A Dataset for Diabetic Foot Assessment

João S. Sequeira[1] [0000−0002−6706−5365] and Gonçalo M. Neves[2] [0000−0003−1961−0282]

[1] Institute for Systems and Robotics/Instituto Superior Técnico, Lisbon, Portugal
joao.silva.sequeira@tecnico.ulisboa.pt
https://sites.google.com/view/joao-silva-sequeira
[2] Institute for Systems and Robotics/Instituto Superior Técnico, Lisbon, Portugal
goncalomneves@tecnico.ulisboa.pt

Abstract. The paper describes a new dataset of images to assess Diabetic Foot using automated procedures. The dataset is being constructed (ongoing work) after images obtained with purpose built hardware and in fully real healthcare settings by healthcare professionals (HP) with minimal training on the system. The system employs four cameras positioned above, below, and on both sides of the foot for full coverage. The device enables healthcare professionals to acquire images at the press of a button, after which they annotate each case as healthy, mild risk, or high risk, in a very simple protocol, and can be installed in local health units and hospitals. This process has been generating a dataset that can be used for training machine learning models to automatically assess the risk of diabetes patients developing ulcerations. The data acquisition procedure and the labelling are detailed, along with metrics that characterize the dataset.

Keywords: Diabetic Foot · Machine Learning.

1 Introduction

Diabetic Foot is a medical condition commonly resulting from Diabetes Mellitus. Recent estimates pointed to 10.5% of population has this condition, with 11.3% in 2030 and 12.2% (around 643 million) in 2045, [5]. However, the updated estimates for 2025 point to 11.1% of global population (age 29-70 years) is living with diabetes and a projected 13%, or 858 million, in 2050, [14].

Among the most serious complications are diabetic foot ulcers (DFUs). Approximately 34 % of people with diabetes will develop a foot ulcer during their lifetime [1]. Moreover, an estimated 80% of diabetes-related lower-leg amputations are preceded by DFUs [2, 8]. In addition, approximately 40% of patients developing a new DFU within one year of healing, with over 75% experiencing recurrence within five years [11], and The 5-year mortality rate for individuals who develop a DFU is approximately 30%, with rates exceeding 70% among those undergoing major amputations [2].

This shows that this medical condition, even considering a possible uncertainty in estimates, is increasingly difficult to control and hence there is a strong motivation to search for tools that can help in the first line of monitoring and prevention. Furthermore, the development of simplified/automated tests for the DF condition may improve the efficiency of human resources usage in the healthcare domain and (ii) simplify the access to an evaluation device to the general population (current medical guidelines indicate that a patient should take a DF test every 3 months and hence it is clear that the increase in the number of people with DF condition is becoming an important economic cost).

The visual analysis of the ulcerations is an area of highly active research in machine learning (ML). Detecting the presence of ulcerations is a task that can be accomplished by standard image classifiers in accordance with medical guidelines. For patients with an already installed ulceration, image analysis will be useful to estimate the area and severity of the ulceration (see for instance [26]). For patients without an explicit ulceration, the analysis of the skin condition may indicate an expected evolution towards an ulceration. The research in this area tends to be harder as skin tones and textures may mask the future development of ulcerations [21] and hence provides a motivation to the development of a dataset focused on the early stage of DF condition.

The DFUC2021 dataset (DF Ulcer Challenge[3]), for instance, provides 15,683 DFU image patches labeled for infection, ischaemia, both, or control, enabling multi-class classification research [27]. Other efforts, such as the DFUC2022 segmentation dataset, provide annotated ulcer boundaries to support wound localization studies [10].

In this work we describe a new dataset, that start to be acquired recently in the context of the DFAA[4] R&D project. Acquisition is still ongoing, though the current content is already usable in machine learning applications.

The paper is organized as follows. Section 2 summarizes ideas on visual analysis of DF. Section 3 describes the acquisition of the images. Section 4 details the annotation done on the raw images. Section 5 presents the main metrics that characterize the dataset. Section 6 presents final remarks.

2 Related work

Visual inspection of the feet by HP is prone to subjective analysis (e.g., depending on how experienced the people performing the analysis are). Current image analysis techniques, namely those resorting to machine learning (ML) techniques. [25] promise to reduce this subjective component.

Galdran et al. [6] compared convolutional neural networks (CNNs) and Vision Transformers for DFU classification using the DFUC2021 dataset, concluding that CNNs generally outperform Vision Transformers in this task, particularly with limited data, and that Sharpness-Aware Minimization (SAM) improves generalization.

[3] https://dfu-challenge.github.io/dfuc2021.html

[4] Diabetic Foot Automated Assessment, https://sites.google.com/view/dfaa-pex/home.

The ulcerations formed under DF conditions depend on multiple factors, from poor glycemic control to ill-fitting footwear, [17]. Characterizations of severity have been proposed in the literature, e.g., the Wagner scale, [18]. Moreover, the pathways for ulceration analysis, following [12], provide the goals for multiple classifiers, e.g., searching for callus and gangrene. In what concerns the classification of images to search for ulcerations, multiple pre-trained networks were already tested (ResNet50, DenseNet121, EfficientNetB2, ResNet101), with the EfficientNetB2 returning the highest Mean Average Precision (MAP), F1-score, and Area Under the ROC Curve (AUC), [15]. This shows the potential of automated visual inspection of the ulcerations, namely the detection of ulcerated areas and the measurement of their size.

A combination of CNNs and Transformers to classify four classes of severity degrees of the ulcerations is presented in [24]. The DFUC2021 dataset was used, which is oriented to the classification of infection and ischaemia of Diabetic Foot ulcers.

The DFUC2021 dataset [27], central to many classification experiments, comprises over 15,600 image patches annotated for control, infection, ischaemia, or both, and serves as a benchmark with backbones such as VGG16, ResNet, InceptionV3, DenseNet, and EfficientNet. The DFUC2022 segmentation dataset [10] includes pixel-level ulcer delineations, enabling research into wound localization, with baseline models such as DeepLabv3+ achieving Dice scores around 0.63.

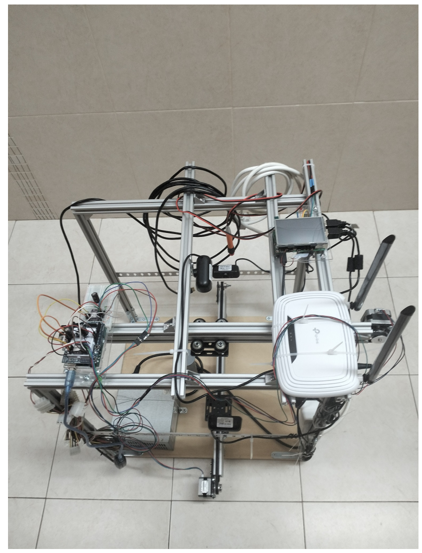
Also using EfficientNet, [23] used a Kaggle dataset. A VGG16 has been claimed to be more effective in [20]. The Kaggle dataset is also used in [19], with the prediction of ulcer pathogenesis being estimated using the classification results. The use of different networks/datasets used by different researchers induced lack of consistency on final results. Away from the deep learning context, [20] also tested several models using different combinations of random forest algorithm, logistic regression, principal component analysis, and recursive feature elimination, to classify ulcers and predict pathogenesis.

3 The DFAA dataset: acquisition of images

In DFAA, the approach taken to obtain images is somewhat different from other datasets. Instead of providing images with maximal information, i.e., in which feet are occupying a large portion of the image, which tend to require some time for proper positioning of the feet, one considers that minimal instructions are provided to the patient and, hence, it is likely that positioning is far from optimal. This intends to mimic realistic situations in healthcare scenarios, in which it will often happen that data acquisition is pressed by time (e.g., because there are multiple people waiting for evaluation). However, at this point this prevents a cross-dataset comparison as no other public access dataset for DF with this same concern is available.

The foot images where acquired using purpose built hardware, shown in Figure 1, composed by four RGB cameras (resolution 1280x720 pixel) arranged in a fixed rig structure. The multi-camera setting aims at increasing the possibility of obtaining images rich enough, i.e., if a region of the foot covered by one of the cameras is not adequately covered then the others may provide useful information. A set of high intensity white LEDs was used to provide adequate lighting conditions whenever necessary.

(a) Device 1 (jpg images)



(b) Device 2 (png images)

Fig. 1. Devices to acquire the images of the feet. The fixed rig on the righthand image is being used to test the integration of additional devices used in the Diabetic Foot assessment.

Acquisition occurred as an additional step to the regular DF checkup, and was performed by the same HP in the ULS. Patients were instructed to sit in a regular chair and put the foot heal at a specific point (indicated in the images in Figure 1). The device was operated through a single button interface. At the pressing of the button four images are acquired in sequence. The whole process for a single foot takes less than 30 seconds, insuring minimum inconvenience for the HP and patients.

The images were acquired with two different camera types (see Figure 2), introducing natural variability in the dataset.

Although the acquisition devices incorporated auxiliary high-intensity white LEDs, the tests were performed in clinical environments where ambient lighting is generally well controlled; as a result, the LEDs had a limited impact on standardizing illumination. Nevertheless, lighting conditions still influence color consistency, contrast, and the visibility of subtle skin features.

Camera characteristics also play a significant role: the cameras on the second device in Figure 2 feature a higher resolution, a wider field of view, and a shorter focal length compared to the first device. These improvements allow for better spatial coverage of the foot, enhanced sharpness at closer distances, and more consistent framing, reducing the likelihood of missing peripheral regions, even though foot positioning remains a critical factor: depending on how the patient places the foot, it may appear partially cropped or differently oriented across images. Additionally, focal distance and camera angle influence image sharpness and introduce perspective distortion, while shadows, reflections, and background heterogeneity may add further variability. Together, these factors contribute to intra-class diversity, reflecting realistic clinical conditions but posing challenges for automated classification.

Furthermore, there is a general consensus that some cleaning and preprocessing of the data must be made prior its usage (see for instance [16]). Figure 3 shows sample images with the background removed. This is a relatively straightforward process (see for instance [22]), using a U-net and a convex-hull filter, that reduces the level of disturbances in the relevant area (though the slight increase in the boundary error).

4 Labelling

The labelling of the images was made at the same time of the acquisition by HP. Consistency is thus directly related with the professional experience of the HP. Junior HP are always coached by experienced colleagues.

Besides the visual analysis, other tests for DF assessment were also registered (vibration, temperature, monofilament, and pedal pulse) were also performed (see for instance [3]). This means that the the labels embed the knowledge of all the tests and not only the visual analysis. Moreover, no information about the images (quality or any aspect of the classification) is reported back to the HP and hence quality does not bias the labels.

The system used by the professionals acquiring the dataset is based on three labels, oriented to prevention (and not ulceration classification as with the Wagner scale, i.e., oriented to the texture of the skin in areas of potential (future) ulcerations. These labels are "low risk", "medium risk", and "high risk". This represents a different strategy from, for example, the system in [24], which aimed at four labels, namely, "Infection", "Ischemia", "Both" and "None" (as in the DFUC datasets).

5 Dataset features

Currently, images in the dataset are available in two formats, namely jpeg and png. The vast majority of the images include both the left and right foot. Each set (left and/or right foot) is of a unique patient i.e., no patient has multiple images.

As in the results of Table 1, the data set, in its current state, is strongly imbalanced. Even combining the medium and high risk classes in a single class yields a skew ratio (see [9]) of $446/48 = 9.3$ (or 40.5 and 12.1 if the medium

(a) Device 1 (b) Device 1

(c) Device 1 (d) Device 1

(e) Device 2 (f) Device 2

(g) Device 2 (h) Device 2

Fig. 2. Image samples

(a) Raw image (b) Segmented image

Fig. 3. Pre-processing of the images to remove useless background features.

Table 1. Dataset features

Total Images	Images per person	Labels		
		Low risk (L)	Medium risk (M)	High risk (H)
503	8	446	11	37

and high risk classes are not merged). Such values are not common, though they have been reported in the literature (see, for instance [9]). Note, however, that skew ratio has been reported to not to be the best dataset metric. [13] suggest the Geometry Mean (GM) or the Bookmaker Informedness, or the Matthews Correlation Coefficient(MCC), as better alternatives.

In general, both the GM and the MCC of a dataset are computed using the accuracies of all classes to produce a measure relating prediction and reality (see, for instance, [4]). In our case, having no accuracies available, we can still use GM, one has $GM = (L\,M\,H)^{1/3} = 56.6$. This value is, clearly, far from the fully balanced dataset, at $(L{+}M{+}H)/3 = 163.7$, for the current size. Moreover, when applied to the raw H,M,L values, GM is affected by the size of the dataset and hence to enable size independent comparisons one can normalize the quantity as $\overline{GM} = GM/(L+M+H)$, with the current value being $\overline{GM} = 0.11$. This value is not that far from the ideal 1/3, though this is a subjective assessment.

As pointed in [7], datasets with "non-representative images", i.e., images without enough information, may compromise performance statistics "in the real world". By training with a dataset obtained in real-world conditions one can avoid this problem, i.e., performance metrics obtained with the DFAA can be expected to have a close correspondence to those obtained in real world conditions.

6 Access and final remarks

Currently available datasets are oriented to ulcerations, e.g., Kaggle, and/or impose harsh constraints to used, e.g., DFUC. The DFAA dataset described in the paper is an ongoing work, property of ULS-LOD[5], a partner of the DFAA project, and the entity responsible for Ethics clearances. Moreover, the corresponding content and statistics are being updated on a regular basis. Information to access the dataset can be requested to the corresponding author. As per the rules of the DFAA project, open access is granted for academic purposes under CC-BY or ODC-BY licencing.

A distinctive feature of this dataset is that is results from a very practical considerations, namely, (i) it is obtained without any guideline to the HP regarding quality of the images and (ii) it minimizes the overhead of the HP that supervise the process.

The imbalance is a natural consequence of the DF distribution among the population. However, given the steady rate increase in the number of images in the dataset it will be possible to select images to form a balanced sub-dataset in a near future.

Moreover, future work will include assessing the performance of baseline classifiers, e.g., using VGG and ResNet.

Acknowledgments. This work was supported by projects DFAA-2023.12465.PEX, from Fundação para a Ciência e Tecnologia, Portugal.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Armstrong, D., Boulton, A., Bus, S.: Diabetic foot ulcers and their recurrence. New England Journal of Medicine **376**(24), 2367–2375 (2017). https://doi.org/10.1056/NEJMra1615439
2. Armstrong, D., Tan, T., Boulton, A., Bus, S.: Diabetic Foot Ulcers: A Review. JAMA **330**(1), 62–75 (2023). https://doi.org/10.1001/jama.2023.10578
3. Boulton, A., et al.: Comprehensive Foot Examination and Risk Assessment: A report of the Task Force of the Foot Care Interest Group of the American Diabetes Association, with endorsement by the American Association of Clinical Endocrinologists. Diabetes Care **31**(8), 1679–1685 (2008). https://doi.org/10.2337/dc08-9021, PubMed: 18663232
4. Cleofas-Sánchez, L., Sánchez, J., García, V., Valdovinos, R.: Associative learning on imbalanced environments: An empirical study. In: Expert Systems With Applications, vol. 54, pp. 387–397. Elsevier (July 2016). https://doi.org/10.1016/j.eswa.2015.10.001
5. Dysted, M., Esztergályos, B., Gautam, S., Helman, B., Pinkepank, M., Randi, A., Salim, A., Wallis, K., Jiménez, B., Ysebaert, M.: IDF Diabetes Atlas 2021. Tech. rep., International Diabetes Federation (2021), 10th edition
6. Galdran, A., Carneiro, G., Ballester, M.: Convolutional Nets Versus Vision Transformers for Diabetic Foot Ulcer Classification. In: Yap, M., Cassidy, B., Kendrick, C. (eds.) Diabetic Foot Ulcers Grand Challenge. DFUC

[5] Unidade Local de Saúde Loures-Odivelas, Lisbon, Portugal, a unit of the Portuguese healthcare system.

2021. Lecture Notes in Computer Science, vol 13183, Springer, Cham. (2022). https://doi.org/10.1007/978-3-030-94907-5_2

7. Goldman, J., Tsotsos, J.: Statistical Challenges with Dataset Construction: Why You Will Never Have Enough Images (August 2024). https://doi.org/10.48550/arXiv.2408.11160
8. Hingorani, A., LaMuraglia, G., Henke, P., Meissner, H., Loretz, L., Zinszer, K., Driver, V., Frykberg, R., Carman, T., Marston, W., Mills, J.: The management of diabetic foot: A clinical practice guideline by the Society for Vascular Surgery in collaboration with the American Podiatric Medical Association and the Society for Vascular Medicine. Journal of Vascular Surgery **63**(2 Suppl), 3S–21S (2016). https://doi.org/10.1016/j.jvs.2015.10.003
9. Jeni, L., Cohn, J., De la Torre, F.: Facing Imbalanced Data - Recommendations for the Use of Performance Metrics. In: Procs. International Conference on Affective Computing and Intelligent Interaction (ACII) (2013). https://doi.org/10.1109/ACII.2013.47, Geneva, Switzerland, September
10. Kendrick, C., et al: Diabetic Foot Ulcer Grand Challenge 2022 Summary. In: Yap, M., Kendrick, C., Cassidy, B. (eds.) Diabetic Foot Ulcers Grand Challenge. DFUC 2022. Lecture Notes in Computer Science, vol 13797, Springer, Cham. (2023). https://doi.org/10.1007/978-3-031-26354-5_10
11. Kitano, T., Yamada, K., Matsumoto, M., Yajima, T., Hoshina, K.: Long-term outcomes after healing of diabetic foot ulcers: a retrospective cohort study. International Journal of Lower Extremity Wounds (2025). https://doi.org/10.1177/15347346251315220
12. Lepantalo, M., et al: Chapter V: Diabetic Foot. European Journal of Vascular and Endovascular Surgery **42**(S2), S60–S74 (2011)
13. Luque, A., Carrasco, A., Martín, A., de las Heras, A.: The impact of class imbalance in classification performance metrics based on the binary confusion matrix. Pattern Recognition **91**, 216–231 (2019)
14. Magliano, D., Boyco, E.: IDF Diabetes Atlas 2025. Tech. rep., International Diabetes Federation (2025), 11th edition
15. Neves, G., Sequeira, J.: Automatic Diabetic Foot Monitoring (2022), Preprint. Available at https://drive.google.com/file/d/1OD4eue2Oz8ZThRilPmphmVDgTrvrqJAA/view
16. Njeri, N.: Data Preparation for Machine Learning Modelling. International Journal of Computer Applications Technology and Research **11**(6), 231–235 (2022). https://doi.org/10.7753/IJCATR1106.1008
17. Oliver, T., Mutluoglu, M.: Diabetic Foot Ulcer. StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing (2023), Available from: https://www.ncbi.nlm.nih.gov/books/NBK537328/
18. Oyibo, S., Jude, E., Tarawneh, I., Nguyen, H., Harkless, L., Boulton, A.: A Comparison of Two Diabetic Foot Ulcer Classification Systems: The Wagner and the University of Texas Wound Classification Systems. Diabetes Care **24**(1), 84–88 (2001). https://doi.org/10.2337/diacare.24.1.84
19. Preiya, V., Ambeth Kumar, V.: Deep Learning-Based Classification and Feature Extraction for Predicting Pathogenesis of Foot Ulcers in Patients with Diabetes. Diagnostics **13**(2) (2023). https://doi.org/10.3390/diagnostics13121983
20. Reddy, S., Alluri, L., Gadiraju, M., Devareddi, R.: Forecasting Diabetic Foot Ulcers Using Deep Learning Models. In: Shakya, S., Balas, V., Haoxiang, W. (eds.) Procs. 3rd International Conference on Sustainable Expert Systems (ICSES 2022), Lecture Notes in Networks and Systems (LNNS), vol. 587. Springer (2023)

21. Reis, S., Pinto-Coelho, L., Sousa, M., Neto, M., Silva, M., Sequeira, M.: Evaluating Skin Tone Fairness in Convolutional Neural Networks for the Classification of Diabetic Foot Ulcers. Applied Sciences **15**(15) (2025). https://doi.org/10.3390/app15158321, article id 8321
22. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, arXiv:1505.04597 [cs.CV]
23. Thotad, P., Bharamagoudar, G., Anam, B.: Diabetic foot ulcer detection using deep learning approaches. Sensors International **4** (2023). https://doi.org/10.1016/j.sintl.2022.100210
24. Toofanee, M., Dowlut, S., Hamroun, M., Tamine, K., Petit, V., Duong, A., Sauveron, D.: DFU-SIAM a Novel Diabetic Foot Ulcer Classification With Deep Learning. IEEE Access (September 2023). https://doi.org/10.1109/ACCESS.2023.3312531
25. Wang, M., Timbol-Cuison, L.: Machine Learning Classification Techniques for Diabetic Foot Ulcers. In: Procs. 4th International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM'22) (2022), Hamburg, Germany, 7-9 October
26. Weatherall, T., Avsar, P., Nugent, L., Moore, Z., McDermottj, J.H., Sreenan, S., Wilson, H., McEvoy, N., Derwin, R., Chadwick, P., Patton, D.: The impact of machine learning on the prediction of diabetic foot ulcers – A systematic review. Journal of Tissue Viability **33**, 853–863 (2024)
27. Yap, M., Kendrick, C., Reeves, N., Goyal, M., Pappachan, J., Cassidy, B.: Development of Diabetic Foot Ulcer Datasets: An Overview. In: Yap, M., Cassidy, B., Kendrick, C. (eds.) Diabetic Foot Ulcers Grand Challenge. DFUC 2021. Lecture Notes in Computer Science(), vol 13183, Springer, Cham. (2022). https://doi.org/10.1007/978-3-030-94907-5_1