

Diabetic foot ulcer segmentation using logistic regression, DBSCAN clustering and morphological operators

Armando Heras-Tang*, Damian Valdés-Santiago*, Ángela León-Mecías*,
Marta L. Bager Díaz-Romañach*, and José A. Mesejo-Chiong*

* *Department of Applied Mathematics, Faculty of Mathematics and Computer Science, University of Havana, Havana, Cuba*

Received 16th of March, 2021; accepted 14th of August 2022

Abstract

Digital images are used for the evaluation and diagnosis of diabetic foot ulcers. Selecting the region of interest in an image is a preliminary step for subsequent analysis. Most of the time, manual segmentation is not very reliable because specialists may have different opinions regarding the ulcer border; which has encouraged researchers to find and test automated segmentation techniques. This paper presents a computer-aided ulcer region segmentation algorithm for diabetic foot images. The proposed algorithm has two stages: ulcer region segmentation and post-processing of segmentation results. The first stage focused on the selection of a trained machine learning model to classify pixels inside the ulcer's region, after comparing five learning models. We performed exhaustive experiments with our own annotated dataset of images of Cuban patients. A second stage was needed due to the existence of some misclassified pixels. In order to solve this, we applied the DBSCAN clustering algorithm, together with dilation and closing morphological operators. The best-trained model after the post-processing stage was the logistic regressor (Jaccard Index 0.81, accuracy 0.94, recall 0.86, precision 0.91, and F1 score 0.88). This trained model was sensitive to irrelevant objects in the scene, but successfully detected the patient's foot. Physicians found these results promising for measuring the lesion area and to monitoring the ulcer healing process during treatments, which would significantly reduce errors.

Key Words: Diabetic foot ulcer, Supervised learning, Image segmentation, DFU.

1 Introduction

Diabetes Mellitus (DM) is a chronic metabolic disease characterized by high blood sugar levels during long periods. It has a high prevalence on developing countries. And, according to the International Diabetes Federation, one in 11 adults (globally about 463 million) lives with this condition [22]. As for Cuba, according to the Cuban Ministry of Public Health, by the end of 2018 diabetes prevalence was 64.3 per 1 000 people, which represents approximately 6.5% of the population [10].

Foot ulcers are a significant complication of DM. Diabetic foot ulcers (DFUs) are open wounds that can develop on the feet of people with diabetes. In Cuba 3.5% of diabetic patients develop DFUs, meaning that

Correspondence to: <dvs89cs@matcom.uh.cu>

Recommended for acceptance by Angel D. Sappa

<https://doi.org/10.5565/rev/elcvia.1413>

ELCVIA ISSN:1577-5097

Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

more than 18,000 people are affected each year [18]. Improperly treated DFUs often precede lower extremity amputation [14]. Of all diabetic patients who have undergone lower limb amputations, 85% have suffered from ulcers [15].

A thorough and systematic evaluation of DFUs is essential for appropriate treatment. Etiology, location, size, depth, severity, tissue and edges types are the main elements of DFUs clinical evaluation [15]. Periodic monitoring of DFUs, which is currently performed manually by clinicians, is imperative to assess the progress of ulcer healing and avoid amputations [46]. However, manual measurement of the wounds' size is often inaccurate, time-consuming, and can also cause discomfort to the patient [9].

One approach followed by clinicians is to take high-resolution digital images of ulcers during the initial evaluation and also in subsequent reviews, allowing comparisons spanning distinct healing stages at different periods of ulcer progression. Visually, the ulcer tissues have three different colors: black, indicating necrosis; white or yellow, indicating septicity; and red, indicating granulation and healing [15, 49]. Unlike other medical images there is no specific equipment to capture DFU images; therefore, pictures are usually taken using digital cameras or smartphones.

Selecting the DFU region in a digital image refers to locating the boundary between the ulcer area and the surrounding skin. This task, known as DFU segmentation, is a crucial preliminary step for subsequent image analysis. However, due to the lack of a standard protocol regarding the position of the camera and the ambient light conditions while taking the pictures, the images may have different scales and lighting quality. Moreover, the problem is usually complex because ulcers can have different shapes and their position can cover several areas of the foot [15].

DFU segmentation can be performed manually using common image processing software like ImageJ [35] or GIMP [41]. Also, special purpose software, such as MITK Workbench, offer manual segmentation functionalities [44]. Nevertheless, manual segmentation is highly dependent on the observer and their visual perception [32]. The need to develop image segmentation methods that will work with images acquired in regular clinical conditions and the developments in digital image processing and analysis have led a substantial amount of research in the automatic detection and evaluation of DFUs.

Accurate and automatic DFUs evaluation strongly relies on well-segmented regions. As a result, DFU segmentation has turned out to be an emergent area in the field of image segmentation. Several segmentation algorithms have been developed to perform automated DFU region detection [16, 24, 32]. State-of-the-art machine and deep learning object detection frameworks are increasingly applied to this problem [19, 29, 46, 49].

In [43], the authors used SVM to determine the wound boundaries of DFU images. Those images were taken using a specific capture box created by the authors where patients place their feet and physicians take the picture with a smartphone. A superpixel segmentation was applied using the Simple Linear Iterative Clustering (SLIC), and then a two-stages cascade classifier was used. They obtained a mean sensitivity (recall) of 73.3%, and 94.6% specificity.

The annotation and estimation of DFU boundaries using images were proposed by Dhane et al. [11]. Their approach involved a gray level based fuzzy similarity measure using spatial information of the images. Then, a fuzzy spectral clustering (FSC) was applied on the best color channel found after computing the mean contrast for 26 different color channels of 14 color spaces. Then, the segmented regions were processed using several morphological operations. The FSC approach was tested on 70 images, obtaining an accuracy of 91.5%, a Jaccard Index of 79.0% and a sensitivity (recall) of 87.3%.

The goal in [1] was to develop a method for segmenting, measuring and characterizing the clinically chronic wounds through images. The first step of the proposed method was to create a Red-Yellow-Black-White probability map (RYKW), which guided the segmentation process using an optimal thresholding. They reported an average segmentation precision of 75.1%.

In [42], a deep learning method based on a variant of deep convolutional neural networks (ConvNet) was proposed. For the segmentation task, they reported an accuracy of 95.0% and a Jaccard Index of 47.3%.

This paper proposes a pixel features based two-step hybrid machine learning algorithm for automatic DFU segmentation with results not far from state-of-the-art algorithms reported in literature. We used a group

of pixel-based features to train four machine learning models for classifying pixels that belong to the ulcer. Generally, classification algorithms can assign elements to the wrong class. Therefore, we performed a post-processing step to improve segmentation. First, we removed misclassified pixels from the segmented region using DBSCAN clustering to discriminate isolated pixels and small groups. We chose this algorithm as it is more robust compared to other clustering methods like fuzzy clustering [25]. We also applied the erosion operator followed by the closing operator in order to fill the remaining holes inside the ulcer region, and isolate pixels close to the wound edges [7].

The results of our research are relevant for designing a computer-aid software that helps Cuban physicians monitor DFU healing during treatment. For now, this process is done manually with the corresponding drawbacks of this measuring method, that we already mentioned.

The remainder of this paper is organized as follow: Section 2 explains the proposed pipeline for DFU automatic segmentation. Section 2.1 details the feature extraction, classification models for determining the pixels belonging to the wound, and the results of the classifiers. Section 3 shows why and how we made the ulcer area refinement. Section 4 presents the dataset construction, image pre-processing, and the results of the post-processing stage. Finally, we present conclusions and remarks.

2 Proposed method

This section explains the sequence of steps followed in the DFU segmentation algorithm (see Figure 1). The first and primary step is feature extraction. Many machine learning algorithms, like the one used in this work, require a numerical representation of the objects under study. Since we are working with images, in which the color of the objects is essential for their classification, a vector numerical representation of the pixel's color is sought. This vector is what we define as pixel features. Section 2.1 explains the pixel features extraction process in detail.

The feature extraction step is followed by a supervised classification one. Supervised classification is a machine learning technique that maps an input feature vector to an output label based on a set of correctly labeled training data. The output labels that are defined for our supervised classification process are ulcer and non-ulcer. Given these labels, the segmentation of the image is carried out. Five different classifiers [3] are used: Random classifier (DC), Decision Tree (DTC), Gaussian Naive Bayes (GNB), Logistic Regression (LR), and Support Vector Machine (LSVC). The results of this step are shown in Section 2.2. Finally, a post-processing step is performed. This step is necessary because pixels assigned to the wrong class may appear. Details of this post-processing are shown in Section 3.

Dummy classifier (DC) or Random classifier is a type of classifier that makes predictions without trying to find patterns in the data, and classifies using only simple rules. To predict the class label it uses one strategy, completely ignoring the trends in the training data, which makes the classifier's behavior completely independent of the training data. In this work the stratified strategy was used, it generates predictions by respecting the class distribution of the training data. A DC is used as a baseline to compare with other classifiers performance.

Decision tree classifier (DTC) builds a hierarchical structure to split the dataset into homogeneous groups. It is a finite, acyclic, and directed graph that contains a unique node that has no parent called root. Each node of the tree has one parent, except the root, and there is a unique path from the root to every node of the graph. Each non-leaf node of the tree represents a test on a feature set, also called split. The leaves are associated to labels (or categories). To classify an unseen instance, various tests are conducted, starting from the root, until reaching a leaf node. The computational complexity of the decision tree is associated to its depth and average nodes per level [17].

Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem. It is an easy way to predict classes both in binary (two class) and multiclass classification problems [17]. They belong to the family of learners that use the maximum a posteriori rule to perform the inference process. They give probabilities that a given pattern belongs to a specific class, and in the case that the independence of the data

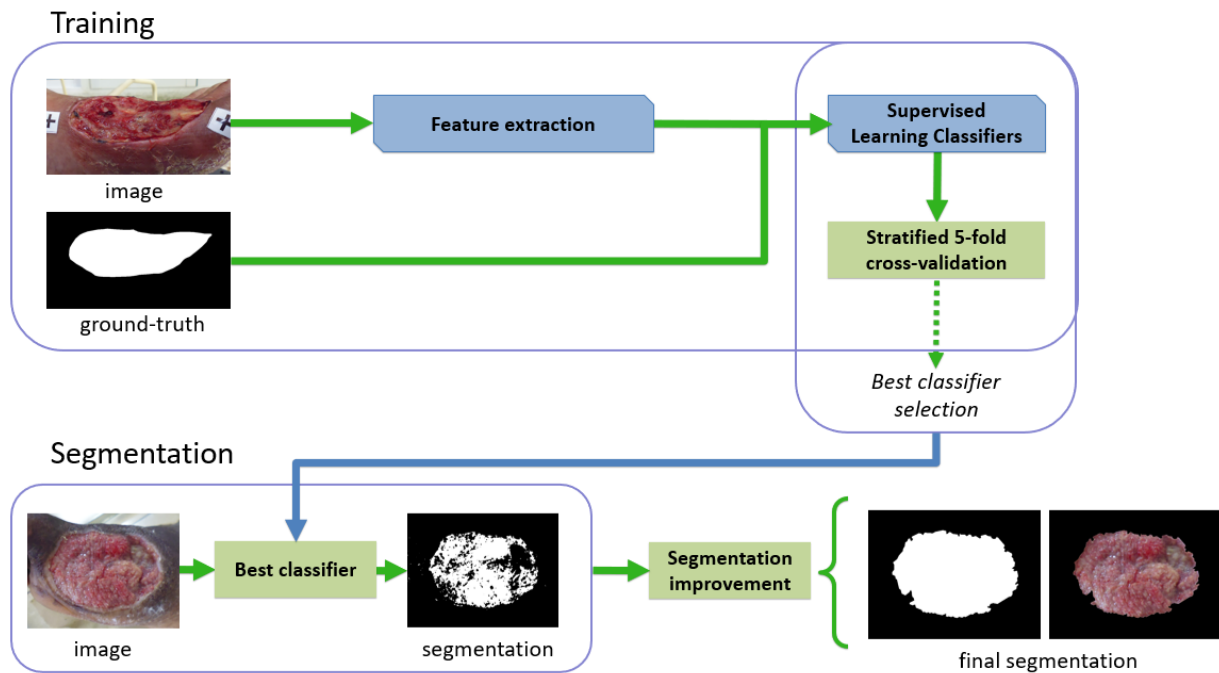


Figure 1: Pipeline of the proposed method. The first stage is the classifier training stage using an stratified 5-fold cross-validation to select the best classifier, which is then used to obtain the segmentation. Then, an improvement stage is performed, obtaining the final segmentation.

attributes assumption fits, the algorithm performs better compared to other classification models. It does not require as much training data [17]. Gaussian Naive Bayes (GNB) is an extension of Naive Bayes to real-valued attributes, most commonly by assuming a Gaussian distribution. The Gaussian or Normal Distribution is the easiest to work with because you only need to estimate the mean and the standard deviation from your training data.

Logistic Regression developed by statisticians is mostly used for binary classification problems. It predicts the probability of an incident based on a set of values used as predictors. The learning function is a logistic one depending on optimal parameters which minimize the classification errors.

Support Vector Machine classifier (LSVC) [27] aims at finding the optimum hyperplane that separates instances of different classes. In a binary classification problem, the hyperplane must maximize the margin between both classes and the margin is defined as the minimum distance from instances of different categories to the hyperplane. The hyperplane is built using a kernel function; in this work a linear function. SVM is a widely used machine learning technique [47].

2.1 Feature extraction

For this work, we used pixels as samples; therefore, we needed to obtain features that represent them and create their features vector. Images store color intensity values for each pixel that depends on a color space [36]. Color plays an important role in DFU image segmentation, because they are related to different tissue features. Mukherjee et al. [28] implemented a classifier for different ulcer tissues. The images were transformed to HSL color space and the S component was selected as feature. Dhane et. al. [11] used 26 color channels for classifying the ulcer pixels.

Hence, we used the following features: intensity values of red, green and blue image components (RGB), normalized RGB or chromaticity (Cro), average intensity of the three color channels [39], red-green channel contrast (rg), red-blue channel contrast (rb) and green-blue channel contrast (gb) [39], the standard deviations

for each RGB channel using a window of radius 2, hue (H), saturation (S) and value (V) in a HSV color space [34]; hue (a^*), saturation (b^*) in a $L^*a^*b^*$ color space [5]; and the luminance (Y), blue chrominance (Cb) and red chrominance (Cr) components of a YCbCr color space [36].

We used combinations of all these features (Table 1) to characterize each pixel and their neighborhood using color information. We did not use all possible features because it could lead to a wrong generalization of the learning model [31].

Identifier	Combination
1	RGB
2	HSV
3	LAB
4	YCbRb
5	RGB,Cro,I,rg,rb,gb
6	HSV,Cro,I,rg,rb,gb
7	LAB,Cro,I,rg,rb,gb
8	YCbRb,Cro,I,rg,rb,gb
9	RGB,HSV
10	HSV,LAB
11	RGB,LAB
12	RGB,YCbRb
13	AB,Cb,Cro
14	RGB,HSV,LAB,YCbCr
15	Cro,H,GB,neigh2std
16	LAB,neigh2std
17	RGB,Cro,neigh2std
18	RGB,Cro,I,rg,rb,gb,neigh2std

Table 1: Feature combinations that define 18 different experiments for each classifier.

2.2 Classifier results

We used a 5-fold stratified cross-validation [20], which balances the proportion of the classes between the created subsets [12, 40] and allows the evaluation of the performance of the learning model during the training stage [40]. We used the following metrics to evaluate the results of the segmentation: accuracy, recall, precision, F1 score, area under the precision-recall curve (AUCPR) [6], and Jaccard Index (JI) [48].

We compared the results of the five classifiers mentioned in subsection ???. They were trained using the feature combinations in Table 1. Figures 2-4 shows its metrics values using 5-fold stratified cross-validation.

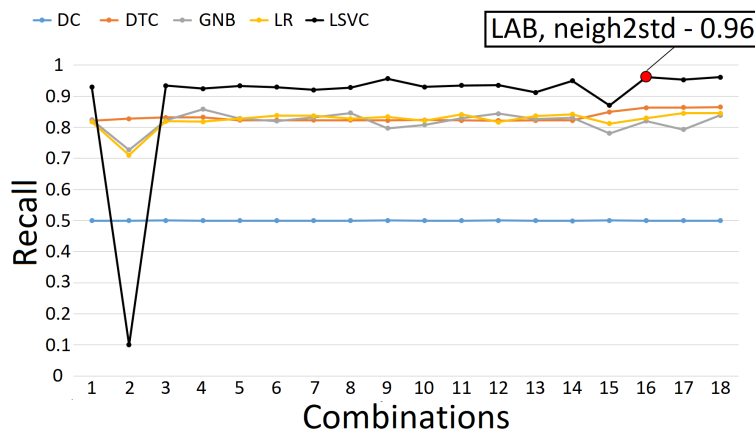


Figure 2: Recall values of each feature combination for every classifier. The maximum recall value (0.96) was reached for L SVC on feature combination 16.

Figure 2 shows that L SVC is the classifier with the best recall value, with all but one feature combination,

reaching a maximum of 0.96 for feature 16 (LAB, neigh2std). However, we cannot select this classifier based only on this result, because if a binary classifier labels all samples as the target class it will have a maximum recall. This would mean that all samples in the target class were correctly classified, but nothing is known about the precision of the classification. There may be many false positives (*FP*), affecting precision.

In Figure 3 we can notice, in general, good precision with DTC, having a maximum of 0.75 with the feature combination 4 (YCbRb). However, combination 16 (LAB, neigh2std) for LSVC, which has the highest recall, has a precision value of 0.41. In this particular case, we preferred to find a balance between recall and precision, and in order to do so, we analyzed the *AUCPR* metric (Figure 4).

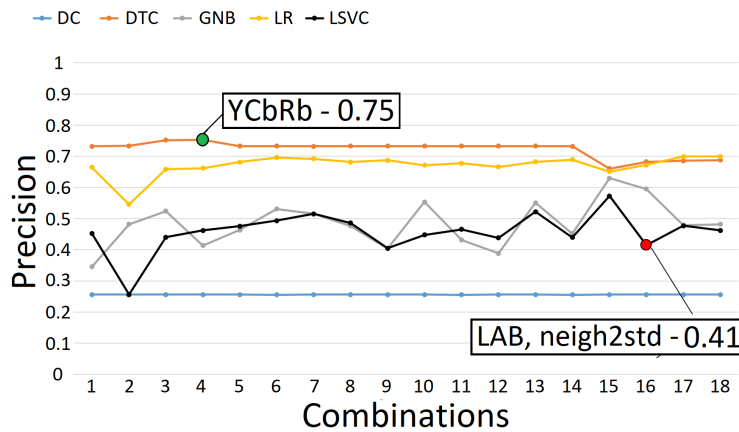


Figure 3: Precision values of each feature combination for every classifier. The maximum precision value (0.75) was reached for *DTC* on feature combination 4.

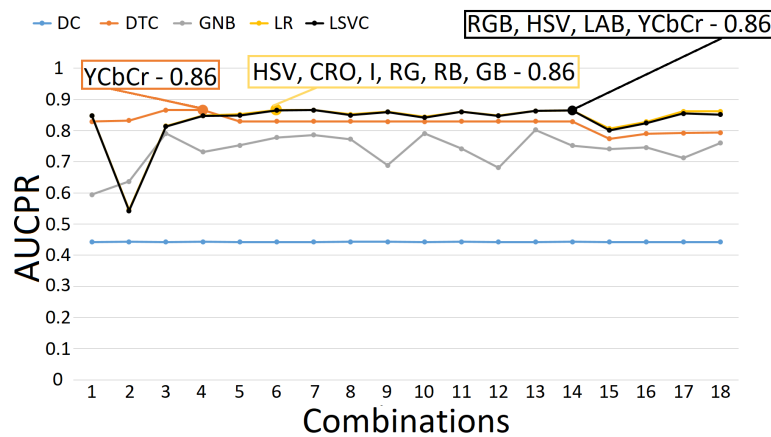


Figure 4: AUCPR values of feature combination for every classifier. The maximum AUCPR value (0.86) was reached for *LR*, *DTC*, and *LSVC* on feature combinations 6, 4 and 14, respectively.

We can see that LR has the best values for the *AUCPR* metric, except for combinations 2, 3, and 4 (HSV, LAB and YCbRb, respectively) which are outperformed by DTC. Also, LSVC shows better results for combination 14 than the other classifiers (RGB, HSV, LAB, YCbCr). Finally, the two best values were obtained by LR with 0.86 for combination 6 (HSV, CRO, I, RG, RB, GB), and DTC with 0.86 for combination 4 (YCbRb).

After analyzing these results, the classifiers DTC, LR and LSVC were selected with feature combinations 4, 6, and 14, respectively, because they were the ones with the best *AUCPR* values. These three models were re-trained with the entire dataset and used on the validation set.

Figure 5 shows that LR has the best results on the validation dataset with combination 6 (HSV, Cro, I, RG, RB, GB). Therefore, it was selected among the ones used in this research as the best supervised binary classifier to perform the pixel-wise segmentation.

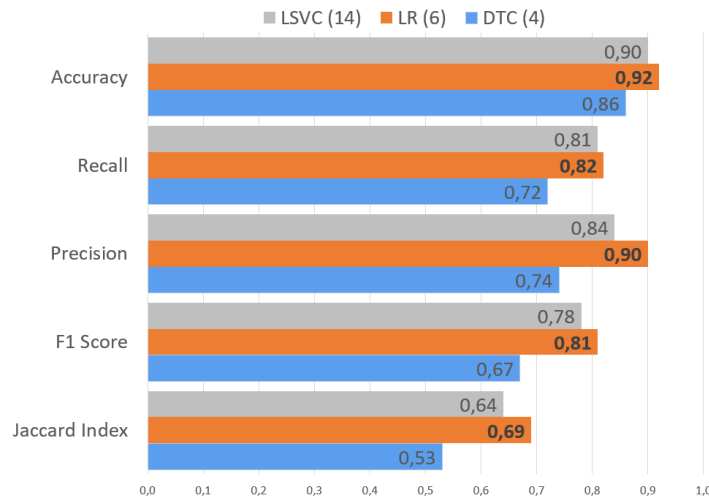


Figure 5: Classifiers results on the validation set. The logistic regression model with the combination 6 had the best results (highlighted in bold).

Figure 6 shows images with segmentation results. We can see that there are groups of isolated points, deformation, and gaps in the regions of interest of the DFU, because of misclassified pixels. We think this is a direct consequence of the lack of uniformity in the image acquisition process (unwanted objects in the frame, elements with color information similar to the ones in the region of interest of the DFUs and their internal tissues). For this reason, we applied a segmentation improvement process, which is described in the next section.

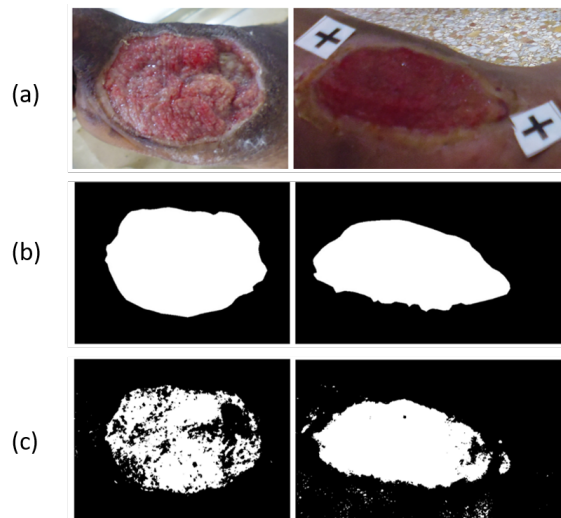


Figure 6: (a) Original images, (b) Segmented images using LR and feature combination 6, (c) Ground-truth images.

3 Post-processing

Binary classifications algorithms are defined as a function approximation. The goal is to estimate that function based on a labeled training set to be able to make predictions on unknown samples [30]. The approximation nature of this process does not always yields perfect results, therefore, we performed a post-processing step to improve segmentation. First, we removed misclassified pixels from the segmented region using an unsupervised clustering algorithm to discriminate isolated pixels and small groups. We use the Density Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm which is a model capable of generating groups with arbitrary form and with high scalability [45]. Finally, we use morphological operators on the final group.

After creating groups using DBSCAN, and with the one with the higher density selected, we proceed to apply morphological operators [4] using an ellipsis shaped structuring element with 5 pixels as diameter. The goal of this process is to fill holes inside the ulcer region, and also remove isolated pixels close to the edges. We apply the erosion operator followed by the closing operator.

3.1 Procedure describing the DBSCAN clustering

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a base algorithm for density-based clustering. It is able of finding clusters of different shapes and sizes from a large amount of data containing noise and outliers [25].

The algorithm proceeds by arbitrarily picking up a point in the dataset (until all points have been visited). If there are at least 10 points within a radius of 0.3 to the point then we consider all these points to be part of the same cluster. The clusters are then expanded by recursively repeating the neighborhood calculation for each neighboring point. We used the sklearn [33] implementation of DBSCAN for Python programming language.

3.2 Morphological operators

Morphological operators allow image processing by introducing transformations from the shape perspective. Set theory is the basic mathematical foundation of morphological operators that were first defined for binary images and later extended for use in grayscale images. In the same way morphological operators can be used in color images, either because they are transformed into grayscales ones or because they work with the intensities in each of the color channels, for example, in RGB space (see Nixon, Mark S.; Aguado, Alberto S., "Feature Extraction and Image Processing for Computer Vision", Third Edition, Elsevier, 2012).

The structuring or structural element defines the shape with which the operation is performed. It is important to define the center, the neighborhood and its dimension. Structural elements could be rectangular or non-rectangular.

3.2.1 Erosion and Dilation

The basic morphological operations are Erosion (\ominus) and Dilation (\oplus). In a binary image the pixels that belong to the image, foreground, have the value 1 and those with value 0 belong to the background. The operations are performed by positioning the center of the structuring element in each pixel of the image. The structural elements also define the neighborhood [38].

In the "eroded" image each visited foreground pixel "goes" to the background (value 0) if any pixel of the neighborhood defined by the structuring element is from the background. When "dilating" the image the visited background pixels remain in the background if none of the neighborhood is from the foreground. If at least one is from the foreground, its value is changed to 1.

It is easy to see that the effect of erosion removes details from the image, thins the contours and often disconnects some structures. On the contrary, dilation increases pixels in the foreground, thickens edges, structures and fills some small spaces in relation to the dimension of the structuring element.

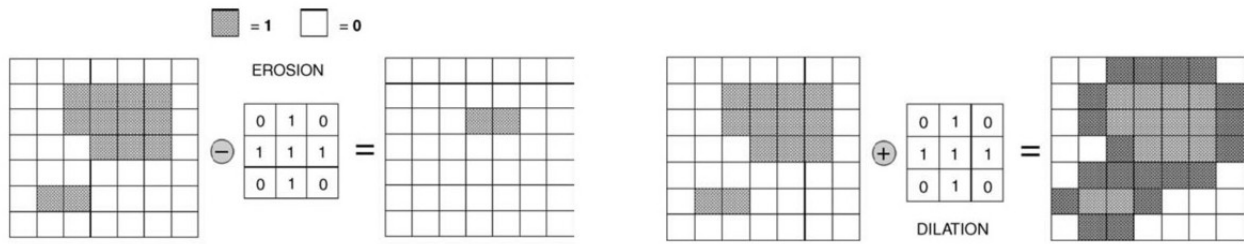


Figure 7: Erosion and Dilation operations by graphics (Figure from [38]).

Thus explained the operations can be seen that when moving to grayscale images the erosion corresponds to the application of the minimum between the pixels. For this reason, by successively eroding the image becomes darker while the dilation that corresponds to the determination of the maximum between pixels makes the image lighter in successive applications.

3.2.2 Opening and Closing

Opening and Closing are combinations of erosions and dilations. Opening = Erosion followed by dilation $A \circ B = (A \ominus B) \oplus B$ and Closure = Dilation followed by Erosion $A \bullet B = (A \oplus B) \ominus B$.

The effect on the image of an Aperture is the removal of small isolated objects and very fine connections between objects. Closure eliminates small gaps, captures small regions, and joins small objects.

3.2.3 Application of morphological operators in post-processing of segmentation results

Frequently segmentation produces images with thin and partially disconnected edges and small objects inside the image. Thus post-processing plays an important role.

To improve the results the obtained image is dilated followed by a closing to remove small unconnected parts from the segment border and to remove small objects also in the interior. Due to the shape of the segment (diabetic foot ulcer) the best results were obtained with an ellipse structuring element. For more details see Figure 8 and the comments in 4.3.

4 Results

In this section we present how the dataset was created and processed in the experiments as well as the performance metrics used. We also report the results of the segmentation task before and after the post-processing improvement step.

4.1 Dataset construction

We used 140 images taken at the Diabetic Foot Clinic at the National Institute of Angiology and Vascular Surgery (INACV, by its Spanish acronym) in Havana, Cuba, from a group of 30 patients during the treatment with Heberprot-P® [2]. The digital camera was attached to the FrameHeber 03 [8], a device developed at the Center for Genetic Engineering and Biotechnology, Cuba (CIGB, by its Spanish acronym) with a resolution of 3648×2736 pixels.

Photographs were not captured in a controlled environment. This lack of uniformity and control causes the existence of color information similar to the ulcer tissues on the image background. This could lead to wrong segmentation results by the proposed method.

After image acquisition and patient's follow-up [42], images were manually cropped so that the foot appeared as the main element of the scene. Then, they were resized to smaller resolutions while maintaining the aspect ratio.

Finally, 37 images with different types of ulcers were selected. Those images were manually segmented by the authors, using the software *DFUlcer-UH: Annotation Tools*® [21]. Then, the constructed dataset was divided into two groups: 70% (26 images) for training, and 30% (11 images) for validation.

4.2 Subsampling

Among the 26 images used in the training set, there is a total of 9,744,190 pixels. The number of pixels belonging to the class "ulcer" represents approximately 25% of the total number of pixels. This imbalance between classes poses a problem, because classifiers cannot generalize the behavior of the minority class, and thus, the performance decreases in terms of prediction of this class [23]. A random subsampling [26] was performed on the majority class in the training set, discarding elements from the "no ulcer" class until both classes had the same number of samples.

4.3 Post-processing results

The unsupervised clustering algorithm DBSCAN [13] was applied to the classification result to create groups of nearby pixels using the Euclidean distance in the two-dimensional space. We only considered those with more than 100 pixels in order to discard groups representing noise. Then, we selected the region with the maximum number of pixels, and dealt with the gaps applying a dilation operator followed by a closing operator (Figure 8) with an elliptical structural element with a diameter of 5 pixels [4, 37]. Figure 9 shows an example of the segmentation improvement process.

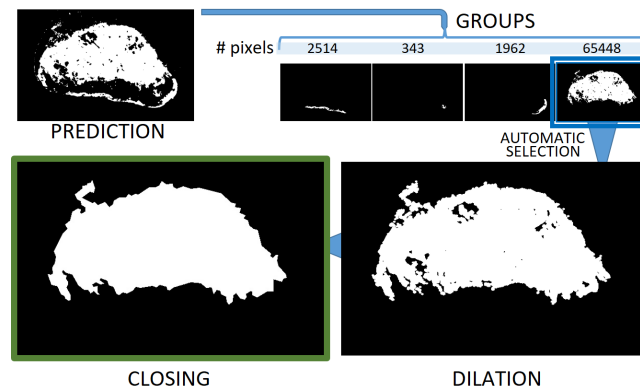


Figure 8: Pipeline of the segmentation improvement algorithm.

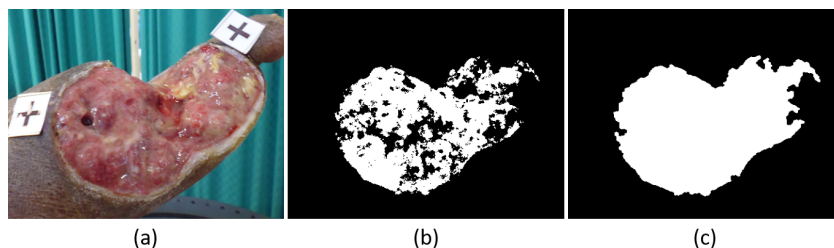


Figure 9: (a) Original image (b) Segmented image (c) Image after segmentation improvement.

To validate the post-processing performance, we improved the binary images obtained by the classifier on the validation set, and compared them with the *ground-truth* image, using the quality measures. Figure 10 compares these results with the ones without improvement. The metrics in Figure 10 make it clear that the post-processing leads to significant improvement on the validation set. The JI showed maximum improvement indicating a better overlap of the segmented region with the *ground-truth*. Figures 11 and 12 show the best and worst results of the refinement process, respectively.

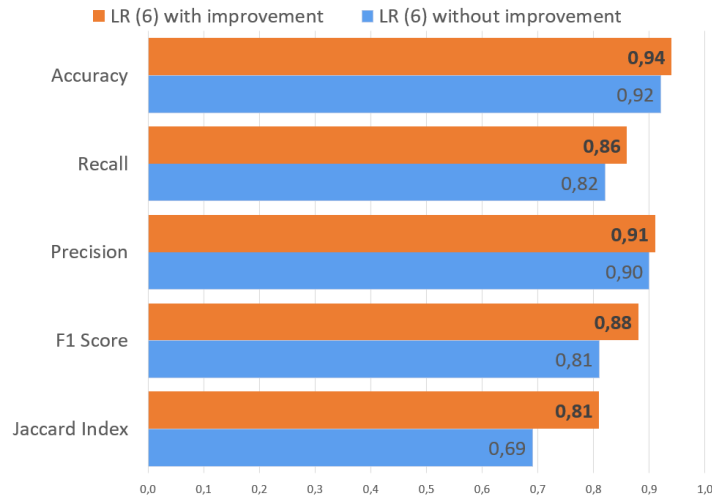


Figure 10: Metrics results for *LR* classifier using feature combination 6 before and after segmentation improvement algorithm.

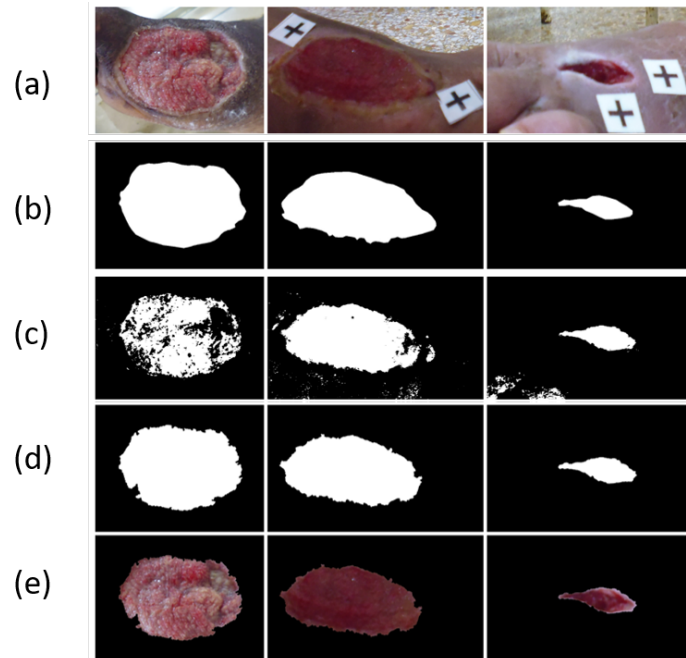


Figure 11: Results with higher *JI*. (a) Original image. (b) *Ground-truth*. (c) Image after segmentation improvement. (d) Refined segmented image. (e) Mask after segmentation improvement applied on the original image.

When compared our results with the state-of-the-art methods, our method has a higher precision and *JI*

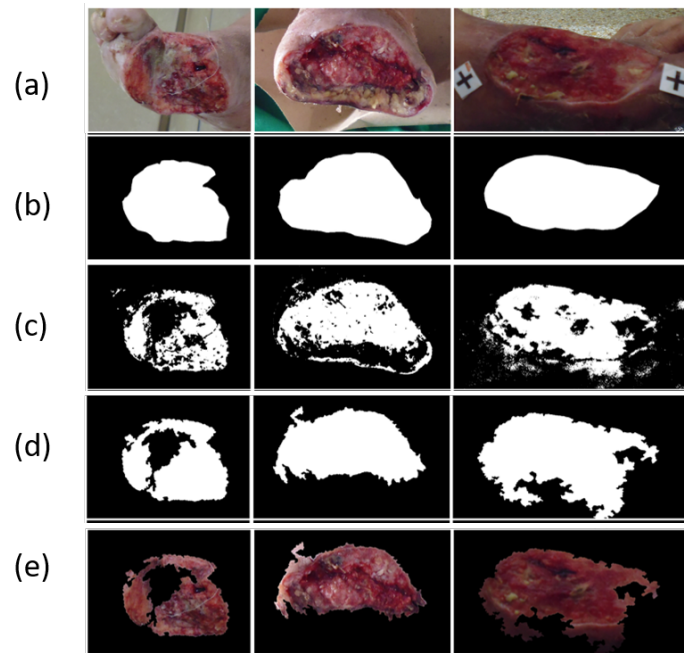


Figure 12: Results with lower JI . (a) Original image. (b) *Ground-truth*. (c) Image after segmentation improvement. (d) Refined segmented image. (e) Mask after segmentation improvement applied on the original image.

(Figure 13). Also, the recall and accuracy results are relatively close to the best ones (Figure 14). Although the algorithms in the literature were not run over our dataset, this comparison is an indicator of the reliability of our results. Figures 13 and 14 show that our results are not far from the state-of-the-art.

5 Conclusions and Discussion

We proposed a novel two-step hybrid algorithm to perform the DFU region segmentation. It combines an automatic image segmentation algorithm based on binary pixel-wise classification with an unsupervised density-based clustering method to improve the first initial segmentation.

We were unable to get access to a free annotated dataset of DFU images. Therefore, we built an annotated dataset to train learning models. We trained and compared four machine learning algorithms to select the best one after the improvement step.

We could not compare our method with other state-of-the-art methods, because the learning models and the dataset used in the cited research were not available.

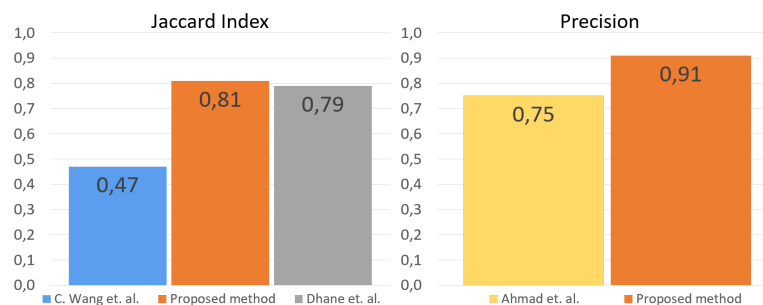


Figure 13: JI and Precision values of the proposed method versus the state-of-the-art reviewed methods.

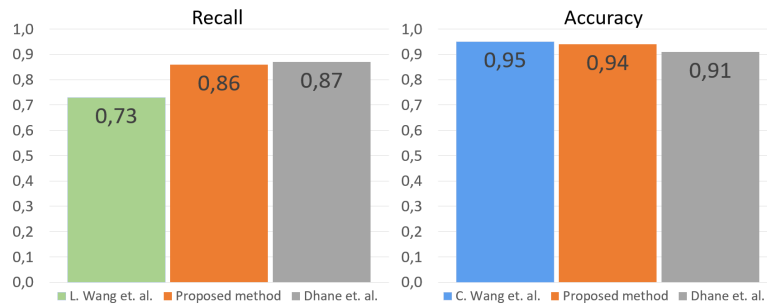


Figure 14: Recall and Accuracy values of the proposed method versus the state-of-the-art reviewed methods.

Even though images were not captured in a controlled environment, the proposed algorithm showed promising results. The logistic regressor gives the best result: with a Jaccard Index of 0.81, an accuracy of 0.94, a recall of 0.86, precision of 0.91, and an F1 score 0.88. Then, our method can be successfully used to perform the DFU region segmentation.

Acknowledgements

We would like to thank the patients, volunteers, and co-workers of The Cuban National Institute of Angiology and Vascular Surgery (INACV, by its Spanish acronym), La Habana, Cuba, specially to Dr. Jose Ignacio Fernández Montequín, Ph.D. The present study was reviewed and approved by the Ethics Committee of the INACV according to the ethical principles of the World Medical Association (Declaration of Helsinki). Written informed consent was obtained from all healthy volunteers (without diabetes) and from all people with DFU included in the study.

This research was developed as part of the project “Development of methodologies and algorithms aimed at supporting the imaging diagnosis of diabetic foot ulcer” from the National Program of Health Determinants, Risks and Prevention of Diseases in Vulnerable Groups, Code 1901100, National Institute of Hygiene and Epidemiology, Ministry of Public Health, Cuba, 2019-2021.

References

- [1] M. F. Ahmad, I. Khansa, K. Catignani, G. Gordillo, C. K. Sen, and M. N. Gurcan. Computerized segmentation and measurement of chronic wound images. *Computers in Biology and Medicine*, 60:74–85, 2015. ISSN 0010-4825. doi: 10.1016/j.combiomed.2015.02.015.
- [2] J. Berlanga-Acosta, J. I. Fernández-Montequín, E. López, P. López, A. Río, C. Valenzuela, Ju. Baldomero, V. Muzio, Ma. Raíces, R. Silva, B. Acevedo-Castro, and L. Herrera-Terrazas. Heberprot-P: A Novel Product for Treating Advanced Diabetic Foot Ulcer. *MEDICC Review*, 15:11–15, 2013. doi: 10.1590/S1555-79602013000100004.
- [3] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag, Berlin, Heidelberg, 2006. ISBN 0387310738. doi: 10.5555/1162264.
- [4] A. C. Bovik. *The Essential Guide to Image Processing*. Academic Press, Inc., USA, 2009. ISBN 0123744571. doi: 10.5555/1717251.
- [5] G. Bradski. The OpenCV Library. *Dr. Dobb’s Journal of Software Tools*, 2000.

- [6] P. Branco, L. Torgo, and R. P. Ribeiro. A Survey of Predictive Modelling under Imbalanced Distributions. *CoRR*, abs/1505.0, 2015. URL <http://arxiv.org/abs/1505.01658>.
- [7] B. Burgeth, A. Kleefeld, B. Naegel, N. Passat, and B. Perret. *Mathematical Morphology and Its Applications to Signal and Image Processing. 14th International Symposium, ISMM 2019 Saarbrücken, Germany, July 8–10, 2019 Proceedings*, volume 11564 of *Lecture Notes in Computer Science*. Springer International Publishing, Cham, 2019. ISBN 978-3-030-20866-0. doi: 10.1007/978-3-030-20867-7.
- [8] C. Cabal and E. González. Manual de usuario FrameHeber 03. *Centro de Ingeniería Genética y Biotecnología*, 2015.
- [9] 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*, 191:105376, 2020. doi: 10.1016/j.cmpb.2020.105376.
- [10] Ministerio de Salud Pública de Cuba. Anuario Estadístico de Salud. *Dirección Nacional de Registros Médicos y Estadísticas de Salud*, 2018.
- [11] D. M. Dhane, M. Maity, T. Mungle, C. Bar, A. Achar, M. Kolekar, and C. Chakraborty. Fuzzy spectral clustering for automated delineation of chronic wound region using digital images. *Computers in Biology and Medicine*, 89:551–560, 2017. ISSN 0010-4825. doi: 10.1016/j.compbimed.2017.04.004.
- [12] N. A. Diamantidis, D. Karlis, and E. A. Giakoumakis. Unsupervised stratification of cross-validation for accuracy estimation. *Artificial Intelligence*, 116(1):1–16, 2000. ISSN 0004-3702. doi: 10.1016/S0004-3702(99)00094-6.
- [13] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, KDD'96*, pages 226–231. AAAI Press, 1996.
- [14] C. M. A. L. García, R. Rodríguez, V. Ruiz, L. Rodríguez, L. Acosta, R. Febles, C. Pancorbo, S. Cantero, O. Vázquez, and M. Moliner. Reducción de las amputaciones con Heberprot-P en la terapia local del pie diabético. *Revista Española de Investigaciones Quirúrgicas*, XIV(1):21–26, 2011.
- [15] L. A. García-Herrera. *Diagnóstico y tratamiento del pie diabético*. Elfos Scientiae, La Habana, 2018.
- [16] G. García Gómez, D. Valdés Santiago, M. L. Bager Díaz-Romañach, W. O. Savigne Gutiérrez, A. Aldama Figueroa, C. Valdés Pérez, A. A. Martínez Hernández, and J. I. Fernández Montequín. Mejoramiento de contraste y segmentación en imágenes de úlceras del pie diabético. *Revista Cubana de Angiología y Cirugía Vasculat*, 20(3):e61, 2019.
- [17] A. Géron. *Hands-on Machine Learning with Scikit-Learning, Keras and Tensorflow. Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly, United States of America, 2nd edition, 2019. ISBN 978-1-492-03264-9.
- [18] R. González, J. Cardentey, and M. C. Casanova. Resultados del Heberprot-P en pacientes con úlceras de pie diabético. *Revista Archivo Médico de Camagüey*, 19(6):590–598, 2015.
- [19] M. Goyal, M. H. Yap, N. D. Reeves, S. Rajbhandari, and J. Spragg. Fully convolutional networks for diabetic foot ulcer segmentation. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 618–623, 2017. doi: 10.1109/SMC.2017.8122675.
- [20] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer, New York, NY, USA, 2001. doi: 10.1007/978-0-387-84858-7.

- [21] A. Heras-Tang, D. Valdés-Santiago, M. L. Baguer Díaz-Romañach, Á. León-Mecías, A. Mesejo-Chiong, and J. I. Fernández-Montequin. DFUlcer-UH: Annotation Tools v2. Registry Number: 3548-10-2019, 2019.
- [22] International Diabetes Federation. IDF Diabetes Atlas, 9th edn. Brussels, Belgium, 2019. URL <https://www.diabetesatlas.org>.
- [23] N. Japkowicz and S. Stephen. The class imbalance problem: A systematic study. *Intelligent Data Analysis*, pages 429–449, 2002. doi: 10.5555/1293951.1293954.
- [24] M. Jawahar, L. J. Anbarasi, S. G. Jasmine, and M. Narendra. Diabetic Foot Ulcer Segmentation using Color Space Models. In *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, pages 742–747. IEEE, 2020. doi: 10.1109/ICCES48766.2020.9138024.
- [25] S. Kockara, M. Mete, B. Chen, and K. Aydin. Analysis of density based and Fuzzy C-means clustering methods on lesion border extraction in dermoscopy images. *BMC Bioinformatics*, 11(SUPPL. 6), 2010. ISSN 14712105. doi: 10.1186/1471-2105-11-26.
- [26] S. Kotsiantis, D. Kanellopoulos, and P. Pintelas. Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering*, 30:25–36, 2005.
- [27] D. Meyer, F. Leisch, and K. Hornik. The support vector machine under test. *Neurocomputing*, 55(1-2): 169–186, 2003. ISSN 09252312. doi: 10.1016/S0925-2312(03)00431-4.
- [28] R. Mukherjee, D. D. Manohar, D. K. Das, A. Achar, A. Mitra, and C. Chakraborty. Automated tissue classification framework for reproducible chronic wound assessment. *BioMed Research International*, 2014, 2014. ISSN 23146141. doi: 10.1155/2014/851582.
- [29] P. L. Munoz, R. Rodríguez, and N. Montalvo. Automatic Segmentation of Diabetic foot ulcer from Mask Region-Based Convolutional Neural Networks. *Journal of Biomedical Research and Clinical Investigation*, 2(1), 2020.
- [30] K. P. Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012. ISBN 0262018020.
- [31] M. Nixon and A. S. Aguado. *Feature Extraction and Image Processing for Computer Vision*. Academic Press, Inc., USA, 3rd edition, 2012. ISBN 0123965497.
- [32] E. Pasero and C. Castagneri. Application of an automatic ulcer segmentation algorithm. In *2017 IEEE 3rd International Forum on Research and Technologies for Society and Industry (RTSI)*, pages 1–4, Modena, Italy, 2017. IEEE.
- [33] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [34] M. Ruela, C. Barata, T. Mendonça, and J. S. Marques. What Is the Role of Color in Dermoscopy Analysis? In *Pattern Recognition and Image Analysis*, pages 819–826, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg. ISBN 978-3-642-38628-2.
- [35] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9:671–675, 2012. doi: 10.1038/nmeth.2089.
- [36] S. Sergyán. Color Content-based Image Classification. In *5th Slovakian-Hungarian Joint Symposium on Applied Machine Intelligence and Informatics*, 2007.

- [37] P. Soille. *Morphological Image Analysis: Principles and Applications*. Springer-Verlag, Berlin, Heidelberg, 2nd edition, 2003. ISBN 3540429883.
- [38] C. Solomon and T. Breckon. *Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab*. Wiley-Blackwell, 2011. ISBN 9780470844724.
- [39] J. B. Sørensen. Support Vector Machines for Pixel Classification: Application in Microscopy Images. Master's thesis, Technical University of Denmark, 2014. URL <http://www.compute.dtu.dk/English.aspx>.
- [40] M. Stone. Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):111–133, 1974. doi: 10.1111/j.2517-6161.1974.tb00994.x.
- [41] The GIMP Development Team. Gimp. URL <https://www.gimp.org>.
- [42] C. Wang, X. Yan, M. Smith, K. Kochhar, M. S. Rubin, S. Warren, J. Wrobel, and H. Lee. A unified framework for automatic wound segmentation and analysis with deep convolutional neural networks. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2415–2418, 2015. doi: 10.1109/EMBC.2015.7318881.
- [43] L. Wang, P. C. Pedersen, E. Agu, D. M. Strong, and B. Tulu. Area Determination of Diabetic Foot Ulcer Images Using a Cascaded Two-Stage SVM-Based Classification. *IEEE Transactions on Biomedical Engineering*, 64(9):2098–2109, 2017. ISSN 1558-2531. doi: 10.1109/TBME.2016.2632522.
- [44] I. Wolf, M. Vetter, I. Wegner, M. Nolden, T. Bottger, M. Hastenteufel, M. Schobinger, T. Kunert, and H.-P. Meinzer. The medical imaging interaction toolkit (MITK): a toolkit facilitating the creation of interactive software by extending VTK and ITK. In Robert L. Galloway, Jr., editor, *Proceedings of SPIE - The International Society for Optical Engineering*, page 16, San Diego, CA, 2004.
- [45] R. Xu and D. Wunsch. *Clustering*. Wiley-IEEE Press, 2009. ISBN 9780470276808.
- [46] M. H. Yap, R. Hachiuma, A. Alavi, R. Brüngel, B. Cassidy, M. Goyal, H. Zhu, J. Rückert, M. Olshansky, X. Huang, H. Saito, S. Hassanpour, C. M. Friedrich, D. B. Ascher, A. Song, H. Kajita, D. Gillespie, N. D. Reeves, J. M. Pappachan, C. O'Shea, and E. Frank. Deep learning in diabetic foot ulcers detection: A comprehensive evaluation. *Computers in Biology and Medicine*, 135:104596, 2021. doi: 10.1016/j.combiomed.2021.104596.
- [47] N. I. R. Yassin, S. Omran, E. M. F. El Houby, and H. Allam. Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review. *Computer Methods and Programs in Biomedicine*, 156:25–45, 2018. ISSN 18727565. doi: 10.1016/j.cmpb.2017.12.012.
- [48] V. Yeghiazaryan and I. Voiculescu. An Overview of Current Evaluation Methods Used in Medical Image Segmentation. Technical Report RR-15-08, Department of Computer Science, Oxford, UK, 2015.
- [49] S. Zahia, M. B. Garcia-Zapirain, X. Sevillano, A. González, P. J. Kim, and A. Elmaghraby. Pressure injury image analysis with machine learning techniques: A systematic review on previous and possible future methods. *Artificial Intelligence in Medicine*, 102:101742, 2020. ISSN 0933-3657. doi: 10.1016/j.artmed.2019.101742.