2022 44th Annual International Conference of
the IEEE Engineering in Medicine & Biology Society (EMBC)
Scottish Event Campus, Glasgow, UK, July 11-15, 2022

2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 978-1-7281-2782-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/EMBC48229.2022.9871502

Classification of Chronic Venous Disorders using an Ensemble Optimization of Convolutional Neural Networks

Bruno Oliveira, Helena R. Torres, Pedro Morais, António Baptista, Jaime Fonseca, and João L. Vilaça

Abstract— **Chronic Venous Disorders (CVD) of lower limbs are one of the most prevalent medical conditions, affecting 35% of adults in Europe and North America. The early diagnosis of CVD is critical, however, the diagnosis relies on a visual recognition of the various venous disorders which is time-consuming and dependent on the physician's expertise. Thus, automatic strategies for the classification of the CVD severity are claimed. This paper proposed an automatic ensemble-based strategy of Deep Convolutional Neural Networks (DCNN) for the classification of CVDs severity from medical images. First, a clinical dataset containing 1376 images of patients' legs with CVD of 5 different levels of severity was constructed. Then, the constructed dataset was randomly split into training, testing, and validation datasets. Subsequently, a set of DCNN were individually applied to the images for classification. Finally, instead of a traditional voting ensemble strategy, extracted feature vectors from each DCNN were concatenated and fed into a new ensemble optimization network. Experiments showed that the proposed strategy achieved a classification with 93.8%, 93.4%, 92.4% of accuracy, precision, and recall, respectively. Moreover, compared to the traditional ensemble strategy, improvement in the accuracy of ~2% was registered. The proposed strategy showed to be accurate and robust for the diagnosis of CVD severity from medical images. Nevertheless, further research using an extensive clinical database is required to validate the potential of this strategy.**

Clinical Relevance— **An automatic classification of CVD to reduce the probability of underdiagnoses and promote the treatment of CVD in the early stages.**

I. Introduction

Chronic venous disorders (CVD) are one of the most common medical conditions in the world adult population, with prevalence among adults around 50% [1]. Due to the variety of signs and symptoms associated with CVD severity, correct diagnosis is essential to provide accurate treatment to the patients [2]. The signs of CVD are typically evaluated in terms of a structured clinical classification protocol named CEAP (Clinical, Etiologic, Anatomic, Pathophysiologic), which incorporates a wide range of signs and symptoms of CVD to describe its severity, ranging from C0 (no visible signs of venous disease) to C6 (active venous ulcer) [2].

To help the diagnosis and monitoring treatment evolution of CVD, digital photographs are regularly captured and stored by healthcare professionals [3]. These photographs allow an easier evaluation of the patient's condition by comparing the CEAP classification in different time periods [4]. Still, the diagnosis of CVD lesions relies on a visual inspection which is time-consuming and dependent on the physician's expertise.

To overcome similar medical problems, automated skin lesion classification methods have been proposed in the literature. For skin images in general, Deep Convolutional Neural Networks (DCNN) , such as EfficientNet, or Resnet architectures, proved to be effective in the diagnostic of skin lesions and reach prediction levels on par with dermatologists [5]. Recently, ensemble-based approaches have been demonstrated to perform better than individual 2D models for skin lesion classification [6], [7]. Specifically for CVD, DCNN focused on varicose veins [8] and skin ulcers classification and segmentation [4], [9] were proposed with promising results. Still, the performance of a DCNN for the classification of CVD severity has not been reported to date. Moreover, current ensemble strategies are based on an average of individual classification results, which neglects the performance of individual networks on the final result.

In this work, we proposed a fully automatic strategy based on an ensemble of DCNN for the classification of CVD severity levels according to the CEAP protocol. Instead of traditional ensemble strategies, the CVD classification result was obtained using an optimization-based ensemble strategy of all the network classification results. This strategy has the potential to improve the current performance of DCNN for skin lesion classification by learning the optimal combination of the classification results per image. To quantify the added-value of this strategy, a validation of the proposed pipeline against state-of-the-art classification methods was performed. Such tool can then be used to promote self-monitoring and aid the diagnosis and treatment follow-up of CVDs.

II. Methods

A. General Overview

The proposed framework relies on a DCNN methodology for CVD severity classification. Figure 1 shows an overview of the proposed methodology divided into three conceptual

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) and SFRH/BD/136670/2018 (H. Torres). 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 LASI-LA/P/0104/2020, UIDB/00319/2020, 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.

B. Oliveira and H. Torres are with 2Ai – School of Technology, IPCA, Barcelos, Portugal, with Algoritmi Center, School of Engineering, University of Minho, Guimarães, Portugal, with Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Braga, Portugal, and with ICVS/3B's - PT Government Associate Laboratory, Braga/Guimarães, Portugal (email: htorres@ipca.pt, boliveira@ipca.pt).

J. C. Fonseca is with Algoritmi Center, School of Engineering, University of Minho, Guimarães, Portugal (email: jaime@dei.uminho.pt).

P. Morais, L. Baptista, and J. L. Vilaça are with 2Ai – School of Technology, IPCA, Barcelos, Portugal (email: pmorais@ipca.pt, jvilaca@ipca.pt).

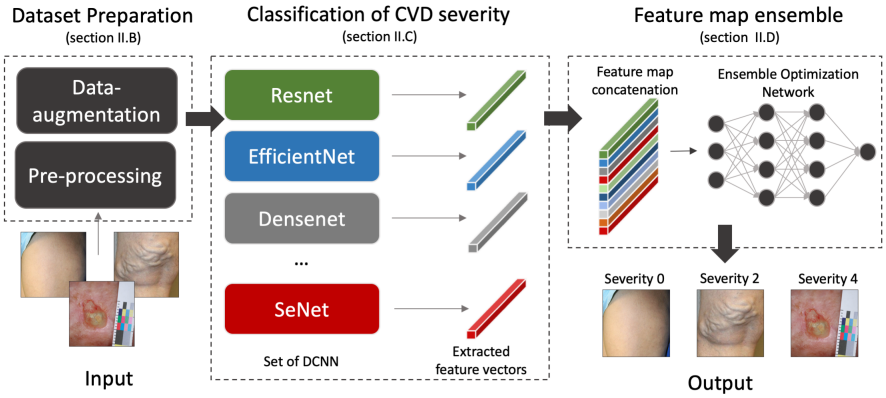

Figure 1 – Overview of the proposed CVD classification methodology.

blocks. In the first (section B), the dataset pre-processing and data augmentation techniques used to uniformize and increase the dataset size, respectively, are explored. This data is then fed to the classification blocks, whose DCNN architectures are fully explained in the second block (section C). Here, an RGB image is fed into a set of convolutional networks to extract representative high-level features. Lastly, the final ensemble optimization strategy that combines the results from different networks is explained in the third block (section D). The loss function developed to allow ensemble optimization strategy to learn the classification task is also described (section E). In the end, a severity probability vector of the patient condition according to 5 levels of severity are obtained: level 0 - no visible signs of venous disease; level 1 - telangiectasias and reticular veins; level 2 - varicose veins; level 3 - edema and skin changes including pigmentation and eczema; and level 4 - healed/active venous ulcers (Figure 2).

B. *Dataset preparation*

1) *Pre-processing*

The first step of the proposed methodology corresponds to the pre-processing of the data. Since the images were captured with different camera sensors, acquisition settings, and light conditions, they present different sizes. Thus, the images were initially resized to have the same minimum side. Then, a center crop for a fixed image size and a normalization of intensities are also applied.

2) *Data augmentation*

To deal with CVD variability, DCNN-based approaches require a large amount of data to correctly perform a specific task [4], [10]. Thus, to improve the generalization capacity and overcome overfitting, data augmentation was performed. Here, spatial-based (i.e. random flip, rotation, scaling, grid, distortion, and elastic transformations) were used to give robustness to the variability of digital cameras and lesion shapes. Moreover, pixel-based augmentation techniques (i.e. random Gaussian noise, brightness, contrast, and gamma transformations) were used to potentiate the robustness to lighting variability and lesions' appearance. The proposed augmentation is performed on the fly during DCNN training stage.

C. *Classification of CVD severity*

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 [11]. Recently, for skin images classification, an ensemble of DCNN has shown to improve the classification performance of individual DCNN [6]. Inspired by these methodologies, we propose to use an ensemble of DCNN to address CVD severity classification. In detail, to introduce variability in the final ensemble, a set of different networks already used for skin lesion classification tasks, were selected [6], [11], namely EfficientNeB3/B4 [12], ResNext [13], Resnet101/50 [14], DenseNet161 [15], InceptionV3 [16], SeNet50/101 [17], and WideResNet50 [18]. Each DCNN was individually optimized for the CVD classification task. Then, a set of high-level features for CVD classification extracted from each network was created.

D. *Ensemble strategy*

From the extracted vector of features, an deep ensemble strategy was implemented [19]. Traditionally, the final ensemble of DCNN for skin classification rely on an voting strategy (i.e. average of classification results), which is suboptimal since gives the same weight to all the networks [6], [7], [20]. In opposition, we propose to concatenate all the feature vectors and further train a simple neural network to learn the best ensemble of the feature vectors. In detail, the output features of the different DCNNs are concatenated and fed into a deep ensemble block composed of a ReLU layer, an adaptive average pooling layer, a fully connected layer, and a softmax layer to get the diagnostic probability of each CVD severity level.

E. *Loss function*

The proposed architecture is an end-to-end DCNN where a classification task must be learned. Similar to other classification frameworks [14], [15], we use the multi-class cross-entropy (CE) for the classification loss for training each DCNN and the final ensemble network. Thus, with $c \in (1,2 \ldots, C)$ being the class index, and C representing the number of CVD severity classes, the classification loss, L_{class} is given by:

$$L_{class}(\boldsymbol{y},\hat{\boldsymbol{y}}) = -\sum_{c=1}^{C} y_c * \log(\hat{y}_c), \qquad (1)$$

where y_c represents the ground truth label for each class c, and $\hat{y}_c$ is the network softmax output of the classification head for the same class c.

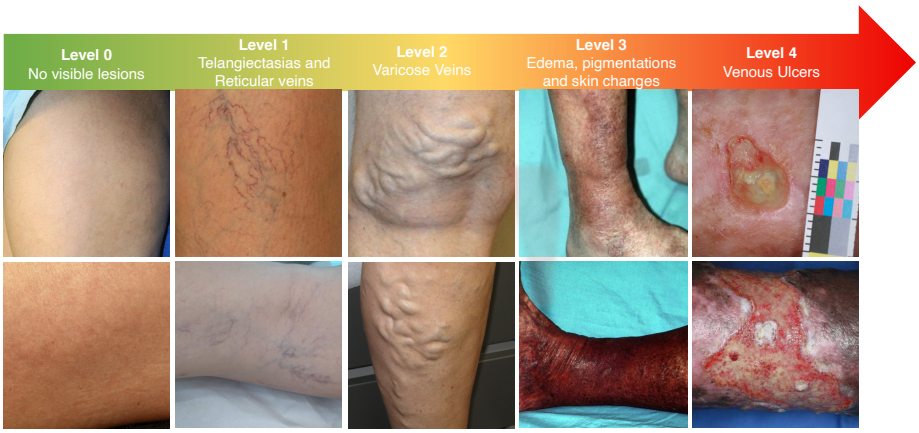

Figure 2 – Example of images from patients with different levels of CVD lesions, ranging from no visible or palpable lesions (level 0) to healed or active venous ulcers (level 4).

III. 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 x 384]. Such size was chosen as a trade-off between computational cost and DCNN requirements. After, an intensity normalization to the range [0, 1] was performed. All the augmentations were performed using the Albumentations framework [21]. The training of each classification network was performed in 500 epochs with a mini-batch size of 4 and using the Adam optimizer with an initial learning rate of 0.001 and a learning rate decay following the "poly" learning rate policy with the power of 0.9. To deal with feature vectors from DCNN with different sizes, an average pooling layer was used to pre-process each feature vector to a common vector size of 1500. For the ensemble optimization network, the training was performed with a mini-batch size of 8 and using the Adam optimizer with an initial learning rate of 0.01.

IV. Experiments

A. Data description

A new clinical database containing 1376 photographs of patients with CVD in lower limbs was constructed. The database was constructed from the normal clinical practice, where 502 images were collected from two public datasets, namely the 217 images from ULCER [22] and 305 from the SD-198 [23] datasets. From SD-198 dataset, 305 images of patients' legs with level 3 severity lesions were selected. These images correspond to lesions from the 5 different levels of CVD severity: level 0 – 223; level 1 – 237; level 2 – 127; level 3- 546; level 4 – 243 images. Figure 2 illustrates examples of the different images.

B. Evaluation strategies

To evaluate the accuracy and robustness of the proposed ensemble strategy (henceforward called EnsembleOPT), a comparison against individual DCNN and the traditional voting ensemble strategy, i.e. majority voting (henceforward called EnsembleMV) was performed. First, the dataset was randomly divided into training, testing, and validation datasets, respectively 80%, 10%, and 10% of the initial dataset. Then, all the networks evaluated were trained using the training dataset with the parameters defined in section 3. For each network, the convergence was assessed according to the network performance on the validation dataset. Finally, all the final results were computed on the testing dataset. The results obtained were computed using Python code running on an Intel (R) i7 CPU at 2-8 GHz with 16 GB of RAM and an NVIDIA GXT 1070 GPU with 8 GB of memory and cuDNN 10.1 libraries. The proposed strategy and all the other conventional neural network architectures were implemented using the Pytorch library.

V. Results

Table 1 shows the comparison between the performance of the proposed strategy and other conventional DCNN architectures for the classification task. Generally, EnsembleOPT presents the best average results against all the other networks, with an accuracy, precision, and recall of 93.8%, 93.4%, and 92.4% respectively. Looking for the confusion matrix in Figure 3, it is perceptible the improvement on the diagnosis using the proposed methodology for all severity levels against EfficientNetB4 (i.e. the DCNN with the best individual score) and EnsembleMV. Moreover, it can also be visualized that, both ensemble strategies considerably reduced the errors of mislabelled images from the highest CVD severity level in comparison with the EfficientNetB4.

VI. Discussion

In this work, we proposed an ensemble optimization strategy of DCNN for the robust classification of CVD severity on medical images. Overall, the proposed strategy showed competitive results in comparison with conventional DCNN and robust results for CVD severity classification.

Table 1 – Comparison of the proposed EnsembleOPT against DCNN and conventional EnsembleMV for CVD classification (mean±S.D.)

Model	Precision	Recall	Accuracy	AUC
EfficientNetB4	91.2±9.3	88.8±6.9	90.1	98.2±1.1
EfficientNetB3	89.8±11.9	89.1±11.9	89.4	98.4±2.9
Resnet50	86.1±12.5	83.6±6.7	85.1	96±1.9
Resnet101	84.1±13.0	84.4±6.2	84.5	95±3.1
Resnext	83.4±11.3	83.2±6.6	83.9	97±1.5
SENet50	89.1±10.9	88.5±6.8	88.8	96.6±2.5
SENet101	83.1±6.2	82.9±5.9	83.9	96.8±2.0
WideResNet50	82.3±19.5	76.5±13.1	78.3	94.1±3.6
DenseNet	87.8±11.4	85.9±7.8	87.6	97.1±1.9
InceptionV3	87.2±14	84.6±11.5	85.1	93.4±9.0
EnsenbleMV	92.4±11.8	91.6±7.4	91.9	**99±0.9**
EnsembleOPT	**93.4±5.5**	**92.4±7.6**	**93.8**	98.7±1.3

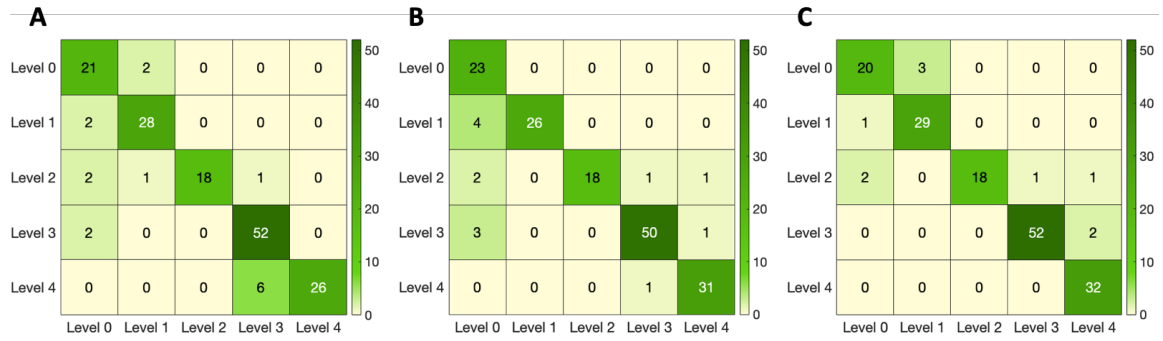

Figure 3 – Confusion matrix of the EfficientNetB4 (A), EnsembleMV (B), and proposed EnsembleOPT (C) for the classification of CVD images from the testing dataset. Confusion matrix in each row represent the true label, while the column represents the predicted classification.

Table 1 validates the performance of proposed strategy for the classification of CVD severity. Specifically, the performance of both ensemble strategies shows that the ensemble of multiple DCNN performed better than the individual strategies. When comparing specifically against EnsembleMV, better results were achieved by the proposed strategy, with an improvement of around 2% on the overall accuracy. Interestingly, narrow S.D, were also obtained for the ensemble strategy. These results proves that our optimization-based strategy better complemented the information extracted by each DCNN than the traditional ensemble strategy.

In Figure 3, it is possible to see that the best individual DCNN (EfficientNetB4) had difficulty on the classification of images from the highest CVD severity level. Contrarily, both ensemble strategies outperformed this strategy, mainly on the classification of such high severity images (i.e. venous ulcers). In computer-aided diagnosis tools, underdiagnoses have a superior cost for the patient, which highlights the importance of this result. These results corroborate the high accuracy of the method, demonstrating the advantage of the described strategy for the normal clinical practise.

VII. Conclusion

In this work, a novel pipeline for the automatic classification of CVD lesions according to the levels of severity described in CEAP is presented. The accuracy and robustness of this technique were initially validated in a new clinical database, demonstrating its potential for the current clinical practice. Overall, this strategy has the potential to aid the diagnosis, reducing the probability of underdiagnoses and promoting the treatment of CVD in the early stages.

References

[1] E. Rabe, J. J. Guex, A. Puskas, A. Scuderi, F. Fernandez Quesada, and VCP Coordinators, "Epidemiology of chronic venous disorders in geographically diverse populations: results from the Vein Consult Program," *Int Angiol*, vol. 31, no. 2, pp. 105–115, Apr. 2012.

[2] V. Vitalelewis, "Aesthetic Treatment of Leg Veins," *Aesthetic Surgery Journal*, vol. 28, no. 5, pp. 573–583, Sep. 2008.

[3] A. A. Meesters, L. H. U. Pitassi, V. Campos, A. Wolkerstorfer, and C. C. Dierickx, "Transcutaneous laser treatment of leg veins," *Lasers in Medical Science*, vol. 29, no. 2, pp. 481–492, Mar. 2014.

[4] D. Y. T. Chino, L. C. Scabora, M. T. Cazzolato, A. E. S. Jorge, C. Traina-Jr., and A. J. M. Traina, "Segmenting skin ulcers and measuring the wound area using deep convolutional networks," *Computer Methods and Programs in Biomedicine*, vol. 191, p. 105376, Jul. 2020.

[5] M. A. Al-masni, M. A. Al-antari, M.-T. Choi, S.-M. Han, and T.-S. Kim, "Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks," *Computer Methods and Programs in Biomedicine*, vol. 162, pp. 221–231, Aug. 2018.

[6] N. Gessert, M. Nielsen, M. Shaikh, R. Werner, and A. Schlaefer, "Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data," *MethodsX*, vol. 7, p. 100864, 2020.

[7] B. Harangi, A. Baran, and A. Hajdu, "Classification Of Skin Lesions Using An Ensemble Of Deep Neural Networks," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Honolulu, HI, Jul. 2018, pp. 2575–2578.

[8] R. Zhu, H. Niu, N. Yin, T. Wu, and Y. Zhao, "Analysis of Varicose Veins of Lower Extremities Based on Vascular Endothelial Cell Inflammation Images and Multi-Scale Deep Learning," *IEEE Access*, vol. 7, pp. 174345–174358, 2019.

[9] G. Blanco *et al.*, "A superpixel-driven deep learning approach for the analysis of dermatological wounds," *Computer Methods and Programs in Biomedicine*, vol. 183, p. 105079, Jan. 2020.

[10] E. Pérez, O. Reyes, and S. Ventura, "Convolutional neural networks for the automatic diagnosis of melanoma: An extensive experimental study," *Medical Image Analysis*, vol. 67, p. 101858, Jan. 2021.

[11] C. Barata, M. E. Celebi, and J. S. Marques, "A Survey of Feature Extraction in Dermoscopy Image Analysis of Skin Cancer," *IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS*, vol. 23, no. 3, p. 14, 2019.

[12] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *arXiv:1905.11946*, Sep. 2020.

[13] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated Residual Transformations for Deep Neural Networks," *arXiv:1611.05431 [cs]*, Apr. 2017.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *arXiv:1512.03385 [cs]*, Dec. 2015.

[15] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," *arXiv:1608.06993 [cs]*, Jan. 2018.

[16] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," *arXiv:1512.00567 [cs]*, Dec. 2015.

[17] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-Excitation Networks," *arXiv:1709.01507 [cs]*, May 2019.

[18] S. Zagoruyko and N. Komodakis, "Wide Residual Networks," *arXiv:1605.07146 [cs]*, Jun. 2017.

[19] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, "Ensemble deep learning: A review," *arXiv:2104.02395*, Mar. 2022.

[20] A. A. Milton, "Automated Skin Lesion Classification Using Ensemble of Deep Neural Networks in ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection Challenge," p. 4.

[21] A. Buslaev, A. Parinov, E. Khvedchenya, V. I. Iglovikov, and A. A. Kalinin, "Albumentations: fast and flexible image augmentations," *Information*, vol. 11, no. 2, p. 125, Feb. 2020.

[22] Ederson. A. G. Dorileo, M. A. C. Frade, A. M. F. Roselino, R. M. Rangayyan, and P. M. Azevedo-Marques, "Color image processing and content-based image retrieval techniques for the analysis of dermatological lesions," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vancouver, BC, Aug. 2008, pp. 1230–1233.

[23] X. Sun, J. Yang, M. Sun, and K. Wang, "A Benchmark for Automatic Visual Classification of Clinical Skin Disease Images," in *Computer Vision – ECCV 2016*, vol. 9910, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 206–222.