Neurocomputing 164 (2015) 112–122

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Wound image evaluation with machine learning

Francisco J. Veredas [a,*], Rafael M. Luque-Baena [b], Francisco J. Martín-Santos [c], Juan C. Morilla-Herrera [c], Laura Morente [d]

[a] Departamento de Lenguajes y Ciencias de la Computación, Universidad de Málaga, Málaga 29071, Spain
[b] Departamento de Ingeniería de Sist. Informáticos y Telemát., Universidad de Extremadura, Centro Universitario de Mérida, Mérida 00680, Spain
[c] Servicio Andaluz de Salud, Junta de Andalucía, Málaga 29001, Spain
[d] Escuela Universitaria de Enfermería, Diputación de Málaga, Málaga 29071, Spain

ARTICLE INFO

Article history:
Received 7 January 2014
Received in revised form
12 December 2014
Accepted 14 December 2014
Communicated by Sandoval D. Francisco
Available online 14 March 2015

Keywords:
Machine vision
Medical imaging
Computer-aided diagnosis
Wound evaluation

ABSTRACT

A pressure ulcer is a clinical pathology of localized damage to the skin and underlying tissue caused by pressure, shear or friction. Diagnosis, care and treatment of pressure ulcers can result in extremely expensive costs for health systems. A reliable diagnosis supported by precise wound evaluation is crucial in order to succeed on the treatment decision and, in some cases, to save the patient's life. However, current clinical evaluation procedures, focused mainly on visual inspection, do not seem to be accurate enough to accomplish this important task. This paper presents a computer-vision approach based on image processing algorithms and supervised learning techniques to help detect and classify wound tissue types that play an important role in wound diagnosis. The system proposed involves the use of the k-means clustering algorithm for image segmentation and compares three different machine learning approaches—neural networks, support vector machines and random forest decision trees—to classify effectively each segmented region as the appropriate tissue type. Feature selection based on a wrapper approach with recursive feature elimination is shown to be effective in keeping the efficacy of the classifiers up and significantly reducing the number of necessary predictors. Results obtained show high performance rates from classifiers based on fitted neural networks, random forest models and support vector machines (overall accuracy on a testing set [95% CI], respectively: 81.87% [80.03%, 83.61%]; 87.37% [85.76%, 88.86%]; 88.08% [86.51%, 89.53%]), with significant differences found between the three machine learning approaches. This study seeks to provide, using standard classification algorithms, a consistent and robust methodological framework as a basis for the development of reliable computational systems to support ulcer diagnosis.

1. Introduction

The European Pressure Ulcer Advisory Panel (EPUAP) defines a pressure ulcer (PU) as an area of localized damage to the skin and underlying tissue caused by pressure, shear, friction or a combination of these factors [1–3]. Prevention, care and treatment of PU pathology represent high costs for health services [4] and have important consequences for the health of the affected population, especially for those populations at risk such as elderly people. Recent studies [5] have also shown how mortality rates associated with this pathology have increased in the last few years.

Accurate diagnosis of the wounds is critical in order to proceed with the right diagnosis and appropriate treatment, which in some cases can require surgery. This crucial evaluation is carried out by clinicians using standardized scales or indexes consisting mainly of a visual inspection of the ulcer. However, this technique has been proven to be an inaccurate way to deal with diagnosis of this sort of wound [6,7].

One of the most challenging factors to cope with when working with PU images lies in the very heterogeneous colorations they present, which are related with a patient's skin color and other several anomalies that may be observed in the images, such as erythemas and skin striation. Moreover, boundaries between different tissue regions are usually extremely irregular and vague, which increases the complexity of the tissue segmentation process. Image processing and computational intelligence techniques have been applied in several current studies to address different aspects of this particular problem of wound diagnosis. One of these aspects involves the partial problem of wound area identification, which has been tackled with different

* Corresponding author. Tel.: +34 952137155; fax: +34 952131397.
E-mail addresses: fvn@uma.es (F.J. Veredas), rmluque@unex.es (R.M. Luque-Baena), fja.martin.sspa@juntadeandalucia.es (F.J. Martín-Santos), jmorilla29@gmail.com (J.C. Morilla-Herrera), lmorente@malaga.es (L. Morente).

http://dx.doi.org/10.1016/j.neucom.2014.12.091
0925-2312/

techniques such as contour detection with histogram segmentation, active contours modelling, region growing, clustering approaches or skin texture models [8–10]. Unlike these proposals, other approaches focus on detecting the different tissues existing in the wound, by using diverse segmentation methods—such as histogram thresholding, watersheds, mean-shift smoothing, region growing, classification or graphs—sometimes combined with machine learning (ML) strategies [11–14].

A few studies can be found in the specific literature on PU tissue recognition. Pioneering work by Berriss [15] addressed the design of an image processing tool based on simple histogram thresholding to separate granulation and slough tissue regions on wound images. For their part, Perez et al. [16] designed a semi-automatic tool that started with a previous calibrating and filtering stage and required user intervention to select sample regions in wound-bed and surrounding areas. With similar objectives, Galushka et al. [17] used a case-based tissue classification approach to classify square regions of interest from a grid-split structure on wound unsegmented images. In their experiments, color and texture patterns from square regions were firstly classified in three tissue types, i.e., granulation, slough and necrosis, and finally categorized into six tissue categories which include epithelial, tendon and haematoma. Lourega et al. [18] presented a hybrid approach, which they named *MeSegHi*, to image segmentation with potential application to tissue recognition tasks such as detection of carcinoma cells. On the other hand, ML strategies have been proposed by several authors as approaches to wound tissue segmentation and recognition. Pioneering work was done by Belem in his Ph.D. thesis [19], where different techniques such as logistic regression, support vector machines (SVM) and feed-forward neural networks (NN) were compared with a manual approach by clinicians with different skills. His results showed SVMs as the best performing classification engines, which gave more consistent judgements than most clinical practitioners. In a similar research area, Serrano et al. [20] designed a computer assisted diagnosis tool that used a NN to classify burn wounds by detecting tissues that determine the depth of the injury. In another ML based study, Kosmopoulos and Tzevelekou [11] presented some exploratory results from an approach which used graph-based image segmentation and SVMs for PU diagnosis. In [21] Wannous et al. used three different methods for region segmentation as well as a SVM to classify tissues in images of pressure sores and ulcers, but their approach required a preliminary manual selection of the area of interest in the wound; moreover, their results show high classification error scores, which are even greater than 50% for some critical tissues such as necrosis (see [21], table IV). More recently, Wannous's team has published [14] an innovative approach that combines 3-D wound surface measurements with tissue classification based on a SVM model to achieve enhanced wound healing assessments. Finally, in [12] we presented a mean-shift procedure along with a region-growing strategy for effective region segmentation, combined with a complex hybrid approach based on acascade of single-class classifiers for wound tissue recognition. Although we obtained high efficacy rates, our work presented some significant limitations, since the classifiers were based on an elaborate architecture consisting of NNs and Bayesian classifiers, which were combined to recognize patterns of color and texture features. Additionally, a complex set of heuristic techniques had to be applied to the classifiers' outputs to get the final classification results.

This study proposes a settled and reliable methodology to support pressure ulcer diagnosis and to serve as a starting point for more complex models. Although previous studies above have highlighted the adequacy of diverse ML approaches for classifying PU tissues from wound segmented images, most of these works lack an exhaustive analysis that addresses robust model fitting. In our study, parameter fitting is supported by *k*-fold cross-validation analysis and a rigorous statistical methodology has been followed (using statistical tests, p-values and confidence intervals) to measure the degree of significance of the differences in efficacy obtained with different ML models and parameters. Therefore, this study combines image analysis techniques with the most widely used classification methods, which are rather settled in the scientific community. Thus the efficacy of three different ML approaches for PU tissue recognition is exhaustively analyzed: SVM, feed-forward NNs and random forest (RF). Moreover, a new highly effective image processing procedure is also proposed and devised to extract a richer descriptor set than those obtained in other similar studies [11,21,13,12], since not only texture and color features are taken into account, but also morphological and topological characteristics are considered. Future development of clinical applications for PU diagnosis and treatment could benefit from both our approach and analysis results. In this vein, PU diagnosis tools based on efficient image segmentation and effective tissue recognition could be further integrated in portable devices, such as smart-phones or tablets, and give support to clinical decision-making as they could provide the clinicians with real-time diagnosis capabilities on the spot, which could contribute to the efficacy of treatments and care interventions.

This paper is structured in six sections, including the introduction given in this Section 1. In Section 2, the methodology followed in this approach is described. Experimental results achieved by different ML approaches are shown in Section 3 (for wound-bed tissue recognition) and Section 4 (for peri-ulcer tissue recognition) and discussed in Section 5. Finally, conclusions and further works are discussed in Section 6.

2. Methodology

Clinicians took color photographs of PUs of patients with home-care assistance. Sacrum and hip PUs were photographed under non-controlled illumination conditions by using a Sony Cybershot® W30 digital camera. The images were acquired with flash-light to get well-illuminated scenes, and at a distance of approximately 30–40 cm from the wound plane. Macro-photography focusing was used to ensure well-focused pictures within these short distances. A group of clinical experts selected a total of 113 photographs that were considered to be an appropriate data set for analysis because of the presence of all significant tissue types in PU evaluation.

2.1. Segmentation process

After an initial median filtering for noise reduction, a preprocessing procedure is applied to each photograph in order to previously detect and extract those spurious regions consisting of flash-light reflections, blood stains, shadows, clinical materials, clinicians' fingers, normalization markers, etc. (see Fig. 1). To that end, an appropriate set of color space transformations and standard deviation filters—which depend on the sort of invalid region needed to be detected—are applied to the images as a previous step to image segmentation.

The objective of the segmentation module is to divide the image into groups of pixels with similar characteristics, in order to conduct the subsequent classification of the obtained regions. This segmentation process is arranged in three main sequential stages (scheduled in Fig. 1), which are based on the specific 'center-surround topology' of the ulcer images, with a centered wound-bed that consists of a variable proportion of granulation, slough and necrotic tissues, which is immediately surrounded by epithelial healing tissue (that could present different conditions, such as

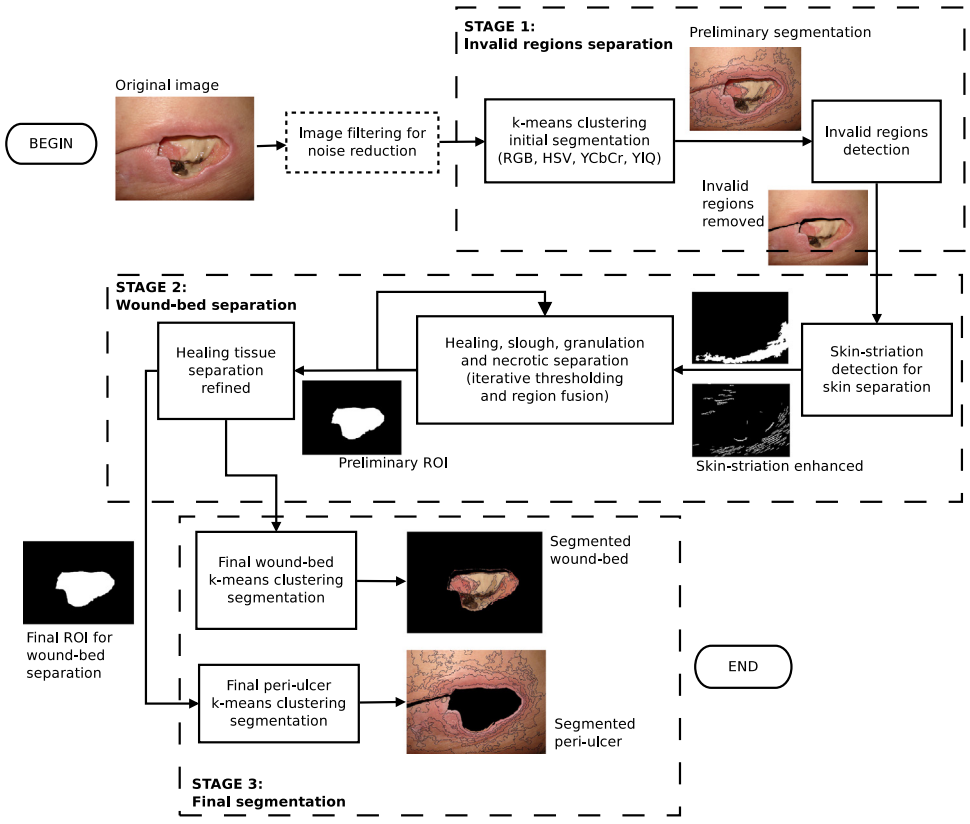

Fig. 1. The image segmentation process, arranged in three main stages: (1) preprocessing for invalid regions detection, (2) separating the wound-bed area from the peri-ulcer area, and (3) a specific segmentation process is independently applied to each of the two main areas obtained.

intact and healthy, macerated, lacerated, eczematous, etc.) and a final periphery mostly consisting of healthy skin.

In the first stage, a simple k-means algorithm has been used (number of groups $k=12$) to get a preliminary segmentation, by clustering the pixels in the image into separate groups consisted of adjacent pixels with similar color features. It is important to note that this process is performed without the intervention of any human operator. This step makes it possible to work at region-level instead of doing it at pixel-level. Components from four different color spaces—RGB, HSV, YCbCr and YIQ—have been used for this clustering-based pre-segmentation, in particular S (Saturation), Y (luminance), Q (crominance) and H (hue) components from those spaces. The number of $k=12$ groups for the k-means algorithm has been chosen as the minimum number that maximizes the number of resulting segmented regions in the images that include a single tissue type (i.e., homogeneous regions).

In the following segmentation stage, some image processing steps have been designed to separate the wound-bed area from the peri-ulcer area (healing and skin regions). In the final stage, a specific segmentation process is independently applied to each of the two main areas obtained, wound-bed and peri-ulcer.

2.1.1. Wound-bed area estimation stage

At this point, a subsequent iterative processing stage has been carried out on the regions resulting from the k-means clustering, with the main objective of locating those regions consisting of healing tissue, slough, granulation and necrotic areas. This tissue detection supplies us with information about the wound-bed shape, since healing tissue is mainly arranged along wound-bed edges and delimits the wound-bed area consisted of granulation, slough and necrotic regions. This process starts with the definition of a new set of color channels obtained by linear and non-linear transformation of the 12 components from the four color spaces above. Thresholding strategies based on the distribution of these color channels in each different tissue type allow us to obtain a ROI (Region of Interest) mask corresponding to an *estimated wound-bed* area. This preliminary separation between wound-bed and peri-ulcer allows us to tackle wound-bed tissue recognition in a more efficient manner.

2.1.2. Final segmentation stage

Once the estimated wound-bed area is fully separated from the peri-ulcer area, pixels inside and outside the wound-bed are processed in two separated pools, respectively, to get the final segmented image. For pixels in each one of these pools, the k-means ($k=12$) clustering algorithm is applied again, but using different combinations of color spaces, channels and transformed channels to enhance the separation between clusters in the multi-dimensional feature spaces of the pixels in both areas. The k-means algorithm separates the pixels in each area (estimated wound-bed and peri-ulcer) into 12 different groups. Even though there could be only four different tissue types inside the *estimated wound-bed* (granulation, slough, necrotic and healing regions) as well as two different tissue classes in the *estimated peri-ulcer* (skin and healing regions), that higher number of clusters than tissue types is used in order to ensure that one and only one tissue type can be found in every region generated by the k-means clustering. At the same time, although the number of resulting regions can be high, region sizes are large enough to extract significant descriptors to be identified as the right tissue type by the subsequent classification process.

A total of 16,449 regions are finally obtained from the 113 images after segmentation, with an average of 145.4 regions per image (standard deviation 52.3 regions), where we got an average of 81.36 regions per image in the wound-bed area (standard deviation 55.3 regions) and an average of 64.2 regions per image in the peri-ulcer area (standard deviation 27.67 regions).

Once the images have been segmented, a set of patterns consisting of topological, morphological, color and texture descriptors from each one of the segmented regions is extracted. A brief enumeration of these features can be summarized as follows:

1. Color components:
   (a) Components from different color spaces:
      (i) RGB
      (ii) YIQ
      (iii) YCbCr
      (iv) HSV
   (b) Linear and non-linear combinations of the different components from the color spaces above, to get transformed color features.
   (c) Image-level descriptors:
      (i) Percentiles of the intensity levels of the transformed color components.
      (ii) Estimated threshold for the Cr channel (from the YCbCr color space).
   (d) Region-level descriptors:
      (i) Mean intensity levels of the transformed color components.
      (ii) Standard deviations of the transformed color components.
      (iii) Mean of the standard deviation filter applied to the region, for all the transformed color components.
2. Topological and morphological components:
   (a) Image-level descriptors:
      (i) Estimation of the radius of the wound-bed in pixels units.
   (b) Region-level descriptors:
      (i) Percentiles of the distribution of the distance of the pixels in the region to the wound-bed border.
      (ii) Percentiles of the distribution of the distance of the pixels in the region to the wound-bed centre.
      (iii) Ratio of the region perimeter belonging to the wound-bed border.
      (iv) Ratio of the region perimeter belonging to the image border.
      (v) Ratio of the region area belonging to the preliminary ROI (see Fig. 1).
      (vi) Ratio of the region area belonging to the first wound-bed area estimation (see Fig. 1).
      (vii) Ratio of the region area belonging to the final wound-bed area estimation (see Fig. 1).

Features 2a, 2(b)i and 2(b)ii were computed after image scaling using the $1\,\text{cm} \times 1\,\text{cm}$ normalization markers included in the images (Fig. 2). Those descriptors included in the 2. *Topological and morphological components*' category provide 'contextual' information associated to each PU image as a whole. Because of the fact that characteristics from those regions in 113 different PU images are included in the pattern sets, each feature pattern is provided with some global information that 'links' its corresponding segmented region with the characteristics of the original PU image that region comes from, which is expected to improve the accuracy of the resulting classifiers. In total, 102 and 104 features are used as inputs to the classifiers for wound-bed and peri-ulcer tissue recognition, respectively (characteristics 2(b)iv and 2(b)v are omitted in wound-bed region classification). To remove distributional skewness, BoxCox transformations [22] have been applied to the input variables, which gave mean lambda estimates .5549 and .3164 from patterns in the wound-bed and the peri-ulcer, respectively. Finally, the classifiers are also supplied with the output value representing the *tissue-class* during the supervised learning phase. A group of five expert clinicians from the *Málaga Province Health Services* reached a consensus (via majority opinion) to label each region from the set of 113 segmented images, by assigning one of the four possible tissue types—*granulation*, *slough*, *necrotic*, *healing*—to each region in the estimated wound-bed area; peri-ulcer regions were labelled as *healing* or *skin* tissues by the experts (Fig. 2). These labelled images become the *gold-standards* and give the expected outputs from the classifiers during its supervised training phase. The tissue class shows an unbalanced distribution in our data, since the prevalence of the different tissue types is quite heterogeneous (see Fig. 3).

3. Wound-bed tissue recognition

Wound evaluation is mainly based on detection and measurement of wound-bed tissues, which are the most significant tissues for PU diagnosis [23]. The image processing and segmentation stages explained in Section 2.1 allow the separation between estimated wound-bed and peri-ulcer areas, so that wound-bed tissues can be identified in a phase independent from that aimed at identifying the peri-ulcer tissues. Three different ML approaches are comparatively used to tackle the multi-class classification problem of wound-bed tissue recognition: SVM [24], NN [25] and RF [26]. The pattern set is divided into two independent sets, 80% patterns for training and 20% for testing, by means of a stratified random sampling that tries to preserve the balance of the class distribution within the splits (see Table 1). Before training the SVMs and the NNs, all patterns in the testing and training sets are standardized by subtracting the mean and dividing by the standard deviation of those patterns in the training set. For each particular ML model fitting, a grid of parameters is created and the model is trained on slightly different data for each candidate combination of tuning parameters. Across each data set, the performance (accuracy and Cohen's *Kappa* coefficient) of held-out samples for 10-fold cross-validation is calculated and the mean and standard deviation summarized for each combination. The combination with the optimal re-sampling accuracy is then chosen as the final model and the entire training set is used to fit a final model. The *caret* R package [27,28] has been used for model fitting with SVM (package *kernlab* [29]), NN (package *RSNNS* [30]) and RF (package *randomForest* [31]), with five repetitions for each fold in the cross-validation procedure (giving a total 100 re-samplings for each combination of parameters from each ML model). Fig. 4 shows the averaged accuracy for each combination of the tuning parameters in each ML model.

For SVMs, a Gaussian radial basis function (RBF) kernel $k(x,x') = e^{\sigma\|x-x'\|^2}$ is used (being k a function that calculates the inner product $\langle \Phi(x), \Phi(x') \rangle$ of two vectors x, x' for a given projection $\Phi : X \to H$). For this multiclass-classification problem, with $n=4$ classes corresponding to 4 tissue types, we use a 'one-against-one'-approach, in which $n(n-1)/2 = 6$ binary classifiers are trained; the appropriate class is found by a voting scheme. The problem of model selection (parameter tuning) is partially addressed by an empirical observation for the Gaussian RBF kernel, where the optimal values of the hyper-parameter σ are known to lie in between the .1 and .9 quantile of the $\|x-x'\|$ statistics [32,29]. A sample of the training set is used to estimate these quantiles [29], where any value within the quantile interval results in good performance. This way, σ parameter is automatically estimated. Additionally, the optimal hyper-parameter *cost*, that

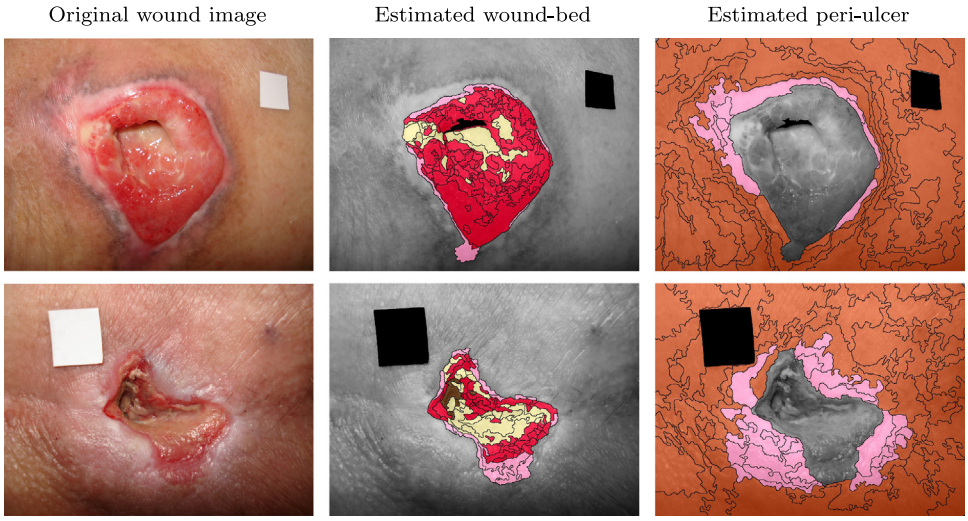

Fig. 2. Examples of two different segmented images, showing the original images (left), the segmented regions in the estimated wound-bed area (centre) and the estimated peri-ulcer (right). The segmented regions have been colored with pseudo-colors representing the different tissue types labelled by the expert clinicians: skin (orange), healing tissue (pink), granulation (red), slough (yellow) and necrosis (brown). Regions corresponding to non-tissue areas, such as shadows or normalization markers, have been given black color. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

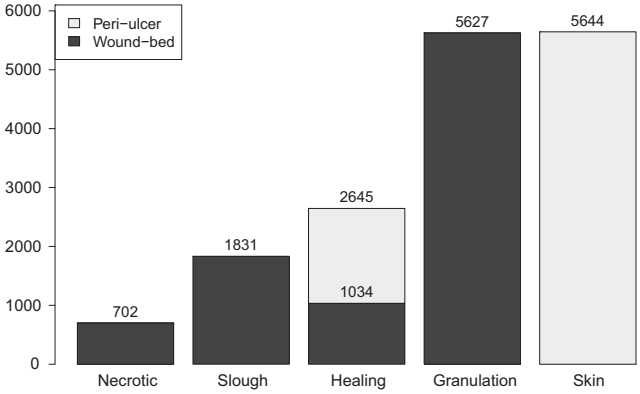

Fig. 3. Distribution of the different tissue types in the 16,449 segmented regions extracted from the 113 PU images, distinguishing between tissues in the estimated wound-bed area of the images (total 9194 regions) and in the peri-ulcer (total 7255 regions). The regions in the images were manually labelled by clinical experts.

Table 1
Distribution of patterns of the different wound-bed tissue types in the training (80%) and testing sets (20%) for ML.

Dataset	Necrotic	Slough	Healing	Granulation	Total
Training	562	1465	828	4502	7357
Testing	140	366	206	1125	1837
Total	702	1831	1034	5627	9194

represents the cost of constraints violation and stands for the 'C'-constant of the regularization term in the Lagrange formulation, is tuned as the one of 10 incremental values in $\{2^i\}_{i=-2}^{7}$ that optimizes the performance of the SVM classifier. Best model's tuning parameters are $\sigma = .0078$, $C = 8$ (Fig. 4A).

Single-hidden-layer feed-forward multilayer perceptron NNs are also constructed and trained with different combinations of parameters to search for the best performance rates in the classification of patterns from the estimated wound-bed areas. Logistic activation function is used and training is performed by error back-propagation with weight decay. The learning rate is set to .2. The network size (i.e., number of *hidden units*) and *weight decay* are the parameters being tuned, selecting the combination of values that gives the highest performance of the NN for *#hidden units* $= \{5, 6, \ldots, 20\}$ and *weight decay* $= \{0, 10^{-6}, 10^{-5}, 10^{-4}\}$. Fig. 4B shows mean accuracy versus number of hidden units for those different weight decay values (displayed with different line styles). Weights are randomly initialized, and maximum number of epochs is fixed to 100 [33]. The optimal parameter values are: *#hidden units* $= 30$, *weight decay* $= 10^{-5}$ (Fig. 4B).

RFs are ensemble learning methods for classification that function by constructing a large pool of decision trees during the training phase, then giving an output that is the mode of the classes given by the individual trees in the pool. The method combines Breiman's 'bagging' idea and the random selection of features (i.e. predictor-set split) in order to construct a collection of decision trees with controlled variation [26]. For RF model-fitting in our experiments for tissue classification, the only tuning parameter is the *number of variables (predictors) randomly sampled* as candidates at each split (Fig. 4C), which is tuned as the one that gives the best performance among the 10 different values in $\{1, 2, \ldots, \lfloor\sqrt{S}\rfloor = 10\}$, being $S = 102$ the number of predictors. The RF parameter *number of trees to grow* has been fixed to 500 trees to ensure that every input pattern gets predicted at least a few times [31]. The optimal *number of predictors randomly sampled* is 7 (Fig. 4C).

3.1. Feature selection

Many feature selection procedures are based on the cooperation of variable importance for ranking and model estimation to generate, evaluate and compare a family of models. Usually, three types of feature selection methods are distinguished [34]: *filter*, for which the score of variable importance does not depend on a given model design method; *wrapper*, which includes the prediction performance in the score calculation; and *embedded*, which combines more closely variable selection and model estimation.

Many feature selection routines use a wrapper approach [35] to find appropriate variables such that an algorithm that searches the feature space repeatedly fits the model with different predictor sets. The best predictor set is determined by some measure of performance (i.e., R^2 and accuracy). An example of one search routine is backwards selection (a.k.a. *Recursive Feature Elimination*, RFE). First, the algorithm tunes and fits the model to all predictors.

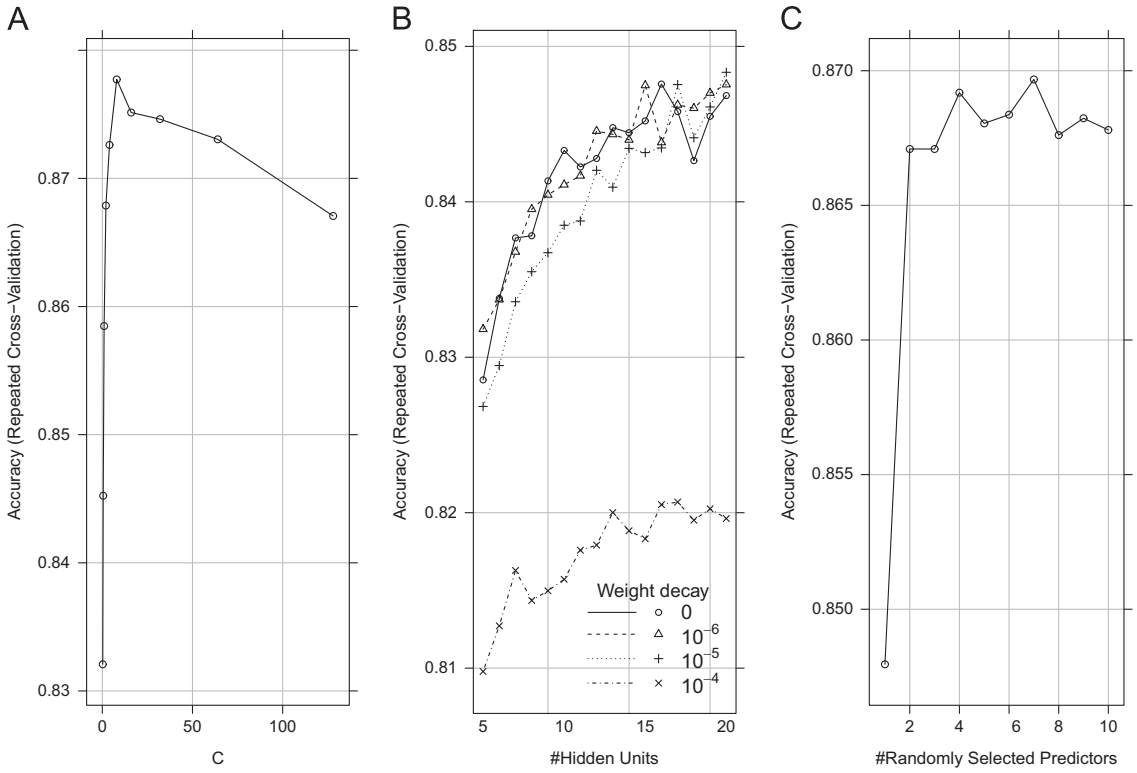

Fig. 4. Plots of the resampling profiles to examine the relationship between the estimates of performance (mean accuracy) and the tuning parameters for three different ML models used in this study: (A) SVM, (B) NN and (C) RF. Notice that axis of ordinates is in different scales.

Table 2
Comparative results of ML models (SVM, NN, RF) fitted on the wound-bed training set with feature selection (S is the number of selected features). Number of re-samples: 50 (10-fold cross-validation with 5 repetitions).

Model	Best-fitting parameters	Accuracy		Kappa	
		Mean (sd)	95% CI (t-Test)	Mean (sd)	95% CI (t-Test)
SVM	$\sigma = .008$ *cost*=8 S=102	.8777 (.0116)	[.8744, .8810]	.7803 (.0210)	[.7744, .7863]
NN	*#Hidden units* = 20 S=37	.8520 (.0133)	[.8482, .8558]	.7303 (.0254)	[.7231, .7375]
RF	*#Randomly selected predictors*=8 S=73	.8725 (.0120)	[.8691, .8759]	.7675 (.0223)	[.7612, .7739]

Each predictor is ranked using its importance to the model. Let S be a sequence of ordered numbers that are candidate values for the number of predictors to retain ($S_1 > S_2, \ldots$). At each iteration of feature selection, the S_i top ranked predictors are retained, the model is re-tuned and re-fit and performance is assessed. The value of S_i with the best performance is determined and the top S_i predictors are used to fit the final model.

The quantification of the variable importance is a crucial issue not only for ranking the variables before a stepwise estimation model but also to interpret data and understand underlying phenomena in many applied problems. The following method in used by RF for estimating the contribution of each variable to the model: for each decision tree, the prediction accuracy on the out-of-bag portion of the data is recorded. Then the same is done after permuting each predictor variable. The difference between the two accuracies is then averaged over all trees, and normalized by the standard error. For NN, the method used here is based on [36], which uses combinations of the absolute values of the weights. For our multi-class classification problem, the class-specific importances are considered to be the same. Finally, for SVM the importance of each variable is individually evaluated using a "filter" approach. ROC curve analysis is conducted on each predictor. For our multi-class classification task, this problem is decomposed into all pair-wise problems and the area under the ROC curve (AUC) is calculated for each class pair. For a specific class, the maximum area under the curve across the corresponding pair-wise AUCs is used as the variable importance measure [37].

Since feature selection is part of the model building process, resampling methods should factor in the variability caused by feature selection when calculating performance. Improper use of resampling to measure performance could result in models that perform poorly on new samples [38,39]. To get performance estimates that incorporate the variation due to feature selection, it is encapsulated inside an outer layer of resampling iterations (10-fold cross-validation with 5 repetitions). Since the aim of this study is to compare the efficacy of three different ML models for PU diagnosis (i.e., SVM, RF and NN), this complete RFE scheme described must be accomplished on each one of these proposed models [37], which becomes computationally burdensome. To tackle this computational limitation and in view of the results obtained from the 102-feature approach for model fitting

and training, the tune parameter grids for the different models have been shrunk with respect to the original ones (see Section 3). Thus, for SVM, only three *cost* values $\{6, 7, 8\}$ have been explored, while σ is left constant to .008; for NN, three different hidden-layer sizes have been tested for model fitting, with *#hidden units* $= \{18, 19, 20\}$ (training is performed by error back-propagation without weight decay); finally, for RF, the *number of variables (predictors) randomly sampled* is set to $\lfloor\sqrt{S_i}\rfloor$ for each subset size S_i. Algorithm 1 shows the complete procedure of feature selection and model fitting.

```
1   begin
2     Data : The training set of patterns with S = 102 features
3     Result : A list of selected features
4     foreach Resampling iteration do
5       Partition patterns into training and hold−back set
6       Tune/Train the model on the training patterns using all features
7       Predict the held−back patterns
8       Calculate variable importance and rankings
9       foreach subset size S_i, i = 1…S do
10        Select the S_i most important variables
11        Tune/Trainthe model on the training set using S_i features
12        Predict the held−back patterns
13    Calculate the performance over S_i using the held−back patterns
14    Determine the appropriate number of features
      Determine the final list of selected features
      Fit final models using the optimal S_i on the original training set
```

Algorithm 1. Recursive Feature Elimination (RFE) with resampling.

Table 2 shows the comparative results from the three tuned ML approaches, obtained via 10-fold cross-validation (with 5 repetitions) on patterns in the training set. Since we did not find any significant difference between the 102-feature and the *S*-feature (feature selection) approach for each ML model, only results *with* feature selection are shown. For each ML model, those parameters giving the best accuracy results are shown in the table, in the column 'Best-fitting parameters'. The column labelled 'Accuracy' is the overall agreement rate averaged over cross-validation iterations. The agreement standard deviation is also calculated from the cross-validation results. The column 'Kappa' is the Cohen's (unweighted) Kappa statistic averaged across the resampling results. On the other hand, in Table 3 the mean differences between the 'paired' models are shown. From these results in Tables 2 and 3 we can observe that the three ML approaches give high accuracy results $>85\%$, with SVM classifiers achieving the best overall performance rates, with significant differences found when compared with NNs and RFs. However, while RFs give mean accuracy that is only $<1\%$ lower than that given by the SVMs, NNs achieve a performance that is $\sim 3\%$ smaller than those achieved by SVMs. Notice that, although NNs give the lowest accuracy rate, they do it by using the minimum set of selected characteristics ($S=37$), while SVMs use the complete set of 102 characteristics instead (i.e., RFE with SVM wrappers resulted in no feature selection). It is should be noted that statistically significant differences in the detection of the tissue type do not involve clinically significant differences, because the diagnosis of an ulcer is based on the analysis of a multitude of tissue regions.

Table 4 shows the associated statistics [40,41] resulting from the cross-tabulation of observed and predicted classes of patterns in the wound-bed testing set. For each tissue type, the accuracy, balanced accuracy, sensitivity, specificity, positive predictive value and negative predictive value are calculated. The detection rate (the rate of true positives) and the detection prevalence (the prevalence of predicted classes) are also shown in the table. For the sake of clarity, given a 2×2 table for each tissue class,

	Reference	
Predicted	Class	No Class
Class	TP	FP
No Class	FN	TN,

where TP, *FP*, *TN* and *FN* stand for *true positive*, *false positive*, *true negative* and *false negative*, respectively, the formulas used in Table 4 are the following:

- $Accuracy = (TP+TN)/(TP+TN+FP+FN)$
- $Prevalence = (TP+FN)/(TP+TN+FP+FN)$
- $Sensitivity = TP/(TP+FN)$
- $Specificity = TN/(TN+FP)$
- $Positive\ predictive\ value\ (PPV) = (sensitivity \times prevalence)/((sensitivity \times prevalence)+((1-specificity)\times(1-prevalence)))$
- $Negative\ predictive\ value\ (NPV) = (specificity \times (1-prevalence))/(((1-sensitivity)\times prevalence)+(specificity\times(1-prevalence)))$
- $Detection\ rate = TP/(TP+TN+FP+FN)$
- $Detection\ prevalence = (TP+FP)/(TP+TN+FP+FN)$
- $Balanced\ accuracy\ (sensitivity+specificity)/2$

As observed in Table 4 and the table's caption, high overall accuracy rates are obtained again with the fitted SVM, NN and RF models when applied on the testing set, with the highest overall efficacy values given by the former (88.08%). However, while the highest performance rates are again given by the SVM and RF models, there is not a clear 'winner'. On the contrary, best results depend on the performance measurement being observed as well as on the wound tissue type considered. Thus, while the RF gets the highest sensitivity rates for granulation tissue, the SVM gives the best sensitivity outcomes for necrotic, slough and healing classes. On the other hand, specificity best outcomes are also heterogeneously distributed among tissue types and ML models, with the SVM and the RF giving the highest values for the different classes. Similar scenarios can be observed in the distribution of results from the rest of the statistics shown in Table 4.

4. Peri-ulcer tissue recognition

Although wound evaluation is mainly based on detection and measurement of wound-bed tissues [23], the image processing and segmentation methods used in this study—and explained in details in Section 2.1—allow the separation between estimated wound-bed and peri-ulcer areas, so that we can deal with peri-ulcer tissue recognition in a phase independent from that aimed at identifying the wound-bed tissues—addressed in Section 3. The peri-ulcer pattern set consists of 7255 samples of two different classes: 5644 are skin samples, while 1611 are labelled as healing regions (see also Fig. 3). The peri-ulcer training set consists of 80% of these patterns: 5805 (skin: 4516; healing: 1289); the peri-ulcer testing set groups together the remaining 1450 samples (skin: 1128; healing: 322).

Following an identical strategy to that accomplished with wound-bed tissues (see Section 3 and 3.1), model fitting (SVM, NN and RF) and feature selection have been carried out on patterns from the peri-ulcer area of the 113 wound images. Unlike wound-bed multi-tissue identification, peri-ulcer tissue recognition is tackled as a simpler binary classification task. For model tuning with the complete 104-feature dataset, the same parameter grids as those given for 102-feature wound-bed classification have been used (see Section 3), giving the following optimal parameter values: SVM: $\sigma = .007$, $C = 16$; NN: *#hidden units* $= 14$, *weight decay* $= 10^{-5}$; RF: *number of variables randomly sampled* $= 9$. However, for model tuning with feature selection, only three *cost* values $\{15, 16, 17\}$ have been explored for SVM tuning, with constant $\sigma = .007$; for NN, three different hidden-layer sizes have been tested for model fitting, with *#hidden units* $= \{13, 14, 15\}$ (training is performed by error back-propagation without weight decay); finally, for RF, the *number of variables randomly sampled* is set to $\lfloor \sqrt{S_i} \rfloor$ for each subset size S_i.

Tables 5 and 6 show the comparative results from model fitting on the testing set and the statistics from the cross-tabulation of observed and predicted classes in the testing set, respectively. Once again, since we did not find any significant difference between the 104-feature and the *S*-feature (feature selection) approach for each model, only results *with* feature selection are shown. As observed in these tables, given the smaller complexity of this binary classification problem, the three ML models analyzed obtain high overall performance rates (with mean accuracy $> 91\%$

Table 3
Differences of means between the three fitted ML models: SVM, NN, RF. Number of resamples: 50 (10-fold cross-validation with 5 repetitions). *p* Value $< .05$ in all cases (*p* value adjustment: Bonferroni).

Models	Accuracy		Kappa	
	Mean of Diff.	98.33% CI	Mean of Diff.	98.33% CI
SVM - NN	.0257	[.0187, .0328]	.0500	[.0369, .0631]
SVM - RF	.0052	[−.0012, .0117]	.0128	[.0011, .0245]
NN - RF	−.0205	[−.0244, −.0166]	−.0372	[−.0442, −.0302]

Table 6
Statistics from the cross-tabulation of observed and predicted classes (confusion matrices) for the three different ML methods (SVM, NN and RF) used for classification of peri-ulcer patterns (of two classes, healing and skin) from a same testing set consisting of 1450 samples, with and feature selection. 'Positive' class: healing (.2221 prevalence). Overall accuracy [95% CI]: SVM: .9214 [.9063, .9347]; NN: .8993 [.8827, .9143]; RF: .9366 [.9228, .9485]. The highest values of each statistic for each tissue type is shown in boldface.

ML method	SVM	NN	RF
Sensitivity	.7640	.7484	**.8230**
Specificity	.9663	.9424	**.9690**
Pos Pred value	.8662	.7876	**.8833**
Neg Pred value	.9348	.9292	**.9504**
Detection rate	.1697	.1662	**.1828**
Detection prevalence	.1959	**.2110**	.2069
Balanced accuracy	.8651	.8454	**.8960**

Table 4
Statistics from the cross-tabulation of observed and predicted classes (confusion matrices) for the three different ML methods (SVM, NN and RF) used for classification of wound-bed patterns from a same testing set consisting of 1837 samples. Overall accuracy [95% CI]: SVM: .8808 [.8651, .8953]; NN: .8187 [.8003, .8361]; RF: .8737 [.8576, .8886]. The highest values of each statistics for each tissue type are shown in boldface.

Tissue (prevalence) ML method	Necrotic (.0762)			Slough (.1992)			Healing (.1121)			Granulation (.6124)		
	SVM	NN	RF	SVM	NN	RF	SVM	NN	RF	SVM	NN	RF
Sensitivity	**.8643**	.8500	.8357	**.7951**	.7404	.7896	**.6505**	.4175	.5728	.9529	.9138	**.9609**
Specificity	.9894	.9853	**.9912**	**.9592**	.9381	.9558	.9730	.9651	**.9792**	**.8638**	.7753	.8343
Pos Pred value	.8705	.8264	**.8864**	**.8291**	.7486	.8164	.7528	.6014	**.7763**	**.9170**	.8653	.9016
Neg Pred value	**.9888**	.9876	.9865	**.9495**	.9356	.9481	**.9566**	.9292	.9478	.9207	.8505	**.9310**
Detection rate	**.0659**	.0648	.0637	**.1584**	.1475	.1573	**.0730**	.0468	.0642	.5836	.5596	**.5885**
Detection prevalence	.0757	**.0784**	.0719	.1911	**.1971**	.1927	**.0969**	.0778	.0827	.6364	.6467	**.6527**
Balanced accuracy	**.9268**	.9176	.9134	**.8771**	.8393	.8727	**.8118**	.6913	.7760	**.9083**	.8445	.8976

Table 5
Comparative results of ML models (SVM, NN, RF) fitted on the peri-ulcer training set with feature selection (*S* is the number of selected features). Number of re-samples: 50 (10-fold cross-validation with 5 repetitions).

Model	Best-fitting parameters	Accuracy		Kappa	
		Mean (sd)	95% CI (t-Test)	Mean (sd)	95% CI (t-Test)
SVM	$\sigma = .007$ *cost* = 15 *S* = 91	.9290 (.0092)	[.9264, .9316]	.7912 (.0267)	[.7836, .7987]
NN	*#Hidden units* = 13 *S* = 37	.9148 (.0119)	[.9114, .9181]	.7494 (.0340)	[.7397, .7590]
RF	*#Randomly selected predictors* = 8 *S* = 67	.9318 (.0092)	[.9292, .9344]	.7986 (.0276)	[.7907, .8064]

in all cases) when identifying patterns from both the training and testing peri-ulcer sets. Although high variability can be found again in the distribution of outcomes from the different performance indicators (sensitivity, specificity, etc.) among the tissue classes and ML methods (see Table 6), the resulting scores are in general higher than those obtained on the wound-bed pattern testing set. The highest accuracy rates are again given by the RF and the SVM, with no significant differences found between these two models ($p \approx .09$). However, as observed in Table 6, the RF gives the highest sensitivity and specificity rates (hence the highest balanced accuracy rate). Finally, although the NN gives the lowest accuracy rate (with significant differences found between the NN and the other two models, $p < .01$), it does it again by using the minimum set of selected characteristics ($S=37$).

5. Discussion

Our segmentation strategy presented in Section 2.1 has some important advantages with respect to other wound image segmentation approaches used in similar studies [14,12]. On one hand, image resolution does not need to be reduced (as it has to in [12], for example) to have acceptable processing times. Thus, the complete segmentation and feature extraction phases in our study gave an average of 69.5 s (standard deviation 12.02 s) per image (with resolution 1622 × 1224 pixels) on a 'standard' modern PC (Intel Core i7®, 2.93 GHz, 6Gb RAM, Windows 7® 64 bits) in Matlab® release 2013. On the other hand, in spite of being an apparently laborious process (Fig. 1), our segmentation approach is particularly adapted to the peculiarities of wound images, and includes fully-automatic processing steps for the detection of invalid regions (such as background and clinical materials) and the effective estimation of wound-bed and peri-ulcer areas. The latter allows us to have two independent image segmentation stages, each one fitted to just one PU area, as well as to extract an enriched feature set consisting of topological and morphological descriptors that join color and texture characteristics to contribute to a higher efficacy of the classifiers. As a proof of the importance of this extended characteristic set, feature selection retains some of these morphological and topological variables as the most contributive predictors for tissue classification (such as the radius of the wound-bed, the maximum distance to the wound-bed centre, and the ratio of the region perimeter belonging to the wound-bed border).

As addressed by previous similar studies, manual delineation of PU images to trace the perimeter of the wound is a very inaccurate procedure, which gives a high inter- and intra-observer variability [9,6,21] and negatively affects its expected performance as a measurement method. Due to this particular nature of PU images, establishing *gold standard* images to be used as references for image segmentation of all the significant tissues in the images is not a realistic goal. For these reasons, *goodness* evaluation methods for image segmentation [42,43,12] are shown to be much more suitable to analyze the quality of the tissue segmentation process developed for wound images. As future work, measurements of region uniformity and contrast have to be done to estimate the performance of the segmentation process on the set of 113 PU images of this study.

This approach proposes a settled and reliable methodology to support pressure ulcer diagnosis, and to serve as a starting point for more complex models. Although previous studies have highlighted the adequacy of diverse ML approaches for classifying PU tissues from wound segmented images [11,44,12–14]—with some authors proposing complex hybrid ML architectures and heuristic models to deal with effective tissue recognition [12]—most of these works lack an exhaustive analysis that addresses robust model fitting. In this study, parameter fitting has been supported by 10-fold cross-validation analysis on a training set, and a rigorous statistical point of view has been followed (using statistical tests, p-values and confidence intervals) to measure the significance of the differences in efficacy obtained with the different ML models and parameters. Fitted models have been finally validated on a separated testing set to account for their degree of adequacy for recognizing tissues from new wound images.

Our feature selection methodology—based on a wrapper approach with recursive feature elimination—has shown to be effective in keeping the efficacy of the classifiers up. On one hand, reducing the number of predictors contributes to the computational efficiency of the overall system for PU diagnosis (since only those 'surviving' features need to be extracted from the images and considered for pattern classification). On the other hand, feature selection addresses the importance and explanatory power of the different predictive factors (extracted characteristics) in the prediction of tissue classes. In this manner, the comprehensibility of the ML system is improved, and its robustness gets reinforced, so that this reduction of dimensionality does not yet depend on the intrinsic statistical properties of the training pattern set, as it would do if a principal component analysis strategy was followed instead (like in [12]). This latter would contribute to the adaptation of the system to new illumination environments or changing clinical image capturing set-ups. It is important to highlight that some features included in the '*Topological and Morphological*' category—which provide 'contextual' information associated to each PU image as a whole—'survive' after feature selection, thus showing their relevance as predictive variables. Finally, it should be also underlined that, although NNs give the lowest accuracy rates in our experiments, they do it by using the minimum sets of selected features ($S=37$ for both wound-bed and peri-ulcer classification), by far followed by RFs (with $S=73$ and $S=67$ for wound-bed and peri-ulcer classification, respectively) and SVMs (which need the full set of features to give high performance rates in wound-bed tissue detection, as well as 87.5% of the characteristics to deal with effective peri-ulcer tissue detection).

In spite of the high overall accuracy rates obtained by the classifiers in this study, specially with SVMs and RFs, a more detailed statistical analysis on each tissue type, by observing different performance indicators such as sensitivity, specificity and positive/negative predictive value, show a high variability in the results, which rejects any preponderance of a particular ML technique over the rest of the models. Our outcomes coincide also with those given by other similar studies [14,12] and might be motivated by the specific nature of this particular problem of wound image diagnosis, given the heterogeneous distribution of tissue-class prevalence. Due to this fact, clinical professionals should be able to count on the predictions given by different MLs to adapt to the specific diagnostic necessities for each wound and/or tissue type. These detailed measures and statistics given in Tables 2–6 contribute to support the design of future clinical decision-making protocols and algorithms. This way, necrotic and slough tissues require high sensitivity detection rates, given the immediate surgical and/or pharmacological treatments that have to be arranged on the wound when these tissues are identified [45,46]. In this sense, SVMs in this work give the best efficacy scores when detecting necrotic and slough tissues. For their part, RFs are the most sensitive classifiers to recognize granulation tissue. Finally, as future work, one may speculate on the idea that even higher performance rates could be obtained by designing committee machines, hybrid approaches or heuristics that adapt even better to this particular computer vision problem.

6. Conclusions

A clustering-based image segmentation approach along with three different machine learning classifiers—support vector

machines, feed-forward neural networks and random forest decision trees—have been presented to accomplish automatic tissue recognition for pressure ulcer diagnosis, with images taken in environments with non-controlled illumination. Each model has been independently fitted and trained, with significant differences found when comparing the overall efficacy from the three approaches. Support vector machines and random forests gave the high performance rates when classifying wound-bed patterns composed of color, texture, region morphology and topology features extracted from the segmented regions in a set of real pressure ulcer images. High efficacy scores have been also obtained in the classification of necrotic and slough tissues, which are extremely important tissues in pressure ulcer diagnosis. Feature selection has provided a way to significantly reduce the number of predictors and keep the accuracy of the classifiers up. Finally, high efficacy rates have been obtained from the three models when identifying patterns from the peri-ulcer areas in the images.

The methodology and results presented in this paper could have important implications to the field of clinical pressure ulcer evaluation and diagnosis. Starting from pressure ulcer digital images, nurses and clinicians could be provided with an objective and reliable computational tool for automatically detecting, estimating and registering important tissue measurements. Moreover, this methodology could be directly extrapolated to other similar environments such as tissue identification on burn wound images or skin tumor pictures.

Acknowledgment

This research has been partially funded by Consejería de Salud y Bienestar Social, Servicio Andaluz de Salud, Junta de Andalucía, project PI-0027/2012. Additionally, the authors acknowledge support from grants TIN2010-16556 from MICINN-SPAIN and P08-TIC-04026 (Junta de Andalucía), all of which include FEDER funds.

References

[1] European Pressure Ulcer Advisory Panel (EPUAP), Guidelines on treatment of pressure ulcers, EPUAP Rev. 1 (1999) 31–33.
[2] S. Tsuji, S. Ichioka, N. Sekiya, T. Nakatsuka, Analysis of ischemia-reperfusion injury in a microcirculatory model of pressure ulcers, Wound Repair Regen. 13 (2) (2005) 209–215.
[3] D. Gawlitta, W. Li, C.W.J. Oomens, F.P.T. Baaijens, D.L. Bader, C.V.C. Bouten, The relative contributions of compression and hypoxia to development of muscle tissue damage: an in vitro study, Ann. Biomed. Eng. 35 (2) (2007) 273–284.
[4] R. Stratton, C. Green, M. Elia, Disease-related Malnutrition: An Evidence-based Approach to Treatment, CABI Publishing, Wallingford, United Kingdom, 2003.
[5] F. Landi, G. Onder, A. Russo, R. Bernabei, Pressure ulcer and mortality in frail elderly people living in community, Arch. Gerontol. Geriatr. 44 (Supplement 1) (2007) 217–223. http://dx.doi.org/10.1016/j.archger.2007.01.030, URL: ⟨http://www.sciencedirect.com/science/article/B6T4H-4N3M3DN-13/2/e3d4d12edc612e2bd418df9cecfd3521⟩.
[6] D. Beeckman, L. Schoonhoven, J. Fletcher, K. Furtado, L. Gunningberg, H. Heyman, C. Lindholm, L. Paquay, J. Verdu, T. Defloor, EPUAP classification system for pressure ulcers: European reliability study, J. Adv. Nurs. 60 (6) (2007) 682–691.
[7] L.E. Edsberg, Pressure ulcer tissue histology: an appraisal of current knowledge, Ostomy/Wound Manag. 53 (10) (2007) 40–49, URL: ⟨http://www.o-wm.com/article/7934⟩.
[8] O. Cula, K. Dana, F. Murphy, B. Rao, Skin texture modeling, Int. J. Comput. Vis. 62 (1–2) (2005) 97–119.
[9] T.D. Jones, P. Plassmann, An active contour model for measuring the area of leg ulcers, IEEE Trans. Med. Imaging 19 (12) (2000) 1202–1210. http://dx.doi.org/10.1109/42.897812, URL: ⟨http://ieeexplore.ieee.org/xpls/absall.jsp?arnumber=897812isnumber=19435⟩.
[10] S.A. Karkanis, D.K. Iakovidis, D.E. Maroulis, D.A. Karras, M. Tzivras, Computer-aided tumor detection in endoscopic video using color wavelet features, IEEE Trans. Inf. Technol. Biomed. 7 (3) (2003) 141–152.
[11] D. Kosmopoulos, F. Tzevelekou, Automated pressure ulcer lesion diagnosis for telemedicine systems, IEEE Eng. Med. Biol. 26 (5) (2007) 18–22. http://dx.doi.org/10.1109/EMB.2007.901786, URL: ⟨http://ieeexplore.ieee.org/iel5/51/4303486/04312663.pdf?tp=&isnumber=4303486&arnumber=4312663&punumber=51⟩.
[12] F. Veredas, H. Mesa, L. Morente, Binary tissue classification on wound images with neural networks and bayesian classifiers, IEEE Trans. Med. Imaging 29 (2) (2010) 410–427.
[13] H. Wannous, S. Treuillet, Y. Lucas, Robust tissue classification for reproducible wound assessment in telemedicine environments, J. Electron. Imaging 19 (2) (2010) 023002.
[14] H. Wannous, Y. Lucas, S. Treuillet, Enhanced assessment of the wound-healing process by accurate multiview tissue classification, IEEE Trans. Med. Imaging 30 (2) (2011) 315–326, URL: ⟨http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5582291⟩.
[15] P. Berris, Adquisition of skin wound images and measurement of wound healing rate and status using color image processing (P.h.D dissertation), The University of Reading, UK, September 2000.
[16] A.A. Perez, A. Gonzaga, J.M. Alves, in: International Workshop on Medical Imaging and Augmented Reality, 2001, pp. 262–266. http://dx.doi.org/10.1109/MIAR.2001.930300. URL: ⟨http://ieeexplore.ieee.org/search/srchabstract.jsp?arnumber=930300&isnumber=20116&punumber=7405&k2dockey=930300@ieeecnfs&query=%28wound+%3Cin%3E+metadata%29+%3Cand%3E+%2820116+%3Cin%3E+isnumber%29&pos=0⟩.
[17] M. Galushka, H. Zheng, D. Patterson, L. Bradley, Case-based tissue classification for monitoring leg ulcer healing, in: 18th IEEE Symposium on Computer-Based Medical Systems, 2005, pp. 353–358. http://dx.doi.org/10.1109/CBMS.2005.39. URL: ⟨http://ieeexplore.ieee.org/iel5/9903/31475/01467715.pdf?arnumber=1467715⟩.
[18] L.V. Lourega, D.M. Ushizima, G.D. Freitas, M.C. Dornellas, A hybrid image segmentation approach using linear and non-linear processing, in: Proceedings of the International Symposium "Vision by Brain and Machines" (VBM), Montevideo, Uruguay, 2006.
[19] B. Belem, Non-invasive wound assessment by image analysis (Ph.D. thesis), School of Computing, Medical Imaging Group, University of Glamorgan, August 2004.
[20] C. Serrano, B. Acha, T. Gmez-Ca, J.I. Acha, L.M. Roa, A computer assisted diagnosis tool for the classification of burns by depth of injury, Burns 31 (3) (2005) 275–281. http://dx.doi.org/10.1016/j.burns.2004.11.019, URL: ⟨http://www.sciencedirect.com/science/article/B6T52-4FHJYHT-1/2/d836d37945f0b1c939f9a59373278070⟩.
[21] H. Wannous, S. Treuillet, Y. Lucas, Supervised tissue classification from color images for a complete wound assessment tool, in: Proceedings of the 29th Annual International Conference of the IEEE EMBS, Cité Internationale, Lyon, France, 2007, pp. 6031–6034.
[22] G.E.P. Box, D.R. Cox, An analysis of transformations, J. R. Stat. Soc. Ser. B (Methodolog.) 26 (2) (1964) 211–252.
[23] C. Sussman, B. Bates-Jensen (Eds.), Wound Care: A Collaborative Practice Manual for Physical Therapists and Nurses, Lippincott Williams & Wilkins, 2001.
[24] V. Vapnik, Statistical learning theory, 1998, Wiley-Interscience, New York.
[25] B.D. Ripley, Pattern Recognition and Neural Networks, Cambridge University Press, Cambridge, 2007.
[26] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32, URL: ⟨http://link.springer.com/article/10.1023/A:1010933404324⟩.
[27] M. Kuhn, Building predictive models in R using the caret package, J. Stat. Softw. 28 (5) (2008) 1–26.
[28] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2013. URL: ⟨http://www.R-project.org/⟩.
[29] A. Karatzoglou, A. Smola, K. Hornik, A. Zeileis, Kernlab – an S4 package for kernel methods in R, J. Stat. Softw. 11 (9) (2004) 1–20, URL: ⟨http://www.jstatsoft.org/v11/i09/⟩.
[30] C. Bergmeir, J.M. Benítez, Neural networks in R using the Stuttgart neural network simulator: RSNNS, J. Stat. Softw. 46 (7) (2012) 1–26, URL: ⟨http://www.jstatsoft.org/v46/i07/⟩.
[31] A. Liaw, M. Wiener, Classification and regression by randomforest, R News 2 (3) (2002) 18–22, URL: ⟨http://CRAN.R-project.org/doc/Rnews/⟩.
[32] B. Caputo, K. Sim, F. Furesjo, A. Smola, Appearance-based object recognition using svms: which kernel should I use? in: Proceedings of NIPS Workshop on Statistical Methods for Computational Experiments in Visual Processing and Computer Vision, Whistler, vol. 2002, 2002.
[33] W.N. Venables, B.D. Ripley, Modern Applied Statistics with S, 4th ed., Springer, New York, 2002, iSBN 0-387-95457-0. URL: ⟨http://www.stats.ox.ac.uk/pub/MASS4⟩.
[34] G. John, R. Kohavi, K. Pfleger, Irrelevant features and the subset selection problem, in: Proceedings of the Eleventh International Conference on Machine Learning, vol. 129, 1994, pp. 121–129.
[35] R. Kohavi, G.H. John, Wrappers for feature subset selection, Artif. Intell. 97 (12) (1997) 273–324, relevance. http://dx.doi.org/http://dx.doi.org/10.1016/S0004-3702(97)00043-X URL: ⟨http://www.sciencedirect.com/science/article/pii/S000437029700043X⟩.
[36] M. Gevrey, I. Dimopoulos, S. Lek, Review and comparison of methods to study the contribution of variables in artificial neural network models, Ecol. Model. 160 (3) (2003) 249–264, modelling the structure of aquatic communities: concepts, methods and problems. http://dx.doi.org/10.1016/S0304-3800(02)00257-0 URL: ⟨http://www.sciencedirect.com/science/article/pii/S0304380002002570⟩.
[37] M. Kuhn, K. Johnson, Applied Predictive Modeling, Springer London Limited, London, 2013.
[38] C. Ambroise, G.J. McLachlan, Selection bias in gene extraction on the basis of microarray gene-expression data, Proc. Natl. Acad. Sci. 99 (10) (2002) 6562

–6566. http://dx.doi.org/10.1073/pnas.102102699 arXiv:http://www.pnas.org/content/99/10/6562.full.pdf+html URL: ⟨http://www.pnas.org/content/99/10/6562.abstract⟩.
[39] V. Svetnik, A. Liaw, C. Tong, T. Wang, Application of Breimans random forest to modeling structure–activity relationships of pharmaceutical molecules, in: F. Roli, J. Kittler, T. Windeatt (Eds.), Multiple Classifier Systems, Lecture Notes in Computer Science, vol. 3077, Springer, Berlin, Heidelberg, 2004, pp. 334–343. http://dx.doi.org/10.1007/978-3-540-25966-4-33, URL: ⟨http://dx.doi.org/10.1007/978-3-540-25966-4_33⟩.
[40] D.G. Altman, J.M. Bland, Diagnostic tests. 1: Sensitivity and specificity, BMJ: Br. Med. J. 308 (6943) (1994) 1552, URL: ⟨http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2540489/⟩.
[41] D.G. Altman, J.M. Bland, Statistics notes: Diagnostic tests 2: predictive values, Br. Med. J. 309 (6947) (1994) 102, URL: ⟨http://www.bmj.com/content/309/6947/102.1⟩.
[42] Y. Zhang, A survey on evaluation methods for image segmentation, Pattern Recognit. 29 (8) (1996) 1335–1346.
[43] M.D. Levine, A.M. Nazif, Dynamic measurement of computer generated image segmentations, IEEE Trans. Pattern Anal. Mach. Intell. 7 (2) (1985) 155–164.
[44] F. Veredas, H. Mesa, L. Morente, A hybrid learning approach to tissue recognition in wound images, Int. J. Intell. Comput. Cybern. 2 (2) (2009) 327–347.
[45] C.P. Langlotz, Fundamental measures of diagnostic examination performance: usefulness for clinical decision making and research, Radiology 228 (1) (2003) 3–9.
[46] A.J. Alberg, J.W. Park, B.W. Hager, M.V. Brock, M. Diener-West, The use of overall accuracy to evaluate the validity of screening or diagnostic tests, J. Gen. Intern. Med. 19 (2004) 460–465.

Francisco J. Veredas was born in Málaga, Spain, in 1970. He received the M.Sc. and Ph.D. degrees in computer science from the University of Málaga, Spain, in 1996, and 2004, respectively. His Ph.D. thesis was about correlated activity in networks of spiking neurons. He has been working as an associated professor in computer programming languages from 1998 at the Department of Languages and Computational Sciences, in the University of Málaga. He develops his current research work in the group of Computational Intelligence in Biomedicine of that university. His research interests relate to applied computational intelligence, machine learning, data mining, neural networks and medical imaging.

Rafael M. Luque-Baena received the M.S. and Ph.D. degrees in Computer Engineering from the University of Málaga, Spain, in 2007, and 2012, respectively. He moved to Mérida, Spain in 2013, and is currently a lecturer at the Department of Computer Engineering in the Centro Universitario de Mérida, University of Extremadura. He also keeps pursuing research activities in collaboration with other Universities. His current research interests include visual surveillance, image/video processing, neural networks and pattern recognition.

Francisco J. Martín-Santos was born in Málaga, Spain, in 1957. He received the M.Sc. degree in Nursing from the University of Málaga, Spain, in 1982. He also received the Master in Public Health and Health Administration from the Andalusian School of Public Health, Granada, Spain, in 1990. He works as a Nursing Director (District of Primary Health Care Malaga-Guadalhorce, Andalusian Health Service, Málaga, Spain) and as an associated professor of the Department of Nursing and Podiatry (Faculty of Health Sciences, University of Málaga, Spain). He develops her current research work in the group of "Chronicity, Dependency Care and Health Services" in the University of Málaga.

Juan C. Morilla-Herrera was born in Santo Domingo, Dominican Republic, in 1962. He received the M.Sc. degrees in Nursing and Health Sciences in 1983 and 2010, respectively, from the University of Málaga, Spain, where he also received the Ph.D. degree in Health Sciences in 2011. He has been working as a clinical associate professor since 2009 in the Nursing Department, Faculty of Health Sciences, of the Málaga Province Council. He works as a Nursing Director of Community Health Care Unit and develops his current research in the group "Chronicity, Dependency Care and Health Services" in the Institute of Biomedical Research in Málaga.

Laura Morente was born in Málaga, Spain, in 1970. She received the M.Sc. degree in Nursing and European Nursing from the University of Málaga, Spain, in 1993 and 2008, respectively, and Ph.D. degree in Health Sciences from this same university in 2012. She has been working as an associated professor in Community Nursing from 1996 at the Department of Communitary Nursing, in the University Nursing School of the County Council of Málaga. She develops her current research work in the group of Applications of Artificial Intelligence to Health-Care in the University of Málaga. Her research interests relate to intelligence systems for health-care, wound management and geriatrics.