Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210919

Automatic Classification of Diabetic Foot Ulcer Images – A Transfer-Learning Approach to Detect Wound Maceration

Jens HÜSERS[a,1], Guido HAFER[b], Jan HEGGEMANN[b], Stefan WIEMEYER[b], Mareike PRZYSUCHA[a], Joachim DISSEMOND[c], Maurice MOELLEKEN[c], Cornelia ERFURT-BERGE[d] and Ursula HÜBNER[a]
[a] *Health Informatics Research Group, Osnabrück University of AS, Germany*
[b] *Christian Hosptial Melle, Niels Stensen Hospitals, Germany*
[c] *Department of Dermatology, Venerology and Allergology, University Hospital of Essen, Germany*
[d] *Department of Dermatology, University Hospital Erlangen, Germany*

Abstract. Diabetic foot ulcer (DFU) is a chronic wound and a common diabetic complication as 2% - 6% of diabetic patients witness the onset thereof. The DFU can lead to severe health threats such as infection and lower leg amputations, Coordination of interdisciplinary wound care requires well-written but time-consuming wound documentation. Artificial intelligence (AI) systems lend themselves to be tested to extract information from wound images, e.g. maceration, to fill the wound documentation. A convolutional neural network was therefore trained on 326 augmented DFU images to distinguish macerated from unmacerated wounds. The system was validated on 108 unaugmented images. The classification system achieved a recall of 0.69 and a precision of 0.67. The overall accuracy was 0.69. The results show that AI systems can classify DFU images for macerations and that those systems could support clinicians with data entry. However, the validation statistics should be further improved for use in real clinical settings. In summary, this paper can contribute to the development of methods to automatic wound documentation.

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

1. Introduction

Diabetes mellitus has a high prevalence and is a global health threat. Among diabetic patients, 2% - 6% of them witness the onset of a diabetic foot ulcer (DFU). The IWGDF defines a DFU as "an infection, ulceration, or destruction of tissues of the foot" of diabetic patients [1]. DFU is a severe late-stage complication of diabetes as it can lead to pain, immobility, infection, and even foot and lower leg amputations. Short-term wound characteristics may indicate delayed healing, such as peri-wound skin

[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.

maceration [2]. Thus, information about the maceration status is essential for planning the wound care and the dressing strategy for health professionals.

In this context, maceration status is part of standardized interdisciplinary wound documentation [3]. The importance of wound documentation correlates with the efforts required to enter and update information. As wound images are easy to obtain and are usually taken by the health care provider anyway, they lend themselves to be used as a data source for automatic detection. For example, AI systems were employed to detect necrotic tissues or infection status of a wound in images [4].

Against this background, particularly the need for digitally supported documentation and the relevance of wound macerations for planning and conducting wound care, this study investigates the automatic classification of DFU images. The main objective is to train an AI system and evaluate its performance.

2. Methods

For this study, we collected 416 wound images that were part of the wound documentation at the Wound Care Center of Christliches Klinikum Melle Germany, a specialized in- and outpatient clinic for patients with DFU. The data preprocessing consisted of two steps. First, the wounds in the images were annotated using bounding boxes, which are frames around the DFU in an image indicating its location. Second, we cropped all DFUs in the images using the bounding box plus 75 pixels as an additional margin. This processing led to 434 images, each showing a single DFU; eleven images contained two, and one image showed three ulcers. Then, the 434 cropped images were classified regarding the maceration status by two health professionals, a physiotherapist and a wound specialist from Christliches Klinikum Melle.

The image classification system relied on the MobileNetV1 model, a convolutional neural network (CNN) for image classification. A key feature of MobileNetV1 among compared to other CNN arcitectures is the flexible adaption of its size to control the complexity of the system which we utilized in this study. The model training used the pre-trained weights based on the *imagenet* dataset, an open image database used for AI development and benchmark, thereby model training uses a transfer-learning approach. The input images were scaled to 224 by 224 pixels (plus three color channels). The top layer of the MobileNetV1 model was replaced with two fully connected layers and a final sigmoid output layer. The final model had 847,014 parameters.

Out of these 434 images, a random subset of 326 images (75%) served as the training set. The remaining 108 images formed the validation set (25%). We used a data augmentation pipeline for model training that randomly transformed the images before each training step to avoid overfitting. The pipeline rotated, sheared, and flipped the images horizontally and vertically. Additionally, the pipeline shifted the brightness randomly. To avoid overfitting, we also defined a dropout rate of 10% in all layers and an early stopping callback to stop training when there is no improvement on validation loss after 20 epochs. The model with the lowest loss was selected as the final model. The model was evaluated on the unaugmented validation set. A GPU (Tesla P100-PCIE-16GB) served as the computational backbone for the model training which was performed using the Python version of the open-source software library *TensorFlow 2.6*.

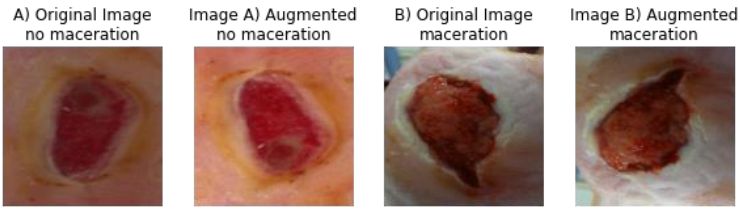

Figure 1. Subset of training images with corresponding labels and an augmented example

3. Results

The model training showed convergence, and the callback triggered early stopping after 93 training epochs. The monitored loss curves of the augmented training and validation losses showed the absence of overfitting. The final model yielded an F1-score of 0.71 on the 108 (unaugmented) validation images. This F1-score corresponds to a recall, also known as sensitivity, of 0.69 and a precision, also known as positive predictive value, of 0.67. The accuracy was 0.69, and the area under the receiver operating curve was 0.78. The images, the code of the training procedure, and the validation statistics are available online at [5].

4. Discussion

This study presents a system for classifying macerations in DFU images using a transfer learning approach. The validation showed that among all images for which the system identified a maceration, 67% were correct (precision). Among the images showing a maceration, 69% were correctly identified (recall). In light of these results, systems using artificial intelligence technologies such as CNNs promise to support the recording of DFU information. The findings are in line with similar initiatives that investigate methods to classify DFU images automatically. For example, the DFU Classification Challenge reached a F1-score of 0.73 for classifying necrotic tissues and wound infections of DFU images [4], which is comparable to our F1-score of 0.71 for macerations.

Although these validation statistics are promising for detecting macerations in DFU images from a scientific point of view, the overall accuracy of 69% is presumably not high enough for real clinical scenarios when used to automatically classify macerations. However, the current version can support semi-automatic recording by proposing the maceration status to a physician by pre-entering the information into the digital wound record, which the physician can accept or decline. Furthermore, the feedback from the physician could contribute to the continuous improvement of the CNN.

When applying classification systems like the one presented here, the context in which it was developed is essential and must be considered. For example, in this study, wound images used for model training showed DFUs without wound dressings and were not covered with cremes or gels. However, when physicians neglect this context, this might lead to unreliable classification. Thus, besides communicating the system's validity to physicians, they must be informed about the system's features and limits.

This study has limitations. Images from a single DFU center were used, and the performance of the final model was validated using the validation set rather than an additional external test set. Internal validity showed satisfactory results. We tried to improve external validity by using transfer learning, data augmentation, dropout, and a sparse model to force the system to learn the general pattern of macerations [6]. Thus, we expect this model to generalize well. Before this system is implemented in wound documentation for clinical use, it should be validated on an expanded image dataset from other clinical centers. Additionally, the number of training images should be increased to further improve the validity of the system.

In summary, the developed classification model showed satisfying validity for classifying wound images for macerations which must be further improved for clinical use in wound documentation to enable automatic wound documentation that promises to curtail the documentation time for clinicians.

Funding

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

References

[1] van Netten JJ, Bus SA, Apelqvist J, Lipsky BA, Hinchliffe RJ, Game F, et al. Definitions and criteria for diabetic foot disease. Diabetes Metab Res Rev. 2020;36(S1):e3268.

[2] Haryanto H, Arisandi D, Suriadi S, Imran I, Ogai K, Sanada H, et al. Relationship between maceration and wound healing on diabetic foot ulcers in Indonesia: a prospective study. Int Wound J. 2017 Jun;14(3):516–22.

[3] Konsensusgruppe, Herberger K, Heyer K, Protz K, Mayer A, Dissemond J, et al. Nationaler Konsensus zur Wunddokumentation beim Ulcus cruris: Teil 2: Routineversorgung – Klassifikation der Variablenausprägungen. Hautarzt. 2017 Nov;68(11):896–911.

[4] Diabetic Foot Ulcer Challenge 2020 - Grand Challenge [Internet]. grand-challenge.org. [cited 2021 Aug 21]. Available from: https://dfu2020.grand-challenge.org/Evaluation/

[5] GitHub Gist for wound image classification for macerations [Internet]. [cited 2021 Aug 27]. Available from: https://gist.github.com/jnshsrs/1d4ce0b4477330b241d61156ab973f64

[6] Shorten C, Khoshgoftaar TM. A survey on Image Data Augmentation for Deep Learning. J Big Data. 2019 Jul 6;6(1):60.