interval from 0 to 1. A mere 885 images compose a very small training set for modern deep learning architectures. Therefore, the set was artificially enlarged by HSV-noising (hue, saturation value), rotating, translating, scaling, cropping, flipping along the horizontal and vertical axis, which makes the resulting classifier much more robust against factors such as different viewpoints and noise. Additionally, mosaics of different images were randomly compiled during training so that a training image may contain both wound types (Fig. 1 left). The mosaic technique reduced site-specific characteristics in combination with the augmentation mentioned above.

Standard performance metrics for single-shot object detection were calculated for all models. The model showing the highest mean average precision (mAP) with a confidence threshold of 0.5 and an intersection over union (IoU) of 0.5 was selected as the final model.

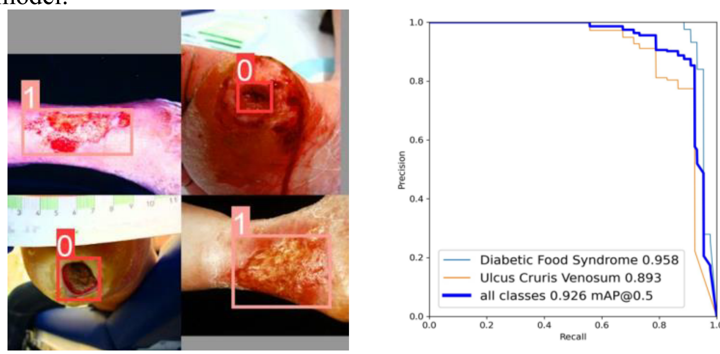

Figure 1 Left: Example of augmented mosaic used for training. In this example, the training image uses four raw images that were randomly augmented and then randomly recompiled into a single training example. The bounding boxes show two diabetic foot ulcers, labelled as 0, and two venous leg ulcers, labelled as 1 **Right**: Recall (x-axis) vs precision (y-axis) curve for the best model (YoloV5m6) for the overall and the class performance with fixed IoU of 0.5 varying confidence levels.

3. Results

Among the 885 images, 789 contained one wound, 85 included two wounds, ten images three wounds, and one image four wounds. The images' average raw width and height were 3374 pixels and 2705 pixels, respectively. All trained models showed convergence and the absence of overfitting. The mean average precision for prediction confidence of 0.5 (mAP_0.5) was at least 0.895 and at most 0.925 when the IoU was set to 0.5. Thus, the models showed low variance and performed similarly for this metric. However, the YoloV5m6, a medium-sized model, had the highest variance for this metric. Compared to the other three models, this model also had the highest precision (0.942) and a high recall value (0.837). Its mAP_0.5:0.95 was 0.642. The model's precision vs. recall curve revealed balanced scores with no sign of overfitting. Furthermore, it shows better predictions for diabetic foot ulcers than venous leg ulcers: the average precision with an IoU of 0.5 (and incrementing confidence levels) was 0.958 and 0.893 (Fig 1 right).

Table 1. Performance metrics per model. The mAP_0.5:0.95 had a fixed confidence at 0.5 and incrementing IoU thresholds for 0.50 to 0.95 in steps of 0.05.

Model	Number of parameters in mio	Recall	Precision	mAP_0.5:0.95	mAP_0.5	F1-score
YoloV5n6	1.9	0.879	0.894	0.638	0.917	0.89
YoloV5s6	7.2	0.908	0.852	0.671	0.920	0.88
YoloV5m6	21.2	0.837	0.942	0.642	0.925	0.89
YoloV5x6	86.7	0.820	0.892	0.652	0.895	0.85

4. Discussion

This study investigated an image-based object recognition system for single-shot wound detection and classification from the YoloV5 family. All models showed satisfying predictive performance, precision, recall, mAP_0.5, and mAP_0.5:0.95. The model with the highest mAP_0.5 metric was the YoloV5m6 model, the medium-sized model with 21.2 million parameters.

This model had a recall (sensitivity) of 0.837, i.e., when a specific wound type is present according to the ground truth, i.e., diabetic foot or venous leg ulcer, the model correctly detects and classifies 83.7% of the present wounds - on average and across both types. Likewise, when a model classifies a detected wound, the fraction of correct detections is 94.2% (precision or positive predictive value). So, when the system predicts and classifies a wound, the probability of a correct prediction is high.

A similar initiative, limited to diabetic foot ulcer detection only [7,8], obtained mAP values in the range of the one in this study. However, they reported lower precision and recall values.

Our results demand cautious interpretation as the images originate from two sites only and, thus, may contain some non-generalizable information. We tried to mitigate this hazard using image augmentation to inflate the dataset and increase its generalizability artificially. Furthermore, our images showed wounds that demand specialized wound care, as both sites focus on the enhanced treatment of complex wound care situations. This may be a limiting factor as less complex wounds and those just on the onset may be underrepresented in the dataset. Therefore, additional images from more sites, including the full spectrum of wound complexity, must be included and tested. Lastly, single-shot detectors may consider background and context information such as anatomical landmarks during detection and classification. For example, when a toe is in the image, odds rise that a diabetic foot ulcer is present rather than a venous leg ulcer. We were able to remedy this effect by compiling mosaic images as training examples to display wounds of both types simultaneously to increase the context independence of the model (Fig 1).

The explainability and transparency of AI algorithms are highly relevant for their safe and reliable application in medicine. In the deep learning architecture used here, it is extremely difficult to extract the features that trigger a specific classification result. However, clinicians can rather easily verify the detection and classification results based on their own medical knowledge, which could help support trust through reproducibility.

It thus could become a potential door opener to accepting clinically more demanding tasks performed automatically. In order to strengthen trust, we are currently enhancing the system with a dashboard where clinicians upload custom images, explore inferences at different confidence thresholds and check the validity of the predictions against their expertise. Furthermore, we are planning to test the findings from the YoloV5 algorithm against other convolutional neural networks for classification and overlay heatmaps that reveal the salient areas.

In conclusion, our results on how to achieve good wound detection and classification may reveal a path towards (semi-) automated entry of wound information in patient records. This functionality may then support clinicians in record keeping and decision support.

Ethical approval

This study was approved by the Ethics Committee of the Osnabrück University of Applied Sciences, Germany (approval no. HSOS/202111/5).

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(12):e045253.
[2] O'Donnell TF, Passman MA, Marston WA, Ennis WJ, Dalsing M, Kistner RL, et al. Management of venous leg ulcers: clinical practice guidelines of the Society for Vascular Surgery and the American Venous Forum. J Vasc Surg. 2014;60(2 Suppl):3S-59S.
[3] Dissemond J, Kröger K, V ICW e.V. Chronische Wunden: Diagnostik – Therapie – Versorgung. 1st ed. München: Urban & Fischer Verlag/Elsevier GmbH; 2019.
[4] Mukherjee R, Manohar DD, Das DK, Achar A, Mitra A, Chakraborty C. Automated tissue classification framework for reproducible chronic wound assessment. Biomed Res Int. 2014;2014:851582. doi: 10.1155/2014/851582.
[5] Nguyen H, Agu E, Tulu B, Strong D, Mombini H, Pedersen P, et al. Machine learning models for synthesizing actionable care decisions on lower extremity wounds. Smart Health (Amst). 2020;18:100139. doi: 10.1016/j.smhl.2020.100139.
[6] Jocher G, Stoken A, Ayush Chaurasia, Borovec J, NanoCode012, TaoXie, et al. ultralytics/yolov5: v6.0 - YOLOv5n "Nano" models, Roboflow integration, TensorFlow export, OpenCV DNN support [Internet]. Zenodo; 2021 [cited 2022 Jan 22]. Available from: https://zenodo.org/record/5563715
[7] Cassidy B, Reeves ND, Joseph P, Gillespie D, O'Shea C, Rajbhandari S, et al. DFUC2020: Analysis Towards Diabetic Foot Ulcer Detection. Eur Endocrinol. 2021;1(1):5.
[8] Yap MH, Hachiuma R, Alavi A, et al. Deep learning in diabetic foot ulcers detection: A comprehensive evaluation. Comput Biol Med. 2021;135:104596. doi: 10.1016/j.compbiomed.2021.104596.