



OPEN

A multi-task convolutional neural network for classification and segmentation of chronic venous disorders

Bruno Oliveira^{1,2,3,4,5}✉, Helena R. Torres^{1,2,3,4,5}, Pedro Morais^{4,5}, Fernando Veloso^{1,2,4,5,6}, António L. Baptista⁴, Jaime C. Fonseca^{3,5} & João L. Vilaça^{4,5}

Chronic Venous Disorders (CVD) of the lower limbs are one of the most prevalent medical conditions, affecting 35% of adults in Europe and North America. Due to the exponential growth of the aging population and the worsening of CVD with age, it is expected that the healthcare costs and the resources needed for the treatment of CVD will increase in the coming years. The early diagnosis of CVD is fundamental in treatment planning, while the monitoring of its treatment is fundamental to assess a patient's condition and quantify the evolution of CVD. However, correct diagnosis relies on a qualitative approach through visual recognition of the various venous disorders, being time-consuming and highly dependent on the physician's expertise. In this paper, we propose a novel automatic strategy for the joint segmentation and classification of CVDs. The strategy relies on a multi-task deep learning network, denominated VENet, that simultaneously solves segmentation and classification tasks, exploiting the information of both tasks to increase learning efficiency, ultimately improving their performance. The proposed method was compared against state-of-the-art strategies in a dataset of 1376 CVD images. Experiments showed that the VENet achieved a classification performance of 96.4%, 96.4%, and 97.2% for accuracy, precision, and recall, respectively, and a segmentation performance of 75.4%, 76.7.0%, 76.7% for the Dice coefficient, precision, and recall, respectively. The joint formulation increased the robustness of both tasks when compared to the conventional classification or segmentation strategies, proving its added value, mainly for the segmentation of small lesions.

Chronic Venous Disorders (CVD) of the lower limbs are one of the most prevalent medical conditions in the adult population worldwide, representing 1–2% of the healthcare budgets in Western European countries and North America^{1–3}. In the initial stages, symptoms associated with CVD include leg pain, discomfort, heaviness, and swelling. Due to the visual impact of this condition on the patient's skin, it also results in low self-esteem, isolation, and depression, all of which impact the patient's quality of life. As the disease worsens to varicose veins, edema, skin changes, and ulceration, the quality of life is reduced and the demand for treatment increases^{2,4,5}. Due to the variety of signs and symptoms associated with CVD severity, correct diagnosis is essential to provide accurate treatment to the patients and to help in the management of medical resources^{6–9}.

The signs of CVD are typically evaluated in terms of a structured clinical classification protocol named CEAP (Clinical, Etiologic, Anatomic, Pathophysiologic)^{6,7}. This protocol incorporates a wide range of signs and symptoms of CVDs to describe their severity, ranging from C0 (no visible signs of venous disease) to C6 (active venous ulcer) (Fig. 1). While the CEAP scoring protocol is useful to classify the stages of CVD, it is relatively static, and thus, insufficient for a physician to determine quantitative changes in severity over time in response to the therapy^{10,11}. Thus, the Venous Clinical Severity Scoring (VCSS) was proposed to supplement the CEAP protocol. The VCSS system evaluates 10 clinical characteristics scored from 0 to 3 (absent, mild, moderate, severe)

¹Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal. ²ICVS/3B's - PT Government Associate Laboratory, Braga/Guimarães, Portugal. ³Algoritmi Center, School of Engineering, University of Minho, Guimarães, Portugal. ⁴2Ai - School of Technology, IPCA, Barcelos, Portugal. ⁵LASI—Associate Laboratory of Intelligent Systems, 4800-058 Guimarães, Portugal. ⁶Department of Mechanical Engineering, School of Engineering, University of Minho, Guimarães, Portugal. ✉email: boliveira@ipca.pt

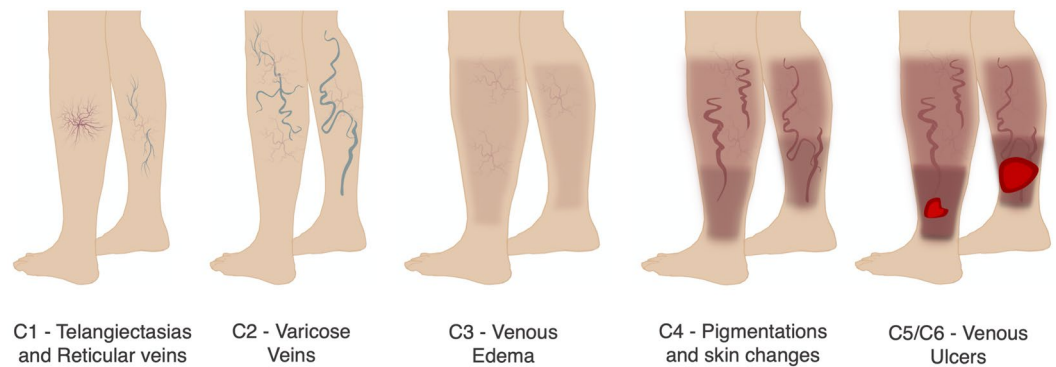


Figure 1. Chronic venous disorders (CVD) stages of development with respective severity classification (CEAP).

to produce a 30-point scale of CVD severity that enables a more sensible evaluation of the treatment response^{10,11}. Although both CEAP and VCSS protocols allow the physicians to report CVD diagnosis, infer the patient condition, and offer indicative information for the treatment, they require a physical examination and visual inspection of the patient's skin to detect the occurrence and extent of all lesions^{4,11}. Nevertheless, this process is challenging and is still totally dependent on the physician's expertise, frequently resulting in incomplete recognition of all CVD signs and, consequently, undertreated CVDs^{7,12}. Moreover, with the current scoring methods, a continuous quantification of the treatment evolution is still hampered.

The research on imaging tools for the examination of skin lesions has already demonstrated its clinical interest with increased diagnostic accuracy^{13–16}. To help the diagnosis and monitoring treatment evolution of CVD, digital photographs are regularly captured and stored by healthcare professionals^{17–19}. These photographs allow an easier evaluation of the patient's condition by comparing the classification (i.e. according to CEAP and VCSS) and lesion extension in different periods^{17,19,20}. References on the clinical images (e.g. ruler, taper) can also be used to ease the correlation with real-world units¹⁷. Nevertheless, skin images are difficult to classify and segment since the CVD lesions vary in shape and size^{13,14}. To overcome similar medical problems, automated skin lesion segmentation and classification methods have been proposed in the literature. Among them, deep convolutional neural networks (DCNN) have achieved significantly improved performance in segmentation and classification tasks on skin images when compared with other image processing strategies^{14,17}.

DCNN architectures such as VGGNet²¹, Fully Convolution Neural networks²², or U-Net²³ architectures, proved to be effective in the diagnosis of skin lesions and reached predicted levels on par with dermatologists^{14,24–27}. However, specifically for CVD images, few automatic strategies were still proposed, targeting mainly skin ulcer segmentation^{17,20,28}, or CVD severity classification^{29–31}. The interpretation of images with these skin lesions is particularly difficult. First, contrary to other medical images (e.g. CT, MRI) where standard and calibrated equipment is available, digital cameras are normally used by physicians to record photos of CVD lesions. Second, CVD presents a wide range of sizes and shapes, ranging from small vessels with 1 mm of caliber to varicose veins often protruding from the skin to skin ulcers with well-defined features. Interestingly, and although not validated in CVD, for high variable segmentation/classification situations, multi-task learning techniques have recently demonstrated their added value to empower network training^{32–36}. These methodologies combine information from different tasks, boosting the generalization ability of the network, and showing improvement in the overall method's performance. For CVD analysis, and in opposition to state-of-the-art skin lesion methods where both segmentation and classification are normally performed independently, this joint strategy could potentiate the results of each task while fulfilling the current requirements of normal clinical practice with both CVD classification and segmentation.

In this study, we propose a novel framework based on DCNN to segment CVD and classify its severity from medical images. Our proposed network, named as VENet, explores the advantages of multi-task learning by jointly learning both classification and segmentation tasks. This strategy is inspired by the U-Net architecture but redirects the high-level features from the deep stages into a new classification branch. Overall, this strategy has the potential to improve the accuracy of both tasks by sharing the same feature map and promoting each other during the training phase. VENet takes CVD images as input and outputs the CVD severity classification (according to CEAP) and segmentation results of the lesions from different severity levels, namely telangiectasias and reticular veins, varicose veins, and ulcers. For the development of the proposed framework, a new clinical dataset of CVD was constructed. The main contributions of this work are described as follows:

- A novel automatic methodology for segmenting and classifying CVD in clinical photographic images. To the best of our knowledge, this is the first attempt to perform simultaneously CVD lesions segmentation and classification;
- A new DCNN, VENet, for multi-class joint segmentation and classification of skin images in an end-to-end manner. With the joint learning of the classification and segmentation task, VENet has the potential to improve the robustness of both tasks;
- A validation of the proposed VENet pipeline against state-of-the-art single-task and multi-task segmentation and classification methods;

This paper is structured as follows. Section “**Methods**” describes the methodology of the proposed framework, with implementation details presented in Sect. “**Implementation details**”. Section “**Experiments**” introduces the validation experiments, followed by results in Sect. “**Results**”. In Sect. “**Discussion**”, the performance of each module of the proposed framework is discussed and compared with state-of-the-art results. The conclusions are given in Sect. “**Conclusion**”.

Methods

General overview. The proposed framework integrates classification and segmentation tasks in an end-to-end DCNN methodology. Initially, an RGB image is fed into the VENet, which produces two different outputs, namely a vector of probabilities and a set of segmentation masks for different CVD lesions. Specifically, the probability vector contains the severity level probability of the patient’s condition, according to 5 levels of severity: level 0—no visible signs of venous disease; level 1—telangiectasias /reticular veins; level 2—varicose veins; level 3—edema and skin changes including pigmentation and eczema; and level 4—healed/active venous ulcers (Fig. 1). Additionally, for each severity, a segmentation mask is generated with the pixel value corresponding to the probability of being a lesion with that severity.

The general overview of the proposed architecture is presented in Fig. 2. It is divided into four conceptual blocks. In the first (Sect. “**Data augmentation**”), data augmentation techniques are explored. This data is then fed to VENet, which architecture is fully explained in the second block (Sect. “**VENet architecture**”). Next, in the third block (Sect. “**Loss function**”), the multi-task loss developed to allow VENet to jointly learn the segmentation and classification tasks are described. Lastly, the post-processing strategy is presented in the fourth block (Sect. “**Post-processing**”).

Data augmentation. To deal with CVD variability, DCNN-based approaches require a large amount of data to correctly perform a specific task^{17,24}. However, access to a generalized medical dataset is challenging. Thus, to improve the generalization capacity of VENet and overcome overfitting problems, data augmentation was performed. In fact, data augmentation techniques have been shown to improve the efficacy of DCNN for the diagnosis of skin lesions^{14,24}. Two different types of data augmentation were implemented, namely spatial-based (i.e. random flip, rotation, scaling, grid distortion, optical distortion, and elastic transformations) and pixel-based techniques (i.e. random gaussian noise, brightness, contrast, and gamma transformations). While spatial-based augmentation aims to deal with the variability of lesion shapes as well as the usage of different digital cameras, pixel-based augmentations are used to improve the robustness of the VENet concerning the acquisition conditions (e.g. lighting variability) and lesions’ appearance.

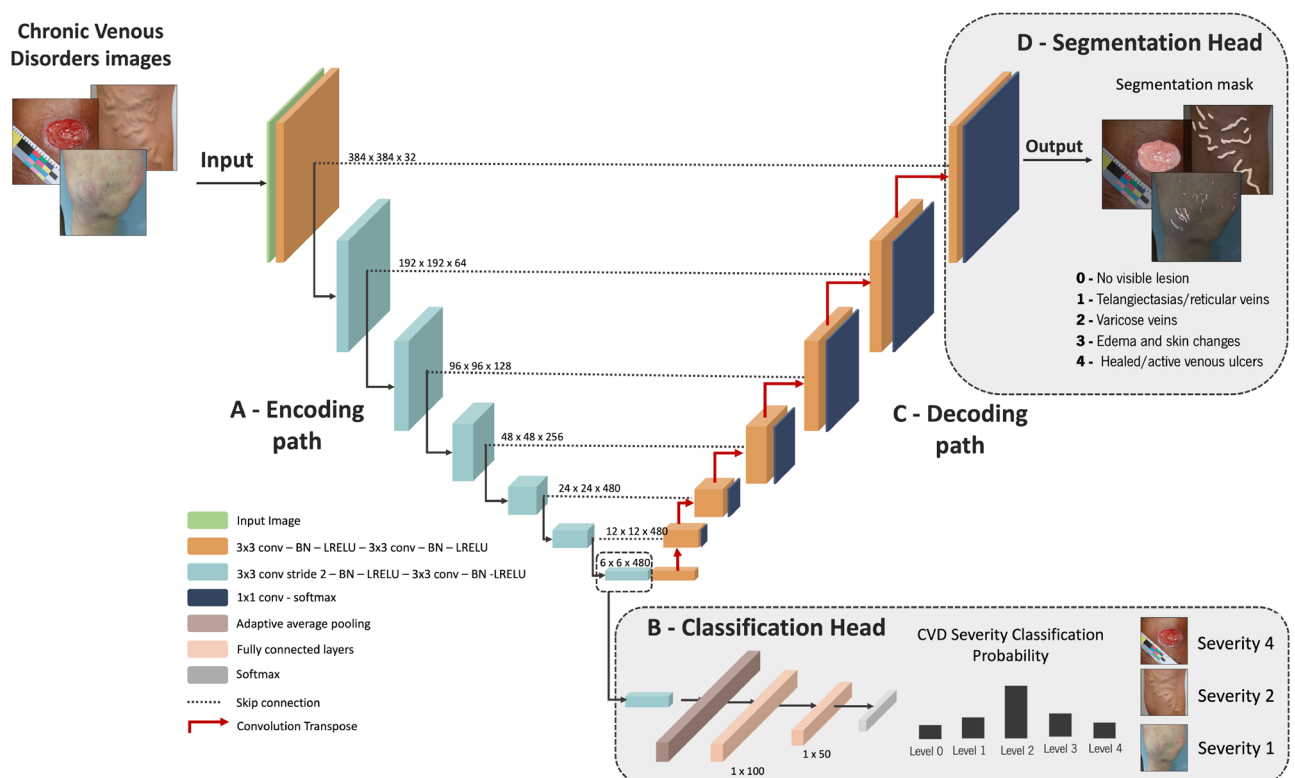


Figure 2. Overview of the proposed VENet for CVD severity classification and lesion segmentation. (A) Encoding path; (B) Classification head; (C) Decoding path; (D) Segmentation head.

VENet architecture. The augmented data is sent as the input of the VENet for CVD segmentation and classification. The VENet architecture utilizes the U-Net as the backbone, and it is divided into four parts, namely the encoding path, the classification head, the decoding path, and the segmentation head (Fig. 2).

Encoding path. The VENet encoding path is composed of downsampling blocks to extract high-level features. Each block of the encoding path is composed of two convolutional layers, with each layer consisting of a convolution, followed by batch normalization and a leaky rectified linear unit (Leaky ReLU). The downsampling is implemented using a strided convolution on the first layer of each encoding block. The initial number of feature maps is defined to be 32, which is double in each downsampling strided convolution operation. Moreover, to limit the computational cost of the model, the number of feature maps is limited to 480 (Fig. 2A). Note that, the generated features from the encoding path are shared by both classification and segmentation heads.

Classification head. For image classification tasks, state-of-the-art results are obtained using DCNN for high-level feature extraction followed by a classification head constituted with fully connected layers to get the final classification vector^{25,37}. Here, the introduction of subsampling layers improves the efficacy of the network since it eliminates the redundancy of features, reducing overfitting and minimizing computational cost^{22,38}. We propose to use the VENet encoding path to extract high-level features for CVD classification. Thus, a classification head is added to the bottom of the VENet. The high-level features are fed into the classification head composed of a convolution block, followed by an adaptive average pooling layer to allow VENet to deal with different input sizes. Finally, the features are fed into the final block composed of a fully connected layer, a Leaky ReLU activation function, another fully connected layer, and a softmax layer to get the diagnostic probability of each CVD severity level (Fig. 2B).

Decoding path. The VENet decoding path is constituted by upsampling blocks to restore the feature maps to the original input size, allowing the interpretation of low-level features. Each block of the decoding path is composed of two convolutional layers, a batch normalization, and a Leaky ReLU activation function. Between each decoding block, a subsampling operation is implemented as a convolution transposed. This makes the encoding and decoding path symmetric. To allow the propagation of spatial information, skip connections are established between the encoding path and the decoding path at the same level. To implement the skip connections, a concatenation of the feature maps of encoding blocks and decoding blocks is used (Fig. 2C)²³.

Segmentation head. In the segmentation head of VENet, the feature maps from the last block of the decoding path are fed into three consecutive convolution layers, with each layer consisting of a convolution, followed by batch normalization and a Leaky ReLU. A softmax layer is used to obtain the final instance-level probability maps. The final number of probability maps is defined to match the number of CVD lesions categories. Moreover, to allow deep supervision of the training process, probability maps from each decoding block are generated using a convolution layer followed by a softmax layer. The deep supervision allows the incorporation of the gradients from the backpropagation process deeper into the VENet easing the training process^{39,40} (Fig. 2D).

Loss function. We use a multi-task loss with two parts, namely the classification and segmentation losses^{41,42}. Similar to other classification frameworks^{43,44}, we use multi-class cross-entropy for the classification loss. Thus, with $c \in [1, 2, \dots, C]$ being the class index and C representing the number of CVD severity classes, the classification loss, L_{class} , is given by:

$$L_{class}(\mathbf{y}, \hat{\mathbf{y}}) = - \sum_{c=1}^C y_c \times \log(\hat{y}_c), \quad (1)$$

where y represents the one-hot ground truth label for each class c , and \hat{y}_c is the VENet softmax output of the classification head for the same class c .

For the segmentation task, we combine the cross-entropy loss with the DICE loss. Cross-entropy loss already showed to allow a smooth convergence, although being ineffective in the case of imbalanced data for segmentation (i.e. where the background is larger than the lesions). This limitation is minimized with the inclusion of DICE loss, which focuses on the shape similarity between the predicted mask and the ground truth CVD segmentation maps^{39,41}. The DICE loss (L_{DICE}) is described as:

$$L_{DICE}(\mathbf{g}, \mathbf{p}) = -\frac{1}{C} \times \sum_{c=1}^C \frac{2 \times \sum_{i=1}^I g_c^i \times p_c^i}{\sum_{i=1}^I g_c^i + \sum_{i=1}^I p_c^i + \varepsilon}, \quad (2)$$

where p_c^i is the softmax output of the VENet segmentation head, g_c^i the value of one hot encoding of the ground truth segmentation map for the pixel $i \in [1, 2, \dots, I]$, and class c , with I the number of pixels per image, and ε a constant value to overcome division by 0. Thus, the segmentation loss, L_{seg} is given by:

$$L_{seg} = L_{DICE} + L_{CE}, \quad (3)$$

with $L_{CE}(\mathbf{g}, \mathbf{p}) = - \sum_{c=1}^C g_c^i \times \log(p_c^i)$ being the CE loss for the segmentation task. Additionally, to enable the deep supervision of the training, the final segmentation loss is the weighted sum of the losses from all resolution outputs of VENet (i.e. from the decoding blocks and segmentation head), given by:

$$L_{seg_deep} = \sum_{r=1}^R w_{deep_r} \times L_{seg_r}, \quad (4)$$

where w_{deep_r} is the weight for the softmax output of the r VENet resolution, $r \in [1, 2, \dots, R]$, and R is the number of outputs of the VENet. Thus, $w_{deep} = \{w_{deep_1}, \dots, w_{deep_R}\}$ is the vector of weights for the deep segmentation supervision loss. Note that, since segmentation prediction masks of different stages have different sizes and cannot be compared directly to the ground truth segmentation map, downsampling of the ground truth segmentation for each level is performed. Finally, the multi-task loss function is defined as:

$$L_{VENet} = \lambda_{class} \times L_{class} + \lambda_{seg} \times L_{seg_deep}, \quad (5)$$

where $\lambda_{class}, \lambda_{seg}$ is used to weigh each task loss. The minimization of the proposed multi-task loss during the learning problem described in Sect. “Implementation details” allows VENet to jointly learn to classify and segment CVD.

Post-processing. From the output of the segmentation head of the VENet, the final predictions are obtained by selecting the label with maximum probability in each pixel. After, isolated segmented pixels are eliminated. In the end, a severity classification and lesion segmentation of the CVD images are obtained.

Implementation details

All the images were initially resized to have a minimal side equal to 384 pixels, followed by a center crop to have the final size of $[384 \times 384]$. Note that, the cropping operation was reviewed to guarantee that pixels corresponding to a lesion were not removed. After, normalization for the range $[0, 1]$ was performed. For the spatial-based augmentation, the input images were randomly flipped, either horizontally, vertically or both, with a probability of 0.3. A rotation by an angle selected randomly from a Gaussian distribution in the range $[-180^\circ, 180^\circ]$ with a probability of 0.2, and an image scale selected randomly from a Gaussian distribution in the range $[-0.3, 0.3]$ with a probability of 0.2, was also applied. A grid distortion with a probability of 0.5 with distortion limits in the range $[-0.3, 0.3]$, an optical distortion with a probability of 1 with distortion limits in the range $[-0.3, 0.3]$, and a shift limit in the range $[-0.05, 0.05]$ were also used. Finally, elastic transformations with an alpha of 120° , sigma of 6, and alpha affine in the range $[-6, 6]$ were also implemented. For the pixel-based augmentations, Gaussian noise was added to the input image with a variance range of $[0, 0.1]$ with a probability of 0.1. Random brightness and contrast augmentations selected randomly from factors in the range $[-0.25, 0.25]$ with a probability of 0.15 were also introduced. Finally, random gamma transformation with a probability of 0.3 selected from gamma limits in the range $[70, 150]$ was also used. The augmentations were performed using the Albumentations framework⁴⁵.

For the multi-task loss, λ_{class} and λ_{seg} were defined to be 0.5 and 1, respectively, which showed to achieve stable performance. Moreover, the weights w_{deep} are defined according to $w_{deep_r} = (1/2r) / \sum_{i=1}^R 1/2i$ thus giving less importance to segmentation predictions of lower resolution. The training of VENet was performed simultaneously for both classification and segmentation tasks in 1000 epochs with a mini-batch size of 4 and using the Adam optimizer with an initial learning rate of 0.0001 and a learning rate decay following the “poly” learning rate policy ($lr = 1 - (1 - epoch/\max_epoch)^{power}$) with power defined to 0.9⁴⁶.

Experiments

Data. A clinical database containing 1376 photographs of patients with CVD in lower limbs was constructed. The data was collected from normal clinical practice, where 522 images were obtained from two public datasets, namely 217 images from ULCER⁴⁷ and 305 from the SD-198⁴⁸. These images correspond to lesions from 5 different levels of CVD severity. All images were obtained using a digital camera. The distribution of images per severity level is described in Table 1. The included CVD images, excepting level 3 related images (i.e. lesions may cover the entire leg), contain pixel-level lesion annotations obtained by manual analysis of one experienced observer which were further validated by a licensed physician. Figure 3 illustrates examples of the different image sources and corresponding segmentations.

Evaluation metrics. The proposed strategy was evaluated through the following metrics: Accuracy (ACC), Recall (REC), Precision (PRE), and Area Under the Curve (AUC) of Receiver Operating Characteristic (ROC).

Severity level	Ours	ULCER	SD-198	Total
0 – No visible or palpable signs of venous disease	223	0	0	223
1 – Telangiectasias or Reticular veins	237	0	0	237
2 – Varicose veins	127	0	0	127
3 – Edema, Pigmentation and other changes in skin and subcutaneous tissue	241	0	305	546
4 – Venous ulcers	26	217	0	243
TOTAL	854	217	305	1376

Table 1. Distribution of the dataset images per severity level.

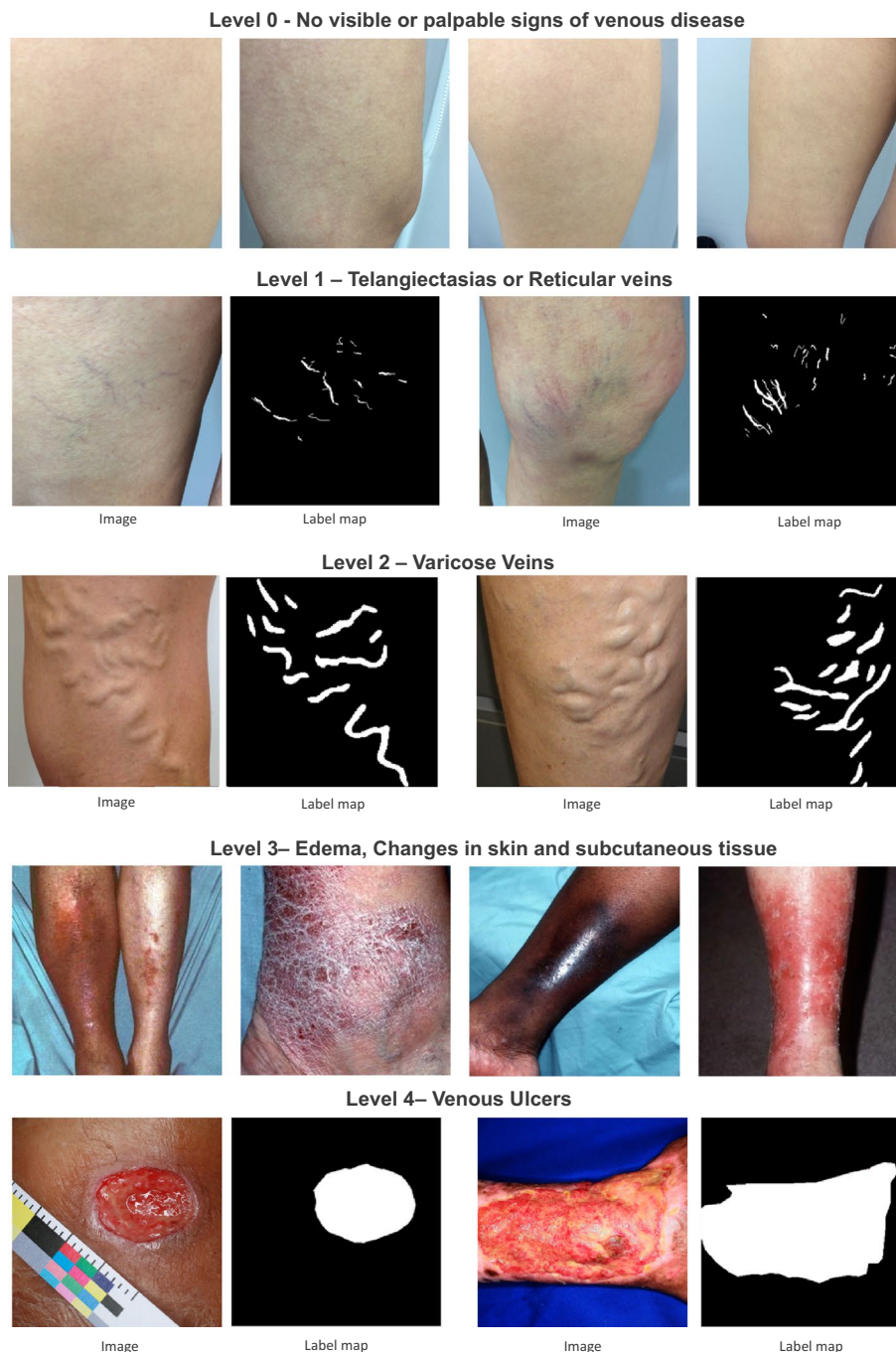


Figure 3. Examples of images used for the development of the proposed framework. Each row represents a different level of CVD severity. Note that, for severity levels with pixel annotations, two images are showed with the corresponding label map. For severity levels 0 and 3 with no manual annotations, four different images were presented.

Moreover, for the classification, the balanced F1-score (F1-score) was used to evaluate the performance of the classification task. For the segmentation task evaluation, the Jaccard (JAC) and Dice coefficient (DICE) was also used to evaluate the similarity between the predicted and ground truth label map. Due to the multi-class segmentation and classification tasks, the average of the metric score from all classes was computed.

Evaluation strategies. To evaluate the accuracy and robustness of the proposed VENet, several experiments were conducted, namely ablation studies (Sect. “[Ablation studies](#)”), a comparison against other conventional DCNN methodologies for lesion classification (Sect. “[Classification DCNN](#)”) and lesion segmentation tasks (Sect. “[Segmentation DCNN](#)”), and a comparison against other multi-task methodologies (Sect. “[Multi-task DCNN](#)”). All the networks were trained using the parameters described in Sect. “[Implementation details](#)”,

except for the learning rate, which was adapted depending on the training loss curve of each network. The results were computed using Python code running on an Intel (R) i7 CPU at 2–8 GHz with 16 GB of RAM and an NVIDIA GTX 1070 GPU with 8 GB of memory and cuDNN 10.1 libraries. The proposed VENet and all the other conventional neural network architectures were implemented using the MONAI framework⁴⁹ with the Pytorch library backend⁵⁰. To measure the statistical significance of the classification results, a one-sided McNemar's test in 2×2 tables ($p < 0.05$) was used. For the comparison of overall correctness, a joint 2×2 table was generated, which included all samples and showed the numbers of samples where none, one, or both methods produced a correct diagnosis. The statistical significance of the segmentation results of VENet was measured against all the other strategies in terms of DICE, PRE, and REC using a one-sided paired t test ($p < 0.05$). The D'Agostino-Pearson test was used to test the normality of the data.

Ablation studies. Ablation experiments were performed to analyse the effect of the optimization approach, image augmentation techniques, loss function, deep supervision training, batch size, and the number of convolution kernels of VENet. First, the dataset was randomly divided into training, validation, and testing datasets, respectively 80%, 10%, and 10% of the initial dataset. Then, for each experiment, one component of the proposed method was replaced while keeping the remaining ones. Next, the modified VENet was trained and evaluated on the training and validation datasets, respectively. Note that, for the level 3 images (i.e. without mask), the segmentation loss component is neglected.

Classification DCNN. The performance of the EfficientNet-b4⁵¹, Resnet101⁴³, DenseNet161⁴⁴, and the encoding block of VENet (i.e. henceforward called VENetC) on the classification task was evaluated against our classification method with the multi-task approach. During these experiments, the dataset from the ablation studies mentioned above was used for training, while the testing dataset was used to obtain the final results.

Segmentation DCNN. The VENet's segmentation performance was compared against the SegResNet⁵², DeepLabV3⁵³, Fully Convolution Networks (FCN)²², and the VENet segmentation block only (i.e. henceforward called VENetS). The training of all networks was performed on the training dataset, while the results presented were computed on the testing dataset. Note that, for the level 3 images (i.e. without mask), the segmentation loss component is neglected during training.

Multi-task DCNN. The proposed VENet was compared against state-of-the-art multi-task learning methods in this experiment. Specifically, four different methodologies were tested, namely MTCNN³⁵, DSI-Net³⁶, Le et al.³⁴, and Che et al.⁵⁴. All the multi-task frameworks were trained with a similar configuration to guarantee a fair comparison among methodologies. Note that, the convergence of all networks was manually verified. All results were measured on the testing database.

Results

Ablation studies. Table 2 depicts the method's performance for the different methodological steps. Overall, the proposed strategy presented the best performance for all metrics, except for the PRE. The proposed method showed results in the test dataset of 78.3% DICE, with an ACC of 81.7% and REC of 77.6% for the segmentation task. Regarding the classification, an ACC of 97.8%, PRE of 97.6%, REC of 98.5%, and F1-score of 97.8% were obtained. The ADAM optimizer showed slightly superior performance than the SGD, while better results were

Model	Segmentation				Classification			
	DICE	JAC	PRE	REC	ACC	PRE	REC	F1-score
Proposed	78.3	67.0	81.7	77.6	97.8	97.6	98.5	97.8
SGD	77.1	65.7	81.6	76.3	97.8	97.6	97.9	97.8
Initial conv = 16	74.1	62.4	82.1	72.3	97.1	96.7	97.5	97.1
Batch size = 2	67.3	54.8	81.6	63.5	89.8	89.6	87.4	89.9
Instance normalization	76.1	64.4	80.8	75.0	94.9	93.7	94.4	94.9
Loss seg. only DICE	69.8	58.0	85.0	65.2	97.1	97.7	96.6	97.1
Loss seg. only CE	65.0	53.0	89.4	56.7	94.9	95.3	95.0	94.9
Dropout 0.2	76.9	65.4	82.1	75.7	95.6	96.0	96.9	95.5
No contrast augmentation	77.4	66.1	82.3	75.9	95.6	96.3	95.9	95.6
No geometric augmentation	72.8	60.7	80.7	70.8	92.7	93.2	91.8	93.1
No loss deep supervision	75.4	63.5	80.9	73.5	94.9	95.6	94.5	95.0

Table 2. Segmentation and classification performance of the ablation studies (mean). SGD Stochastic gradient descent optimizer with Nesterov momentum ($\mu = 0.95$). Loss Seg. only DICE – without cross entropy loss on Eq. (3) Loss Seg. only CE – without dice loss on Eq. (3) For each ablation experiment, one component of the proposed method was replaced or disabled. For each row, the model column describes the component that was changed. Bold values indicate best performance.

obtained with the data augmentation strategies. Moreover, it is also shown that the introduction of deep supervision improved the VENet performance.

Classification DCNN. Table 3 shows the comparison between the performance of the proposed VENet and other conventional DCNN architectures for the classification task. Generally, VENet presents the best average results against all the other networks, with an ACC, PRE, and REC of 96.4%, 96.4%, and 97.2% respectively. When comparing against VENetC, an improvement in the overall accuracy was shown. Still, no significant differences were observed between VENet and VENetC. Figure 4 shows the confusion matrix of VENet, showing that most of the classification errors originated from images with severity level 3. Figure 5B compares the loss curves of the training of single-task and multi-task networks for the classification task. VENet presented a lower loss value after convergence on both training and testing datasets.

Segmentation DCNN. Table 4 presents the results of the performance of the proposed VENet when compared against other conventional segmentation strategies. Overall, VENet showed the best performance in comparison with all the other evaluated strategies with an average DICE of 75.4%, a PRE of 76.7%, and a REC of 76.7%. The average DICE of VENet was 2.7% higher than the VENetS architecture. The segmentation results for each lesion separately show that VENet presents the best results against all the other DCNN architectures, with an average DICE of 70.8%, 62.8%, and 92.5% for the telangiectasias and reticular veins, varicose veins, and venous ulcers, respectively. Figure 5A compares the loss curves of the training of single-task and multi-task networks for the segmentation task. Again, analysing Fig. 5B, a better generalization ability of VENet against the single-task segmentation strategies was observed. Figure 6 shows example results for all the evaluated DCNNs.

Multi-task DCNN. Table 5 compares the performance of the VENet methodology against other state-of-the-art multi-task DCNNs. Overall, the VENet architecture presents the best performance for the segmentation of the CVD lesions, with an increment of 4% against DSI-Net, which achieved the second-best performance. Nevertheless, for the classification task, both DSI-Net and Che et al. frameworks achieved better performance with an accuracy of 97.8 and 97.1 against the 96.4 achieved by VENet.

Model	ACC	PRE	REC	F1-score	AUC
VENet	96.4	96.4 ± 5	97.2 ± 2.6	96.3	99.6 ± 0.5
VENetC	94.9	93.8 ± 5.8	96 ± 3.9	94.9	99.1 ± 0.8
DenseNet161	93.5 ^a	93.3 ± 6.5 ^a	92.5 ± 5.8 ^a	93.5 ^a	96.5 ± 5.1 ^a
Resnet101	94.2	92.6 ± 8.5	95.4 ± 4.8	94.1	99.2 ± 0.7
EfficientNetB4	93.5 ^a	93.3 ± 8.3 ^a	93.5 ± 2.2 ^a	93.4 ^a	98.5 ± 1.1 ^a

Table 3. Comparison of the VENet performance against conventional DCNN for CVD classification task (mean ± S.D.). ^ap < 0.05. Two-sided McNemar's test against the proposed VENet strategy. Bold values indicate best performance.

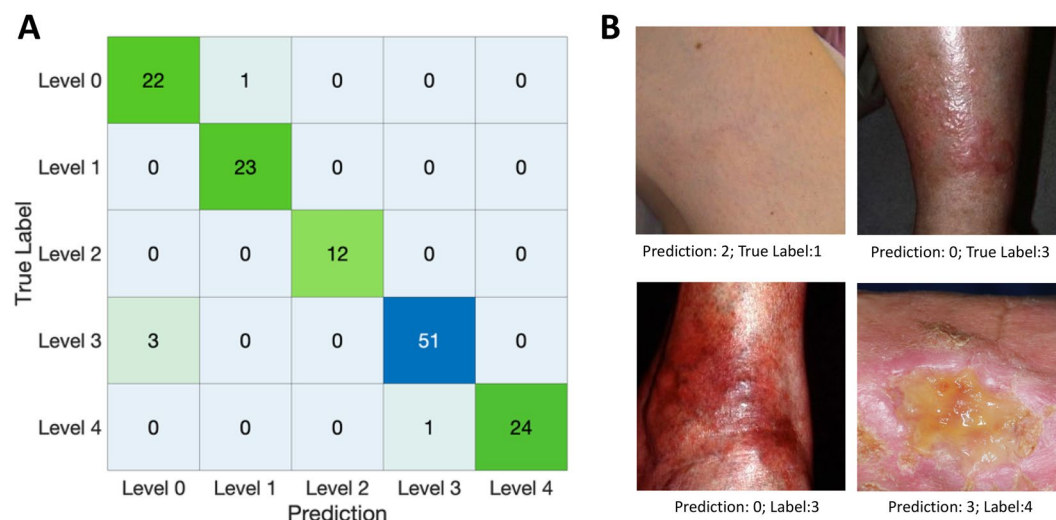


Figure 4. Confusion matrix of the proposed VENet for the classification of CVD images from the validation dataset. (A) Confusion matrix in each row represent the true label, while the column represents the VENet predicted classification. (B) Example of wrong classified images by VENet.

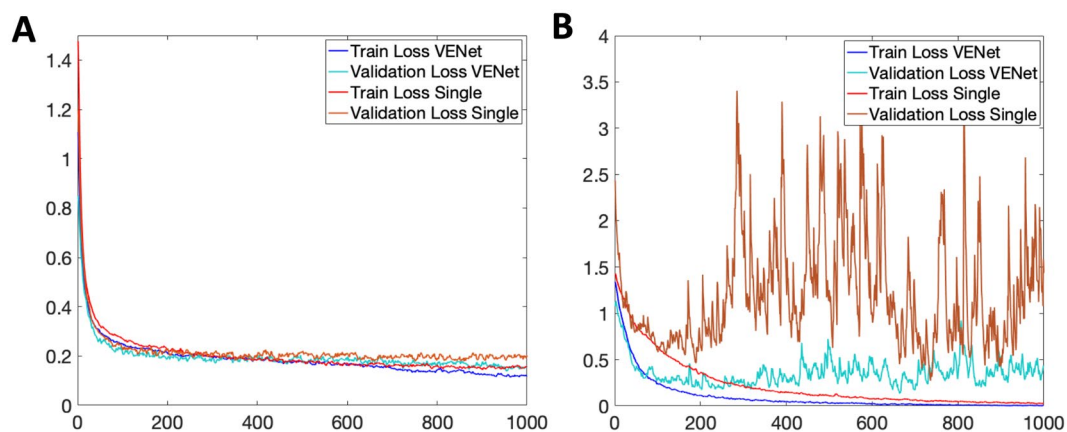


Figure 5. Comparison of loss curves of the training of single task DCNN (VENetS and VENetC for segmentation and classification, respectively) against the proposed multitask VENet. (A) Loss curves on the training/validation dataset for the segmentation task; (B) Loss curves on the training/validation dataset for the classification task. Note that, to ease interpretation, the segmentation loss curve was shifted up to have 0 as the minimum value.

Model	Telangiectasias reticular veins			Varicose veins			Venous ulcers			Overall		
	DICE	PRE	REC	DICE	PRE	REC	DICE	PRE	REC	DICE	PRE	REC
VENet	70.8 ± 10.6	70.3 ± 16.4	75.3 ± 13.6	62.8 ± 12.1	65.4 ± 8.9	63.1 ± 16.7	92.5 ± 6.3	94.4 ± 4.4	91.6 ± 9.9	75.4 ± 9.7	76.7 ± 9.9	76.7 ± 13.4
VENetS	68.1 ± 11.7 ^β	72 ± 16.4	68.7 ± 16.2 ^β	58.8 ± 16.9 ^β	67.4 ± 12.5	55.8 ± 21.9 ^β	91.3 ± 6.1	93.4 ± 6.2 ^β	90.1 ± 9.7	72.7 ± 11.6 ^β	77.6 ± 11.7	71.5 ± 15.9 ^β
SegResnet	68.3 ± 11.5 ^β	73 ± 17.5	69.8 ± 17.3 ^β	62 ± 13.3	62.4 ± 10.6	64.2 ± 19.6	91.7 ± 8.3	92.8 ± 10.6	91.8 ± 9	74 ± 11 ^β	76 ± 12.9	75.2 ± 15.3 ^β
Deeplabv3	63.3 ± 12.2 ^β	60.4 ± 15.7 ^β	69.9 ± 13.7 ^β	50.5 ± 24.1 ^β	60.9 ± 9 ^β	51.3 ± 25.7	88.6 ± 20	92.4 ± 4.1	86 ± 21.4	67.5 ± 18.8 ^β	71.2 ± 9.6 ^β	69.1 ± 20.2 ^β
FCN	61 ± 12.9 ^β	67.8 ± 15.5 ^β	59.5 ± 17 ^β	60.7 ± 10.2 ^β	64.2 ± 8.1	60.4 ± 17.5	92.3 ± 6.6	96.3 ± 3.5^β	89.4 ± 10.9 ^β	71.3 ± 9.9 ^β	76.1 ± 9	69.8 ± 15.1 ^β

Table 4. Comparison of the VENet performance against conventional DCNN for the CVD segmentation task. The results for the segmentation of each studied CVD lesion are presented separately. Moreover, the overall column presents the average of the individual lesion results (mean ± S.D.) ^βp < 0.05, Paired *t* test against the proposed VENet strategy. Bold values indicate best performance.

Discussion

The proposed study presents a fully automated DCNN approach for severity classification and lesion segmentation from CVD images. Such a tool can potentiate diagnostic accuracy and enable the quantification of the treatment evolution. To the best of our knowledge, this is the first attempt to perform simultaneously CVD lesions segmentation and classification. Overall, the presented method showed high accuracy against state-of-the-art CVD lesion classification and segmentation techniques (i.e. single-task and multi-task). Indeed, the joint learning promoted the encoding path of VENet to extract more relevant features from the input images, leading to the generalization of the method, and ultimately resulting in increased robustness in both tasks. Specifically, for the segmentation, a statistically significant improvement in the performance was obtained, mainly for the segmentation of small lesions. For the classification task, although no statistical differences were measured, more stable training and an increase in classification robustness were also registered.

To validate the described pipeline, ablation analysis was performed. Table 2 proves the importance of each described component, achieving better results for most of the studied parameters of the base configuration, except for the PRE where isolated loss segmentation functions (i.e. only cross-entropy or DICE) showed superior performance. Indeed, cross-entropy is less susceptible to lesion size variations, allowing to have a more stable loss convergence during training⁵⁵. Still, its low performance with imbalanced segmentation data also resulted in worse performance for the DICE metric. The introduction of data augmentation resulted in a significant impact on the performance of the VENet. Interestingly, geometric-based augmentation techniques proved to be more relevant for the final method's performance when compared to contrast-based augmentation ones. The original dataset images were obtained from different sources and under different light conditions. This already introduces variability, which reduces the impact of contrast-based augmentation techniques on the method's performance. On the contrary, venous disorders have high shape variability that is not represented in the original dataset. This limitation is reduced with the introduction of geometric-based data augmentation techniques. Note that, besides the deep supervision is focused on the decoding path, and consequently on the segmentation task of VENet, its absence resulted in a decrement in the accuracy of both tasks. This demonstrates that the joint learning of both tasks promoted the quality of the shared features extracted along the encoding path.

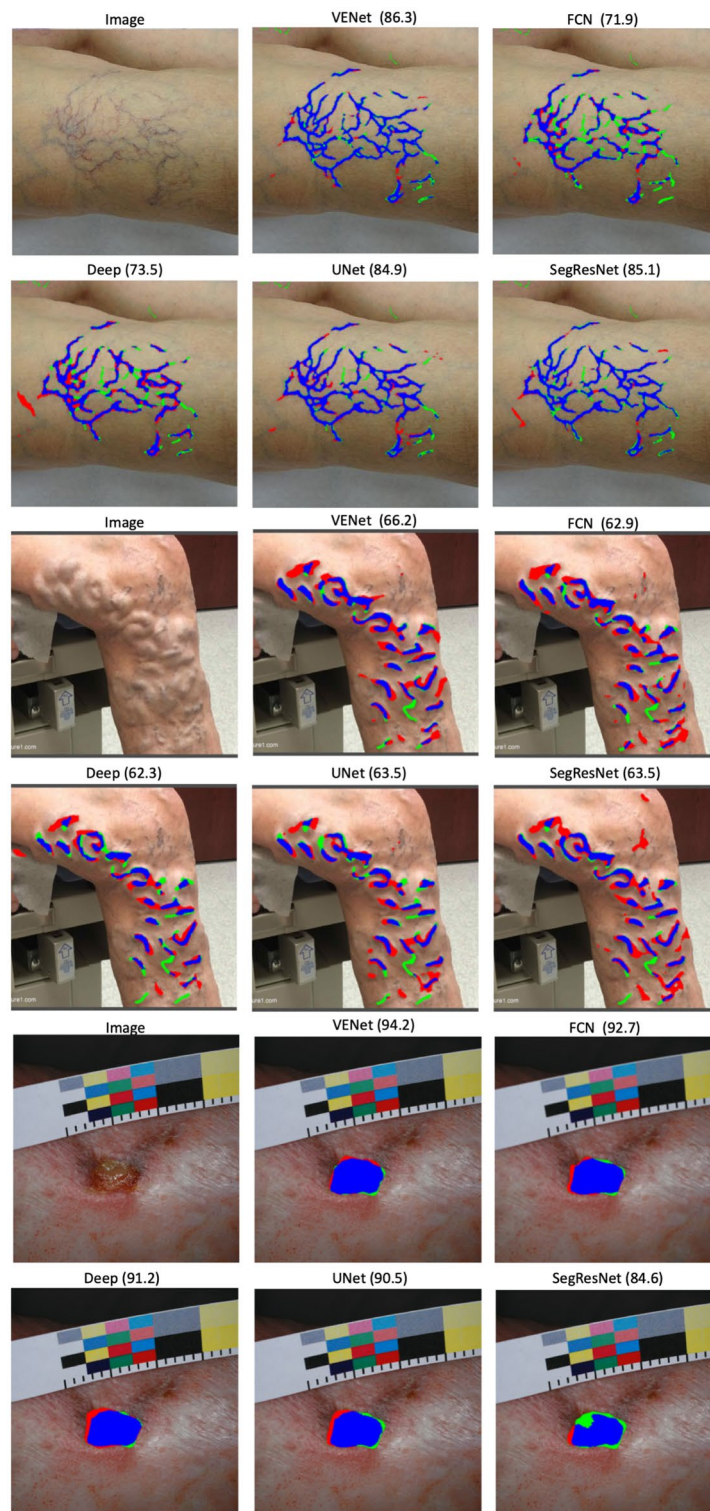


Figure 6. Segmentation examples of the proposed VENet and other conventional DCNN for the segmentation of CVD lesions, namely telangiectasias and reticular veins, varicose veins, and venous ulcers. The DICE metric is also provided for each segmentation result. The blue, red, and green colors represent the true positives, false positives, and false negatives, respectively.

Table 3 validates the performance of VENet for the classification of CVD severity. Comparing both VENet and VENetC (i.e. encoding path plus classification head), although an improvement in the classification performance

Model	Segmentation			Classification			
	DICE	PRE	REC	ACC	PRE	REC	F1-score
Proposed	75.4 ± 9.6	76.7 ± 10.1	76.7 ± 13.3	96.4	96.4 ± 5.0	97.2 ± 2.6	96.3
Le et al.	61.7 ± 11 ^β	62.4 ± 12 ^β	66.1 ± 16 ^β	96.4	95.9 ± 5.2	97.1 ± 3.6	96.3
DSI-Net	71.4 ± 11.4 ^β	77.2 ± 10.8	68.7 ± 15.7 ^β	97.8	97.5 ± 3.3	97.9 ± 3.4	97.8
MTCSN	60.8 ± 14.6 ^β	62.5 ± 14.4 ^β	64.6 ± 21.2 ^β	95.6	95.4 ± 3.5	95.9 ± 2.8	95.6
Che et al.	70.8 ± 10.1 ^β	73.4 ± 10.7 ^β	70.5 ± 13.8 ^β	97.1	96.7 ± 3	97.5 ± 3.4	97.1

Table 5. Comparison of the VENet performance against other multitask DCNN for CVD classification and segmentation (mean ± S.D.). ^βp < 0.05, Paired *t* test against the proposed VENet strategy. ^αp < 0.05, Two-sided McNemar's test against the proposed VENet strategy. Bold values indicate best performance.

was achieved, no significant differences were found. Still, a better generalization ability of the VENet during training was verified (Fig. 5B). With the multi-task approach, the different tasks can provide regularization to each during the training phase^{25,42}. Thus, the common features that lead to the optimized result of the VENet for both tasks are searched. This increases the robustness against the high variability between lesion types. Since the CVD lesions' size are usually minor in the early stages, only a small region of the image can be occupied, which may lead the classification network to neglect them. Thus, by combining both tasks, the segmentation head can focus the classification network on the lesion's spots, improving the overall robustness of the classification task during training.

Regarding the segmentation task, when looking individually for each lesion, VENet outperformed all the evaluated DCNN with an overall significantly better DICE and REC, with an improvement of approximately 2.5%, 4%, and 1.2% when comparing with VENetS for the segmentation of telangiectasias and reticular veins, varicose veins, and ulcers, respectively. Specifically, the second-best performance was obtained by SegResNet, suggesting that the incorporation of such architecture as the backbone of the multi-task framework could be an interesting solution. Thus, VENetS was ineffective in detecting equally good features among the different lesion types when compared with VENet. By assessing Fig. 6, it is possible to observe the advantages and consistency of the proposed methodology in the detection of all lesions. Altogether proved that the joint training of both tasks potentiate the generalization ability of the network, enabling the extraction of more relevant information from the CVD images. Thus, the benefit of jointly performing both classification and segmentation is demonstrated^{32,33,41,56}. Note that, this benefit is stronger for the segmentation task than for the classification task where no statistical differences were found when compared against the single-task strategy. Comparing VENet against VENetS per lesion performance, significantly better results were obtained for the segmentation of telangiectasias/reticular veins, and varicose veins. Thus, this suggests the superior performance of VENet for the segmentation of these smaller lesions.

Making a quantitative analysis of the results, an ACC of 96.4% with a PRE of 96.4% and REC of 97.2% was achieved for the classification of CVD severity, demonstrating the advantage of the described strategy for normal clinical practice. In Fig. 4, we can see that the images wrongly classified with severity 0 presented some artifacts related to the light condition which could have misled VENet. Most of the errors were from mislabelled images of severity level 3. This severity level comprises a wider range of lesions than the other ones, which seems to hamper the performance of VENet for the classification of these images. Moreover, the dataset showed an imbalanced nature between classes. Interestingly, the impact of using a class-balanced loss on VENet performance was also evaluated (for more details, the reader is kindly directed to the supplementary material, Appendix A). Although significant differences were not found, the incorporation of a class-balanced focal loss improved the classification results. Nevertheless, this also leads to a decrement in the segmentation performance, which can be the result of the optimization process starting to focus on minimizing specific classification errors that now have increased weight. Still, when the focus is on the performance of the lesion classification stage, class-balanced losses proved their added value. Note that, for the diagnosis of medical conditions, underdiagnoses have a higher cost for the patient's condition than overdiagnosis. This feature can be learned by VENet by weighting mislabelled images of higher severity levels during the training process. With such a strategy, one is expected to reduce the underdiagnoses of CVD images.

For the segmentation task an overall DICE of 75.4%, with a PRE of 76.7% and a REC of 76.7% was achieved. Specifically, a superior average DICE was achieved for venous ulcers when compared with the results from other lesions. This result is explained by the larger dimensions with well-defined boundaries of the leg ulcers, making their segmentation simpler (Fig. 6). In opposition, vascular lesions such as telangiectasias are smaller lesions and difficult to detect. Varicose veins may not present well-defined boundaries since they are normally deeper and protruding vessels in the skin (i.e. the lesion is not illuminated evenly). This makes the segmentation task more challenging for these lesions. Here, post-processing techniques to fuse components of the same vessel, as proposed for other segmentation problems⁵⁷, can be explored in the future to improve the method's overall performance. Additionally, the segmentation result can notably increase the performance of the clinical diagnosis, enabling the quantification of treating evolution by simply comparing lesion masks of the patients over time. This gives enhanced information on the treatment monitoring to the physician and allowing to overcome the current limitations of the actual scoring strategies.

When comparing VENet against other multi-task frameworks, for the classification task, two networks presented slightly better results than VENet, namely DSI-Net and Che et. al. with an ACC of 97.8 and 97.1, against 96.4 of VENet. Still, no significant differences were obtained. For the segmentation task, a statistically significantly

better performance of VENet was obtained, with an improvement of around 4% of the DICE for the second-best performance. This corroborates again the added value of VENet for the task of segmentation of CVD lesions.

Analysing the evaluated multi-task architectures, they all are similar with a backbone to extract features, followed by two heads for the segmentation and classification tasks. To take advantage of features from both tasks, and guarantee consistency in the prediction the DSI-Net, Che et al. and MTCSN networks proposed feature passing modules between the two branches. This may improve the segmentation performance for both DSI-Net and Che et al. strategies when compared with the Le et al. network. Still, VENet outperformed the results of these networks. VENet also potentiates the interaction between the two tasks with feature-passing modules (i.e. concatenation) between the encoding and decoding paths. Nevertheless, instead of the evaluated multi-task DCNNs that only add such modules after the feature extraction, VENet establishes a connection between both tasks from the beginning of the encoding path. This enhances the propagation of spatial information by minimizing the loss of information during the down-sampling process, which may occur mainly for the smaller lesions, such as telangiectasias, reticular veins, and varicose veins. Indeed, VENet presented an average DICE of 70.8 and 62.8 against the 65.5 and 57.1 of the DSI-Net, for the segmentation of telangiectasias/reticular veins, and varicose veins, respectively. For the segmentation of venous ulcers, VENet presented an average DICE of 92.5 against 91.6 from the DSI-Net. As already shown, this proves the added value of the proposed VENet strategy, mainly for the segmentation of small structures.

Overall, the automatic VENet strategy can ease and fast the diagnosis of CVD lesions, being a relevant tool for physicians as well as a self-care module for patients, thus, potentially reducing the probability of underdiagnoses and promoting the treatment of CVD in the early stages. Notwithstanding, this study presents some limitations that could be addressed in the future. During training, the number of feature maps of VENet has been limited to 480 due to the computational cost. A half-precision training strategy can be explored in the future to minimize this limitation⁵⁸. This may allow for increasing the number of feature maps as well as the batch size during training. Currently, VENet has around 34.15 million parameters, taking 25.8 ms for inference of a CVD image using the computational setup previously described in Sect. “Experiments”. Here, techniques to reduce memory and energy consumption could be explored, e.g. pruning⁵⁹. Next, the results were obtained in a dataset of 1376 images, some obtained from the same clinical center. Therefore, specific intrinsic factors of the training dataset may be common, such as the camera acquisition parameters. Although data augmentation was explored to reduce this dependency, this raises concern about the method's ability to generalize on a new test dataset with images from clinical centers not presented on the training/validation datasets.

Finally, CVD presents high levels of recurrence, which demands a periodic evaluation of the evolution of the patient's condition^{60–62}. For it, a computation of real-world units of all detected lesions is required, which can be accomplished with the introduction of a reference in the images (e.g. ruler, taper) or the usage of an RGB-Depth sensor. In fact, the impact of incorporating references, such as rulers or tapers on the images was tested in a preliminary experiment (please see the supplementary material, Appendix A). Although significant differences were not found for the DICE, the incorporation of references in the images also resulted in an increment in the performance of the VENet. Since they were mainly located on the severity level 4 images, they could be learned by DCNN to aid the classification and segmentation tasks. Note that, although the incorporation of such references has not been the focus of this work, this new development is envisioned as future work. With this extension, VENet can also enable a continuous quantification of patients' conditions over CVD treatment and long-term recurrence of lesions. Indeed, the quantification of the treatment evolution is a desired feature among the treatment of all CVD severity levels, such as quantifying the treatment performance of vascular lasers for the treatment of telangiectasias (i.e. vessel clearance over treatment sessions), the recurrence of varicose veins after treatment (e.g. stripping, laser ablation), or also assess the evolution of healing of venous ulcers¹⁷. The computation of real-world lesion size using VENet work is envisioned as future work.

Conclusion

In this work, we proposed a fully automated method for CVD severity classification and lesion segmentation. This method explores the multi-task learning advantages by learning both tasks simultaneously. The proposed VENet methodology allowed to boost the performance against other conventional DCNNs for the classification and segmentation tasks. While for the classification these resulted in increased robustness and training stability, although without statistical differences, for the segmentation task, statistically significantly better performance was obtained, mainly for the segmentation of small lesions. Overall, the proposed method can be explored in normal clinical practice to aid physicians and patients in the diagnosis/monitoring of CVD, overcoming underdiagnoses, and potentiating the treatment of these lesions in the early stages.

Data availability

The datasets generated and analysed during the current study can be made available at reasonable request to the corresponding author.

Received: 8 June 2022; Accepted: 26 December 2022

Published online: 14 January 2023

References

1. Rabe, E. et al. Epidemiology of chronic venous disorders in geographically diverse populations: Results from the vein consult program. *Int. Angiol.* **31**(2), 105–115 (2012).
2. Bergan, J. J. et al. Chronic venous disease. *N. Engl. J. Med.* **355**(5), 488–498 (2006).

3. Meulendijks, A. M., de Vries, F. M. C., van Dooren, A. A., Schuurmans, M. J. & Neumann, H. A. M. A systematic review on risk factors in developing a first-time venous leg ulcer. *J. Eur. Acad. Dermatol. Venereol.* **33**(7), 1241–1248. <https://doi.org/10.1111/jdv.15343> (2019).
4. Eberhardt, R. T. & Raffetto, J. D. Chronic venous insufficiency. *Circulation* **130**(4), 333–346. <https://doi.org/10.1161/CIRCULATION.NAHA.113.006898> (2014).
5. Meulendijks, A. M., Welbie, M., Tjin, E. P. M., Schoonhoven, L. & Neumann, H. A. M. A qualitative study on the patient's narrative in the progression of chronic venous disease into a first venous leg ulcer: A series of events. *Br. J. Dermatol.* **183**(2), 332–339. <https://doi.org/10.1111/bjd.18640> (2020).
6. Meissner, M. H. *et al.* Primary chronic venous disorders. *J. Vasc. Surg.* **46**(6), S54–S67. <https://doi.org/10.1016/j.jvs.2007.08.038> (2007).
7. Lurie, F. *et al.* The 2020 update of the CEAP classification system and reporting standards. *J. Vasc. Surg. Venous Lymphat. Disord.* **8**(3), 342–352. <https://doi.org/10.1016/j.jvsv.2019.12.075> (2020).
8. Meulendijks, A. M., Franssen, W. M. A., Schoonhoven, L. & Neumann, H. A. M. A scoping review on chronic venous disease and the development of a venous leg ulcer: The role of obesity and mobility. *J. Tissue Viabil.* **29**(3), 190–196. <https://doi.org/10.1016/j.jtv.2019.10.002> (2020).
9. Oliveira, B. *et al.* Characterization of the workspace and limits of operation of laser treatments for vascular lesions of the lower limbs. *Sensors* **22**(19), 7481. <https://doi.org/10.3390/s22197481> (2022).
10. Rutherford, R. B. *et al.* Venous severity scoring: An adjunct to venous outcome assessment. *J. Vasc. Surg.* **31**(6), 1307–1312. <https://doi.org/10.1067/mva.2000.107094> (2000).
11. Vasquez, M. A. *et al.* Revision of the venous clinical severity score: Venous outcomes consensus statement: Special communication of the American Venous Forum Ad Hoc Outcomes Working Group. *J. Vasc. Surg.* **52**(5), 1387–1396. <https://doi.org/10.1016/j.jvs.2010.06.161> (2010).
12. Davies, A. H. The seriousness of chronic venous disease: A review of real-world evidence. *Adv. Ther.* **36**(S1), 5–12. <https://doi.org/10.1007/s12325-019-0881-7> (2019).
13. Lei, B. *et al.* Skin lesion segmentation via generative adversarial networks with dual discriminators. *Med. Image Anal.* **64**, 101716. <https://doi.org/10.1016/j.media.2020.101716> (2020).
14. Yuan, Y., Chao, M. & Lo, Y.-C. Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance. *IEEE Trans. Med. Imaging* **36**(9), 1876–1886. <https://doi.org/10.1109/TMI.2017.2695227> (2017).
15. Thomsen, K., Iversen, L., Titlestad, T. L. & Winther, O. Systematic review of machine learning for diagnosis and prognosis in dermatology. *J. Dermatol. Treat.* **31**(5), 496–510. <https://doi.org/10.1080/09546634.2019.1682500> (2020).
16. Iqbal, I., Younus, M., Walayat, K., Kakar, M. U. & Ma, J. Automated multi-class classification of skin lesions through deep convolutional neural network with dermoscopic images. *Comput. Med. Imaging Gr.* **88**, 101843. <https://doi.org/10.1016/j.compmedimag.2020.101843> (2021).
17. Chino, D. Y. T. *et al.* Segmenting skin ulcers and measuring the wound area using deep convolutional networks. *Comput. Methods Progr. Biomed.* **191**, 105376. <https://doi.org/10.1016/j.cmpb.2020.105376> (2020).
18. Anisuzzaman, D. M., Patel, Y., Niezgodna, J., Gopalakrishnan, S., Yu, Z. A Mobile App for Wound Localization using Deep Learning. **8**.
19. Cazzolato, M. T. *et al.* The UTrack framework for segmenting and measuring dermatological ulcers through telemedicine. *Comput. Biol. Med.* **134**, 104489. <https://doi.org/10.1016/j.compbiomed.2021.104489> (2021).
20. Chino, D. Y. T., Scabora, L. C., Cazzolato, M. T., Jorge, A. E. S., Traina, C., Traina, A. J. M. ICARUS: Retrieving skin ulcer images through bag-of-signatures. In *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*, Karlstad 82–87 <https://doi.org/10.1109/CBMS.2018.00022> (2018).
21. Simonyan, K., Zisserman, A. Very deep convolutional networks for large-scale image recognition. Preprint at <http://arxiv.org/abs/1409.1556> (Accessed 15 March 2021) (2015).
22. Long, J., Shelhamer, E., Darrell, T. Fully convolutional networks for semantic segmentation. Preprint at <http://arxiv.org/abs/1411.4038> (Accessed 22 December 2020) (2015).
23. Ronneberger, O., Fischer, P., Brox, T. U-Net: Convolutional Networks for biomedical image segmentation. Preprint at <http://arxiv.org/abs/1505.04597> (Accessed 15 March 2021) (2015).
24. Pérez, E., Reyes, O. & Ventura, S. Convolutional neural networks for the automatic diagnosis of melanoma: An extensive experimental study. *Med. Image Anal.* **67**, 101858. <https://doi.org/10.1016/j.media.2020.101858> (2021).
25. Zhang, J., Xie, Y., Wu, Q. & Xia, Y. Medical image classification using synergic deep learning. *Med. Image Anal.* **54**, 10–19. <https://doi.org/10.1016/j.media.2019.02.010> (2019).
26. Thomsen, K., Christensen, A. L., Iversen, L., Lomholt, H. B. & Winther, O. Deep learning for diagnostic binary classification of multiple-lesion skin diseases. *Front. Med.* **7**, 574329. <https://doi.org/10.3389/fmed.2020.574329> (2020).
27. Harangi, B. Skin lesion classification with ensembles of deep convolutional neural networks. *J. Biomed. Inform.* **86**, 25–32. <https://doi.org/10.1016/j.jbi.2018.08.006> (2018).
28. Wang, C. *et al.* Fully automatic wound segmentation with deep convolutional neural networks. *Sci. Rep.* **10**(1), 21897. <https://doi.org/10.1038/s41598-020-78799-w> (2020).
29. Shi, Q. *et al.* An automatic classification method on chronic venous insufficiency images. *Sci. Rep.* **8**(1), 17952. <https://doi.org/10.1038/s41598-018-36284-5> (2018).
30. Rostami, B. *et al.* Multiclass wound image classification using an ensemble deep CNN-based classifier. *Comput. Biol. Med.* **134**, 104536. <https://doi.org/10.1016/j.compbiomed.2021.104536> (2021).
31. Oliveira, B., Torres, H. R., Morais, P., Baptista, A., Fonseca, J., Vilaca, J. L. Classification of chronic venous disorders using an ensemble optimization of convolutional neural networks. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* 516–519 <https://doi.org/10.1109/EMBC48229.2022.9871502> (2022).
32. Wu, Z. *et al.* ELNet: Automatic classification and segmentation for esophageal lesions using convolutional neural network. *Med. Image Anal.* **67**, 101838. <https://doi.org/10.1016/j.media.2020.101838> (2021).
33. Gao, K. *et al.* Dual-branch combination network (DCN): Towards accurate diagnosis and lesion segmentation of COVID-19 using CT images. *Med. Image Anal.* **67**, 101836. <https://doi.org/10.1016/j.media.2020.101836> (2021).
34. Le, T.-L.-T., Thome, N., Bernard, S., Bismuth, V., Patoureaux, F. Multitask classification and segmentation for cancer diagnosis in mammography. Preprint at <http://arxiv.org/abs/1909.05397> (Accessed 10 October 2022) (2019).
35. Kong, Z. *et al.* Multi-task classification and segmentation for explicable capsule endoscopy diagnostics. *Front. Mol. Biosci.* **8**, 614277. <https://doi.org/10.3389/fmolb.2021.614277> (2021).
36. Zhu, M., Chen, Z. & Yuan, Y. DSI-Net: Deep synergistic interaction network for joint classification and segmentation with endoscopy images. *IEEE Trans. Med. Imaging* **40**(12), 3315–3325. <https://doi.org/10.1109/TMI.2021.3083586> (2021).
37. Litjens, G. *et al.* A survey on deep learning in medical image analysis. *Med. Image Anal.* **42**, 60–88. <https://doi.org/10.1016/j.media.2017.07.005> (2017).
38. Al-masni, M. A., Al-antari, M. A., Choi, M.-T., Han, S.-M. & Kim, T.-S. Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks. *Comput. Methods Progr. Biomed.* **162**, 221–231. <https://doi.org/10.1016/j.cmpb.2018.05.027> (2018).

39. Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J. & Maier-Hein, K. H. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18**(2), 203–211. <https://doi.org/10.1038/s41592-020-01008-z> (2021).
40. Xue, C. *et al.* Global guidance network for breast lesion segmentation in ultrasound images. *Med. Image Anal.* **70**, 101989. <https://doi.org/10.1016/j.media.2021.101989> (2021).
41. Zhou, Y. *et al.* Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images. *Med. Image Anal.* **70**, 101918. <https://doi.org/10.1016/j.media.2020.101918> (2021).
42. Song, L., Lin, J., Wang, Z. J. & Wang, H. An end-to-end multi-task deep learning framework for skin lesion analysis. *IEEE J. Biomed. Health Inform.* **24**(10), 2912–2921. <https://doi.org/10.1109/JBHI.2020.2973614> (2020).
43. He, K., Zhang, X., Ren, S. and Sun, J. Deep residual learning for image recognition. Preprint at <http://arxiv.org/abs/1512.03385> (Accessed 15 March 2021) (2015).
44. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K. Q. Densely connected convolutional networks. Preprint at <http://arxiv.org/abs/1608.06993> (Accessed 15 March 2021) (2018).
45. Buslaev, A., Parinov, A., Khvedchenya, E., Igllovikov, V. I. & Kalinin, A. A. Albumentations: Fast and flexible image augmentations. *Information* **11**(2), 125. <https://doi.org/10.3390/info11020125> (2020).
46. Kingma, D. P., Ba, J. Adam: A method for stochastic optimization. Preprint at <http://arxiv.org/abs/1412.6980> (Accessed 21 March 2021) (2017).
47. Ederson, A. G. Dorileo, M. A. C. Frade, A. M. F. Roselino, R. M. Rangayyan, Azevedo-Marques, P. M. Color image processing and content-based image retrieval techniques for the analysis of dermatological lesion. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vancouver, BC, 1230–1233 <https://doi.org/10.1109/IEMBS.2008.4649385> (2008).
48. Sun, X., Yang, J., Sun, M. & Wang, K. A benchmark for automatic visual classification of clinical skin disease images. In *Computer Vision—ECCV 2016* Vol. 9910 (eds Leibe, B. *et al.*) 206–222 (Springer International Publishing, 2016).
49. The MONAI Consortium. Project MONAI 4323059. <https://docs.monai.io>, 10.5281/zenodo (2020).
50. Paszke, A. *et al.* PyTorch: An imperative style, high-performance deep learning library. Preprint at <http://arxiv.org/abs/1912.01703> (Accessed 25 March 2021) (2019).
51. Tan, M., Le, Q. V. EfficientNet: rethinking model scaling for convolutional neural networks. Preprint at <http://arxiv.org/abs/1905.11946> (Accessed 25 March 2021) (2020).
52. Myronenko, A. 3D MRI brain tumor segmentation using autoencoder regularization. Preprint at <http://arxiv.org/abs/1810.11654> (Accessed 25 March 2021) (2018).
53. Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A. L. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. Preprint at <http://arxiv.org/abs/1606.00915> (Accessed 21 March 2021) (2017).
54. Chen, S., Wang, Z., Shi, J., Liu, B., Yu, N. A multi-task framework with feature passing module for skin lesion classification and segmentation. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* 1126–1129 <https://doi.org/10.1109/ISBI.2018.8363769> (2018).
55. Asgari Taghanaki, S., Abhishek, K., Cohen, J. P., Cohen-Adad, J. & Hamarneh, G. Deep semantic segmentation of natural and medical images: A review. *Artif. Intell. Rev.* **54**(1), 137–178. <https://doi.org/10.1007/s10462-020-09854-1> (2021).
56. He, K., Gkioxari, G., Dollár, P., Girshick, R. Mask R-CNN. Preprint at <http://arxiv.org/abs/1703.06870> (Accessed 14 March 2021) (2018).
57. Mookiah, M. R. K. *et al.* A review of machine learning methods for retinal blood vessel segmentation and artery/vein classification. *Med. Image Anal.* **68**, 101905. <https://doi.org/10.1016/j.media.2020.101905> (2021).
58. P. Mickevicius *et al.* Mixed precision training. Preprint at <http://arxiv.org/abs/1710.03740> (Accessed 12 October 2022) (2018).
59. Hoefler, T., Alistarh, D., Ben-Nun, T. & Dryden, N. Sparsity in deep learning: Pruning and growth for efficient inference and training in neural networks. *J. Mach. Learn. Res.* **22**, 124 (2021).
60. Wallace, T. *et al.* Long-term outcomes of endovenous laser ablation and conventional surgery for great saphenous varicose veins. *Br. J. Surg.* **105**(13), 1759–1767. <https://doi.org/10.1002/bjs.10961> (2018).
61. Sadick, N. S. Long-term results with a multiple synchronized-pulse 1064 nm Nd:YAG laser for the treatment of leg venulectasias and reticular veins. *Dermatol. Surg.* **27**(4), 365–369. <https://doi.org/10.1046/j.1524-4725.2001.00295.x> (2001).
62. Finlayson, K., Edwards, H. & Courtney, M. Factors associated with recurrence of venous leg ulcers: A survey and retrospective chart review. *Int. J. Nurs. Stud.* **46**(8), 1071–1078. <https://doi.org/10.1016/j.ijnurstu.2008.12.012> (2009).

Acknowledgements

The authors acknowledge Fundação para a Ciência e a Tecnologia (FCT), Portugal and the European Social Found, European Union, for funding support through the “Programa Operacional Capital Humano” (POCH) in the scope of the PhD grants SFRH/BD/136721/2018 (B. Oliveira), SFRH/BD/136670/2018 (H. Torres), and SFRH/BD/131545/2017 (F. Veloso). Moreover, authors gratefully acknowledge the funding of the projects “NORTE-01-0145-FEDER-000045” and “NORTE-01-0145-FEDER-000059”, supported by Northern Portugal Regional Operational Programme (NORTE 2020), under the Portugal 2020 Partnership Agreement, through the European Regional Development Fund (FEDER). It was also funded by national funds, through the FCT and FCT/MCTES in the scope of the project UIDB/05549/2020 and UIDP/05549/2020. The authors would like to thank Ederson A. G. Dorileo, and co-authors for providing the ULCER dataset. Moreover, the authors also would like to thank Xiaoxiao Sun and co-authors for providing the SD-198 dataset. Finally, the authors would like to thank you Neusa Tenreiro and other healthcare that helped on the acquisition of the images.

Author contributions

B.O., H.T. and P.M. did the investigation, methodology development, and validation. B.O. wrote the original draft and revised the article. J.F. and J.V. supervised and secured funding. H.T., P.M., F.V., A.B., J.F. and J.V. reviewed and edited the manuscript. B.O., H.T., P.M., A.B., collected and labelled data.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-27089-8>.

Correspondence and requests for materials should be addressed to B.O.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023