Appendix S1

SUPPLEMENTARY MATERIALS AND METHODS

Case selection

This retrospective study included adults diagnosed with an arteriolosclerotic ulcer of Martorell (ASUM) between 1999 and 2019 at the University Hospital of Vienna/Austria, the University Hospital of Zurich/Switzerland or the University Hospital of Bern/Switzerland. The diagnosis "ASUM" was originally based on the typical clinical as well as histological presentation of the respective patient. Typical clinical presentation included a painful ulcer with necrotic changes located on the lower leg in patients with long-standing arterial hypertension. Typical histological presentation included: (*i*) arteriolosclerosis (ideally combined with arteriolar calcification) and (*ii*) lack of signs typical for other types of leg ulcers, e.g. leukocytoclasia next to the postcapillary venoles. However, in order to exclude ulcers with components of potential other aetiological factors the following exclusion criteria were applied: (*i*) lack of valid medical documentation, (*ii*) lack of histological samples, (*iii*) no long-standing arterial hypertension documented in medical history (or lack of antihypertensive medication), (*iv*) peripheral artery occlusive disease (PAOD) > II (or PAOD with insufficient rest perfusion), (*v*) venous insufficiency with vessels draining the area of the ulcer (demonstrated by Duplex ultrasound) or any chronic venous disease at the affected patient of >C4 (according to the CEAP classification), (*vi*) clinical data or histological findings suggestive of another cause of the ulcer (e.g. pyoderma gangraenosum) or suggestive of a multifactorial cause of the ulcer, or (*vii* moderate-to-severe renal insufficiency (defined by an estimates Glomerular Filtration Rate (according to the Chronic Kidney Disease Epidemiology Collaboration formula) of <45 ml/min). The ASUM study group was compared with patients in the following leg ulcer control groups: venous leg ulcers (Venous), pyoderma gangraenosum (PG) and necrotizing small vessel vasculitis (Vasculitis). The internal Research Documentation & Analysis system was searched for patients with ulceration on the lower legs that had been diagnosed with 1 of these diseases at the Department for Dermatology of the Medical University of Vienna. Patients were included if both histological and clinical diagnosis matched the disease. Patients were excluded if there was no histological sample available or a lack of clinical documentation. Patients were also excluded if histological analysis or clinical documentation were suggestive of another cause of the ulceration or a multi-factorial event.

Patients with venous leg ulcers had to fulfil the following criteria: (*i*) history of chronic venous disease, (*ii*) venous insufficiency on lower leg, diagnosed by ultrasound, (*iii*) CEAP stage C6 on the affected leg (S1).

Patients were included in the group of necrotizing small vessel vasculitis (Vasculitis) of the following criteria were applicable: (*i*) palpable purpura on the lower legs, (*ii*) fibrinoid necrosis of vessel walls in the histology, (*iii*) neutrophilic leukocytoclasis (S2).

The inclusion of patients with pyoderma gangraenosum was based on Delphi Consensus criteria (S3) with a biopsy of ulcer edge demonstrating neutrophilic infiltrate representing the major criterion. In addition, patients had to fulfil at least 4 of the 8 minor criteria, which include: (*i*) exclusion of an infection, (*ii*) positive pathergy, (*iii*) diagnosis of an inflammatory bowel disease or inflammatory arthritis, (*iv*) history of a papule, pustule, or vesicle ulcerating within 4 days of its appearance, (*v*) peripheral erythema, undermining border, and tenderness at the site of ulceration, (*vi*) the presence of multiple ulcerations, of which at least one is located on an anterior lower leg, (*vii*) cribriform or "wrinkled paper" scar(s) at healed ulcer sites and (*viii*) a decrease in ulceration within one month of initiating immunosuppressive medication (S3).

In addition, non-ulcerative skin of non-hypertensive controls (Control N) and non-ulcerative skin of hypertensive controls (Control H) served as control groups. A similar age range compared with the ASUM study group served as an inclusion criterion for the control groups, in order to avoid major age-related biases of the study. The Hospital Information Management System of the General Hospital of Vienna was searched for patients who had undergone a skin biopsy or excision on the lower legs without pathological results (e.g. re-excision of melanomas, fibromas). Fifteen patients with and 15 without arterial hypertension were included in the analysis.

Histological analysis

All histological specimens were first stained with haematoxylin-eosin (H&E). These specimens served as a basis for the basic histomorphological analysis of the arterioles at the dermo-hypodermal junction of each sample. All arterial vessels with a size between 100 µm and 500 µm outer diameter and located at the lower dermis or upper hypodermis were included into the metric analyses. Based on previous publications on histological criteria in ASUM (for a detailed summary of published criteria please see Table SI[1]) the following criteria were selected for in depth assessment: (*i*) quantification of the wall/lumen ratio (as an indicator of the stenosing arteriolopathy) according to a method adapted from a previously published protocol by Faber & Hines (23), (*ii*) quantification of subendothelial (fibro-)hyalinosis, including the calculation of a cellularity index determined by optical quantification of all nuclei per 100 µm^2 vessel wall surface, (*iii*) quantification of the arteriolar calcification, and (*iv*) scoring of periarteriolitis.

Wall-to-lumen ratio

The wall-to-lumen ratio was calculated of all included arteriolar vessels. Vessels were excluded from the analysis if the lumen was completely occluded and no measurement could be made. A formula previously published by Faber & Hines (23), in which the mean of 4 measurements of the wall is divided by the mean of 2 measurements of the lumen, was used (see Fig. 1c). The mean of all measurements per sample was calculated.

Hyalinosis

Subendothelial hyalinosis was defined as a glassy substance in the arteriolar wall between endothelium and adventitial region that is hypo- or acellular in light microscopy. The presence of hyalinosis was evaluated by quantifying the number of patients in whose samples at least one hyalinized arteriolar vessel could be detected, and the mean number of hyalinized arteriolar vessels per sample. In addition, the cellularity was computed from size-normalized images. The nuclei found within the media of the arteriolar wall (defined as the area between endothelium and adventitial region) were counted, and the area of the media was computed using the software Fiji (Fiji is just ImageJ) (S4). The number of nuclei was divided by the area of the media in µm^2. The cellularity is given as nuclei per 100 µm^2.

Calcification

The extent of calcification was evaluated by quantifying the proportion of patients with at least one calcified arteriolar vessel and the mean number of calcified vessels per sample. In addition, the proportion of the calcified area within calcified vessels was computed by determining the ratio of area of calcification and the area of the vessel, which were measured using Fiji (S4).

Periarteriolitis

The extent of periarteriolitis was analysed by quantifying the number of vessels showing periarteriolitis per sample. To evaluate the severity of periarteriolitis a grading system from 0 (no periarteriolitis) to 4 (severe periarteriolitis) was defined. The mean grading per sample was computed by the arithmetic mean of all vessel gradings.

If there was more than one sample per patient, the mean of the 2 or more samples was calculated and used for further analysis.

Convolutional neural network analysis

To evaluate objective histomorphological distinguishability between ASUM and other types of diseases, an analysis was performed using convolutional neural networks (CNN). The learned task of the CNN was to predict the diagnosis of a case based on a series of manually acquired histological image crops showing vessels and their immediate surroundings at the dermal-subcutis junction. We retrained a ResNet34 (S5) architecture, with weights initialized via ImageNet pretraining, to predict probability for "ASUM" from a single squared RGB-image as input, resized to 224×224 pixels. Training was performed for this task within the pytorch v1.3.1 (S6) framework, using the torchvision v.0.4.2 library, with 5-fold cross-validation (CV). Adam (S7) was used as an optimizer, with a weighted cross-entropy loss, and learning rate initialized at 0.00001, which was reduced at plateaus, and a maximum of 30 epochs. Usual image augmentation techniques, such as random rotation, flipping, scaling and color jitter, were applied. Target classes were binarized to "ASUM" and "Not ASUM". No normalization for a mean pixel was performed. Twenty percent of cases were split for final test evaluations, stratified by diagnosis, whereas only cases from a 1 out of 3 centres (Vienna) were used for training. To obtain predictions for a case rather than single images, the predicted probability of "ASUM" was averaged over all images of a single case. The resulting value was measured on cases included in the validation set, and an optimal cut-off set as the one with the highest sensitivity and specificity. Given ASUM is a rather rare type of leg ulcer, more emphasis was put on sensitivity and it was provided twice the weight on this measurement, resulting in a cut-off of 0.25. The test-set included 523 images from 3 centres (Zurich 265, Bern 178, Vienna 80), including 51 cases of ASUM and 14 others, with a mean number of 8 images (SD 6.5) per case. For predictions on this set, a fixed set of 5 crops (all corners and central) including all combinations of horizontal flipping and 90° rotations were applied as Test-Time-Augmentation. To provide objective insight into what might be useful distinctive visual information, GradCAM (26) was applied to the best performing CNN, which provided a class activation map for the diagnosis of ASUM for 7×7 tiles for every input image. All resulting tiles were extracted alongside their activation value and sorted.

Discriminatory power and biostatistical analysis

The overall aim of the present study was to investigate whether certain histomorphological criteria can differentiate ASUM from other types of leg ulcers. In addition, the specificity of potential findings should be examined via comparison with non-ulcerative (healthy or hypertensive) controls. For statistical comparison of ASUM to ulcerative and non-ulcerative control groups a Fisher's exact test (for binary variables) and a Wilcoxon–Mann–Whitney test (for continuous variables) was performed. p-values <0.05 were considered statically significant, and p-values <0.01 highly significant. The significance levels were adjusted in pairwise comparisons for each variable, respectively, using a Bonferroni *post-hoc* correction. Results of continuous variables were depicted as mean ± standard deviation, if not indicated otherwise.

In addition, a logistic regression model was defined, using the diagnosis of ASUM as the target variable and the wall/lumen ratio, calcification, hyalinosis and cellularity as explanatory variables. The same model was computed using only calcification and cellularity. ROC curves were drawn in order to evaluate both models (Fig.3b). The logistic regression models distinguished between ASUM and control groups. For those variables significant in the universal model, Fisher's exact-tests (for binary variables) and Wilcoxon–Mann–Whitney test (for continuous variables) were performed to compare ASUM in terms of said variables with every other diagnosis separately. The significance levels were adjusted in the pairwise comparisons for each variable, respectively. Finally, an optimal cut-off value for cellularity to diagnose ASUM was obtained by determining the optimal probability threshold (Youden index) in the logistic model that comprises only calcification and cellularity as explanatory variables. The cut-off value of the probability was then inserted to the formula of the logistic model. By setting the variable calcification to 0 and 1, respectively, 2 optimal cut-off values for cellularity were obtained by rearranging the formula. In addition, a second cellularity and calcification threshold value were determined using an empirical method aiming at highest sensitivity and specificity. All statistical analyses were performed using IBM SPSS Statistics 24 (IBM Inc. Armonk, NY, USA) and R, Version 3.6.1.

SUPPLEMENTARY REFERENCES

S1. Lurie F, Passman M, Meisner M, Dalsing M, Masuda E, Welch H, et al. The 2020 update of the CEAP classification system and reporting standards. J Vasc Surg Venous Lymphat Disord 2020; 8: 342–352.

S2. Biswas A. Vasculitis. Cambridge: Cambridge University Press; 2016.

S3. Maverakis E, Ma C, Shinkai K, Fiorentino D, Callen JP, Wollina U, et al. Diagnostic criteria of ulcerative pyoderma gangrenosum: a Delphi consensus of international experts. JAMA Dermatol 2018; 154: 461–466.

S4. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nature Methods 2012; 9: 676–682.

S5. He K Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Available from: https://ieeexplore.ieee.org/document/7780459

S6. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. NeurIPS; 2019; 2019.

S7. Diederik P. Kingma JB. Adam: a method for stochastic optimization. Available from https://arxiv.org/abs/1412.6980. 2014: p. 1412.6980.