Comprehensible evaluation of prognostic factors and prediction of wound healing

Marko Robnik-Šikonja[*], **David Cukjati**[◇], and **Igor Kononenko**[*]

[*]University of Ljubljana, Faculty of Computer and Information Science,
Tržaška 25, 1001 Ljubljana, Slovenia
e-mail of corresponding author: Marko.Robnik@fri.uni-lj.si
tel.: +386 1 4768459, fax: +386 1 4768498

[◇]University of Ljubljana, Faculty of Electrical Engineering,
Tržaška 25, 1001 Ljubljana, Slovenia

Abstract

We analyzed the data of a controlled clinical study of the chronic wound healing acceleration as a result of electrical stimulation. The study involved a conventional conservative treatment, sham treatment, biphasic pulsed current, and direct current electrical stimulation. Data was collected over 10 years and suffices for an analysis with machine learning methods.

So far only a limited number of studies have investigated the wound and patient attributes which affect the chronic wound healing. There is none to our knowledge to include treatment attributes. The aims of our study are to determine effects of the wound, patient and treatment attributes on the wound healing process and to propose a system for prediction of the wound healing rate.

First we analyzed which wound and patient attributes play a predominant role in the wound healing process and investigated a possibility to predict the wound healing rate at the beginning of the treatment based on the initial wound, patient and treatment attributes. Later we tried to enhance the wound healing rate prediction accuracy by predicting it after a few weeks of the

wound healing follow-up. Using the attribute estimation algorithms ReliefF and RReliefF we obtained a ranking of the prognostic factors which was comprehensible to experts. We used regression and classification trees to build models for prediction of the wound healing rate. The obtained results are encouraging and may form a basis for an expert system for the chronic wound healing rate prediction. If the wound healing rate is known, then the provided information can help to formulate the appropriate treatment decisions and orient resources towards individuals with poor prognosis.

1   Introduction

Skin is vital organ in a sense that loss of the substantial fraction of its mass immediately threatens life of an individual. Such a loss can result suddenly, either from a fire or mechanical accident, but can also occur in a chronic manner, as in skin ulcers.

In more than a decade of clinical use of an electrical stimulation to accelerate the chronic wound healing at the Institute of the Republic Slovenia for Rehabilitation in Ljubljana each patient and wound were registered and the wound healing process was weekly followed. At the beginning of the study in 1989, wounds were randomly assigned into four treatment groups: conservative treatment, sham treatment, biphasic current stimulation and direct current stimulation. Jerčinović et al. [11] proved that stimulated wounds are healing significantly faster than only conservatively or sham treated wounds and Karba et al. [14] proved that electrical stimulation with direct current is effective only if the positive electrode is placed on the wound surface, which is an invasive method, therefore only stimulation with biphasic current pulses has been in regular use ever since. However, dynamics of the wound healing process does not depend only on the type of the treatment, but also on the wound and patient attributes.

The aims of our study are to determine effects of the wound, patient and treatment attributes on the wound healing process and to propose a system for prediction of the wound healing rate.

So far only a limited number of studies have investigated the wound and patient attributes which affect the chronic wound healing. Skene et al. [26] found that with the presence of the graduated compression the healing occurred more rapidly

for patients with smaller initial ulcer area, shorter duration of ulceration, younger age and when no deep vein involvement was detected on photoplethysmography. The measurement of the ulcer area was found to be the strongest predictor of the ulcer healing. Birke et al. [1] found that a time to complete the wound closure is related to the wound depth and wound diameter. Johnson [12] found four factors influencing the vascular ulcer healing: ABpI (ankle/brachial pressure index), liposclerosis, edema, wound status and ulcer area. Lyman et al. [18] found significant relationship between the wound healing rate and bacterial load.

None of listed studies included treatment attributes. Presently, the quantity of available data permits an application of statistical tools and machine learning methods for analysis of the healing process, as well as of the effects of different therapeutic modalities. In the first step of our analysis we determine which wound and patient attributes play a predominant role in the wound healing process. Then we investigate a possibility to predict the wound healing rate at the beginning of treatment based on the initial wound, patient and treatment attributes. Finally we discuss a possibility to enhance the wound healing rate prediction accuracy by predicting it after a few weeks of the wound healing follow-up.

The paper is organized into five Sections. Section 2 discusses the problem and the collected data. Section 3 describes machine learning tools and algorithms used. In Section 4 we present our findings. Final Section contains discussion.

2   The problem and the data set

During more than a decade of clinical use of the electrical stimulation, data concerning patients, wounds, and their treatment were collected. 266 patients with 390 wounds were recorded in computer database up to date. Unfortunately many patient and wound data are missing and not all wounds were followed regularly or until the complete wound closure which is a common problem of clinical trials. The wound case inclusion criteria was the initial wound area larger than 1cm^2 and at least four weeks (or until the complete wound closure) follow up of the wound healing process. It was fulfilled in 300 wound cases (214 patients).

Among these 300 wound cases in 174 cases the observation period lasted until the complete wound closure and was shorter in 126 cases. For these 126 cases the time to the complete wound closure was estimated from the wound area measurements obtained during the observation period [4, 5]. No significant difference was observed between the actual time to complete the wound closure and the estimated one (from four or more weeks of wound healing observation).

Wounds in the database are described with length, width, depth and grade. Because the time to the complete wound closure is highly dependent on the initial wound extent, a measure of the wound healing rate Θ was defined as an average advance of the wound margin towards the wound center and was calculated as the average wound radius (the initial wound area S_0 divided by the initial perimeter p_0 and multiplied by 2) divided by the time to the complete wound closure T [5]:

$$\Theta = 2\frac{S_0}{p_0 T} \quad \left[\frac{\text{mm}}{\text{day}}\right] \tag{1}$$

Neither the distribution of the wound healing rate was normal nor any of its mathematical transforms was found to have normal distribution; therefore non-parametric statistical analysis was used. Differences in distribution of numeric attributes in groups formed by nominal attributes were determined by Kruskal-Wallis one-way analysis of variance. If the difference was significant, two-sample Kolmogorov-Smirnov test mutually compared distributions of numeric attributes at specific values of nominal attributes. The relationship of nominal attributes were tested with chi-square test. Correlation of two numerical attributes was tested with Spearman's test. The attributes we used in our study are briefly described below, further details and analysis can be found in [6].

2.1 Wound attributes

For an evaluation of the efficacy of a particular treatment modality or for evaluation of influence of the wound and patient attributes on the wound healing it is necessary to periodically follow the wound healing process. In [5] it was demonstrated that following wound area is sufficient to determine wound healing process dynamics and that wound shape can be approximated with an ellipse thus it is enough to periodically follow mutually perpendicular diameters (largest wound diameter and diameter perpendicular to it) of the wound. From the measured diameters the wound area, the perimeter and the width to length ratio are calculated.

The wound depth measurement is invasive, because we have to enter the measuring device into the wound. An alternative measure is grading systems. We used a four stage Shea grading system [25]. The wound depth and grade were collected only at the beginning of the treatment. The wound depth was measured in 43% of cases and the wound grade determined in 94% of the cases. Since the wound depth is strongly correlated to the wound grade and the wound depth values are often missing, the depth was omitted from the further analysis. Due to a strong

correlation between the initial wound area and the perimeter, the perimeter was omitted from the further analysis.

Other wound attributes are a wound type, location (trochanter, sacrum, gluteus and other), elapsed time from spinal cord injury to wound appearance (InjuryAppear) in months, and elapsed time from the wound appearance to the beginning of the treatment (AppearStart). The major wound aetiology was the pressure ulceration (82.7% of the cases). Other aetiologies were the arterial ulceration (1.0%), neurotrophic ulceration (6.3%), traumatic ulceration (6.0%), and vascular ulceration (3.7%).

2.2 Patient attributes

Recorded patient attributes were age, sex, total number of wounds, diagnosis and, in the case of the spinal cord injury, the degree of spasticity. The most frequent diagnosis was the spinal cord injury (71.7%). Trauma appeared in 11.3% of cases, diabetes mellitus in 7.3%, geriatrics in 3.3%, multiple sclerosis in 3.0%, and venous diseases in 3.0% of wound cases.

2.3 Treatment attributes

The wounds were randomly assigned into four treatment groups. 54 (18.0%) wounds were treated only conservatively [10], 23 (7.7%) wounds received also the sham treatment, 42 (14%) wounds were treated with direct current and 181 (60.3%) wounds with biphasic current stimulation. In the sham treated group electrodes were applied to the intact skin on both sides of the wound for two hours daily and connected to stimulators, in which the power source was disconnected and delivered no current. The wounds treated with the direct current were stimulated with 0.6mA for half an hour, an hour or two hours daily. Biphasic current stimulated wounds received biphasic, charge-balanced current pulses for half an hour, an hour or two hours daily with electrodes placed on both sides of the wound [13, 14].

The treatment attributes are the type of the treatment and daily duration of the electrical stimulation in minutes. The electrically stimulated wounds healed at higher rate and extent than other wounds. Over 90% of electrically stimulated wounds healed within 6 weeks. Only 70% of sham treated wounds and 72% of only conservative treated wounds healed within the same period.

2.4 Mathematical model of wound healing rate

Survey of normalized wound area time plots revealed that 51% of wounds which fulfilled the inclusion criteria have exponential wound healing process with an initial delay longer than three and a half days, 40% of wounds have a delay of more than seven days and 9% of wounds have the delay of more then 14 days. A recent study [4] showed that a decrease (or increase) in normalized wound area over time is best described by a delayed exponential model, thus wound healing dynamics is a nonlinear process.

Delayed exponential equation is therefore the structure of mathematical model of the wound healing process and by fitting this model to a particular chronic wound case we can calculate the parameters of the model. We need at least four measurements (performed in at least three weeks) of the wound area before we can estimate the parameters of the mathematical model. We estimated the time to the complete wound closure from the parameters of the mathematical model [5]. We considered the wound healed when its estimated area was smaller than five percent of the initial value and smaller than $1cm^2$. The estimated wound healing rate was calculated with Equation (1) .

The estimated wound healing rates for all wound cases were then compared to the actual ones which were calculated from the observed times to the complete wound closure. We found that the estimated wound healing rate after at least four weeks of wound follow-up did not differ significantly from the actual one [6]. If a wound was followed only three weeks or less, the difference was significant. From the known structure of the mathematical model the wound healing rate can be predicted after at least four weeks of follow-up. In clinical trials four weeks is a short period, but clinical practice prefers shorter time for the treatment outcome prediction.

3 Machine learning algorithms

Presently, the quantity of the available data permits the use of machine learning methods. We used machine learning algorithms for induction of tree based models (regression and classification trees). With the models we predicted the wound healing rate from the initial wound, patient and treatment data. We have also considered the estimated wound healing rate based on the mathematical model [6] and built trees for prediction of the wound healing rate after one, two, three, four, five and six weeks of follow-up. In this manner we have taken temporal

aspect of the data into account.

We tested several algorithms for attribute evaluation which were used in a feature subset selection and for selection of splits in interior nodes of the tree based models.

In this section we present machine learning algorithms and methodology used, and the next Section contains results.

3.1 Tree based models

Tree based models are useful for modeling data where the error minimization is not the only goal, and comprehension of the underlying processes and user's confidence into results are just as important. This is the reason why they are often applied to medical problems.

Nowadays, most statistical packages have some form of tree based modeling, so we only briefly describe this methodology and focus on specifics of our approach. Details and further references can be found in textbooks on machine learning and data mining, for example [19, 27].

To select the appropriate split the learning algorithm has to evaluate several attributes, and decide which would partition the given (sub)problem most appropriately. The criterion in medical problems should not be just the expected error minimization but also the comprehensibility of the tree for human experts. The evaluation of the quality of attributes is of principal importance and has received much attention in the machine learning literature.

3.2 Attribute evaluation

There are two issues in attributes' quality evaluation which we believe are especially relevant for comprehensibility and are often neglected in the context of medical problems. The first is the question of the properties of the split which particular evaluation heuristics favors and the second is the problem of detecting conditionally dependent attributes.

The decision and regression trees are constructed greedily in a top to bottom style. There is no global optimization, therefore splitting decisions made in each node are final. If one would only want to minimize error on the training set, which is an optimistic estimate of the generalized error, then the best greedy heuristic measure is the one which measures error of the splits produced by particular attribute. For decision trees such a measure is defined by Equation (2) and for regression trees it is the mean squared error (MSE) described by Equation

(4). But in practice we use various attributes' quality evaluation measures, each with its own rationale and bias. The majority of such heuristics measure impurity of the splits [2, 20] and assume the conditional independence of the attributes. The idea of Relief algorithmic family [15] is to evaluate partitioning power of attributes according to how well their values distinguish between similar examples. An attribute is given a high score if its values are different for similar observations with different prediction values and it has similar values for similar observations with similar prediction values. Similarity of observations and attributes' values is defined with the distance in the attribute space. Relief algorithms are aware of the contextual information, and can correctly estimate the quality of attributes in problems with strong dependencies between the attributes.

There seems to exist a tendency that trees produced with Relief algorithms are preferred by human experts in terms of comprehensibility [7, 16, 22]. The reason might be in their ability to handle conditionally dependent attributes and their bias which favors attributes capable to distinguish between similar examples with different prediction.

In the following we will shortly present the attributes' quality evaluation measures discussed in this work and applied in the wound healing rate problem. As our application revealed that RReliefF for regression and ReliefF for classification problems were the most effective concerning error and comprehensiveness of the learned models, we devote more space to these measures.

3.2.1 Classification error

Classification error as attribute estimator has some major drawbacks [3] in terms of the structure of the tree. If we classify with the majority class, the classification error as the attribute estimator it is defined as

$$E_C = 1 - \sum_j p(A_j) \max_i p(C_i|A_j), \tag{2}$$

where $p(A_j)$ stands for probability of the j-th value of the attribute A, and $p(C_i|A_j)$ is the probability of the i-th class value given the realization of the j-th attribute value. We did not use this measure due to its proven poor generalization ability [17].

3.2.2 Gain ratio

Gain ratio [20] is the most popular impurity based measure for decision trees. It is entropy based measure defined as

$$gainR(A) = \frac{gain(A)}{entropy(A)} = \frac{\sum_i p(C_i)\log p(C_i) - \sum_i \sum_j p(C_i|A_j)\log p(C_i|A_j)}{\sum_j p(A_j)\log p(A_j)} \quad (3)$$

Its gain part tries to maximize the difference of entropy before and after the split. To prevent excessive bias towards multiple small splits the gain is normalized with the attribute's entropy.

3.2.3 Mean squared error

The MSE criterion is the regression counterpart of the classification error. MSE evaluation of the attribute A is minimum over all possible splits of attribute A of the weighted mean squared error of the predictor ϕ:

$$MSE(\phi, A) = \min_{split\ of\ A} p_L \cdot R_{t_L}(\phi) + p_R \cdot R_{t_R}(\phi), \quad (4)$$

where t_L and t_R are the subsets of cases that go left and right, respectively, by the split based on A, and p_L and p_R are the proportions of cases that go left and right. $R_t(\phi)$ is the mean squared error of the predictor ϕ compared to real prediction values τ_i of instances x_i in the subset t of size N_t:

$$R_t(\phi) = \frac{1}{N_t} \sum_{i=1}^{N_t} (\tau_i - \phi(x_i))^2. \quad (5)$$

$\phi(x_i)$ is the value predicted by ϕ. The usual choice for predictor ϕ is the mean prediction value of examples in each subset. Beside this we used also the linear regression formulas, k-NN method, median value, and kernel regression.

3.2.4 Mean absolute error

The mean absolute error(MAE) criterion is similar to MSE, the only difference being in Equation (5), where instead of squared difference the absolute difference is used.

3.2.5 Relief algorithms

Relief algorithms (Relief, ReliefF and RReliefF) are general and successful attributes' quality evaluation measures [24] and are especially good in detecting conditional dependencies. They provide a unified view on attribute estimation in regression and classification and their quality estimates have a natural interpretation. While they have commonly been viewed as one of the most successful preprocessing algorithms for feature subset selection [8] they have actually been used successfully in a variety of settings: to select splits or to guide constructive induction in the building phase of decision or regression tree learning, as an attribute weighting method, in discretization of numerical attributes, in inductive logic programming, and with association rules based classifier.

The idea of Relief algorithms is to evaluate partitioning power of attributes according to how well their values distinguish between similar observations. An attribute is given a high score if its values separate similar observations with different prediction values and do not separate similar observations with similar prediction values. RReliefF and ReliefF sample the space of observations, compute the differences between predictions and values of the attributes and form a statistical measure for the proximity of the probability densities of the attribute and the predicted value. Assigned quality evaluations are in the range $[-1, 1]$, however, values below zero are assigned to completely irrelevant (random) attributes. The quality evaluation $W(A)$ of attribute A assigned by ReliefF or RReliefF can be interpreted in two ways: as the difference of probabilities and as the portion of the concept explanation.

Equation (6) forms the basis for the difference of probabilities interpretation of the quality estimations of the Relief algorithms: the difference of the probability that two instances have different value of the attribute A if they have different prediction value and the probability that two instances have different value of the attribute if they have similar prediction values. These two probabilities contain additional condition that the instances are close in the problem space and form an estimate of how well the values of the attribute distinguish between the instances that are near to each other.

$$W[A] = P(\text{different value of A}|\text{near instance with different prediction})$$
$$- P(\text{different value of A}|\text{near instance with same prediction}) \quad (6)$$

These two probabilities are mathematical transcription of Relief's idea: the first term rewards attribute if its values separate similar observations with different prediction values and the second term punishes it if it does not separate

similar observations with similar prediction values. As it turned out, this interpretation was quite difficult for human comprehension. Negated similarity (different values) and subtraction of the probabilities are difficult to comprehend for human experts.

Another view on attributes' quality evaluations of ReliefF and RReliefF is possible with the portion of prediction values which the attribute helps to determine [23]. We say that attribute A is responsible for the change of the predicted value of the instance I to the predicted value $b(I)$ if the change of its values is one of the minimal number of changes required for changing the predicted value from I to $b(I)$. We denote this responsibility with $r_A(I, b(I))$. As the number of examples n goes to infinity, the quality evaluation $W(A)$ computed from m cases I from the sample $\mathcal{S}$ for each attribute converges to the number of changes in the predicted values the attribute is responsible for (R_A):

$$\lim_{n \to \infty} W(A) = \frac{1}{m} \sum_{I \in \mathcal{S}} r_A(I, b(I)) = R_A \qquad (7)$$

We interpret Relief's weights $W(A)$ as the contribution (responsibility) of each attribute to the explanation of the predictions. The actual quality evaluations for the attributes in the given problem are approximations of these ideal weights which occur only with abundance of data. On the basis of this interpretation we can present Relief's evaluations in a graphical form as in Figure 1.

In the presented study of wound healing medical experts confirmed the comprehensibility of quality evaluations. This interpretation was preferred also by ecology expert in the problem of modeling of the apparent photosynthesis [7].

4 Results

In the earlier stages of our data exploration we found that RReliefF for regression and ReliefF for classification problems, were the most effective concerning error and comprehensiveness of the learned models, so we present results only for them. The results for mean squared error and mean absolute error as attribute estimators in regression are inferior to the results by Relief algorithms (e.g., average relative squared error of models was 0.64 for RReliefF, 0.73 for MSE and 0.77 for MAE after one week of treatment). The same is true for other settings and for Gain ratio in classification as well (e.g., average classification error was 0.70 for ReliefF and 0.81 for Gain ratio at the beginning of treatment and 0.11 versus 0.32 after 6 weeks of treatment).

4.1 Evaluation of the prognostic factors

Results of the statistical analysis [6] showed that the wound healing rate directly depends on the wound treatment and wound grade, while interactions of the other wound and patient attributes with the wound healing rate are not easy to determine.

The attributes' quality estimates (see Table 1) calculated with RReliefF revealed that the initial wound area, followed by the patients' age and time from wound appearance to treatment beginning, are the most powerful prognostic attributes. Important prognostic attributes are also wound shape (width to length ratio), location, and type of treatment. The attribute "Model estimation" represents the estimated wound healing rate calculated from a model of wound healing dynamics. As expected, its quality estimate increases as we add more and more observations (from 0.0 to 0.67).

Table 1: The quality of wound, patient and treatment attributes assigned by RReliefF.

Attribute	0 weeks	1 week	2 weeks	3 weeks	4 weeks	5 weeks	6 weeks
Area (mm^2)	0.135	0.168	0.171	0.161	0.127	0.123	0.122
Age (year)	0.123	0.114	0.094	0.095	0.096	0.092	0.094
AppearStart (week)	0.119	0.121	0.104	0.131	0.121	0.114	0.115
Width to length ratio	0.096	0.098	0.099	0.095	0.103	0.108	0.113
Location	0.085	0.084	0.085	0.081	0.081	0.081	0.081
Treatment	0.066	0.058	0.051	0.052	0.050	0.051	0.051
InjuryAppear	0.062	0.065	0.044	0.050	0.035	0.040	0.039
Daily treatment (min)	0.046	0.039	0.031	0.035	0.025	0.025	0.026
Grade	0.046	0.039	0.057	0.048	0.048	0.047	0.043
Diagnosis	0.039	0.039	0.038	0.038	0.038	0.038	0.037
Aetiology	0.027	0.025	0.026	0.024	0.024	0.024	0.024
Model estimation	0.000	0.399	0.602	0.626	0.663	0.659	0.670

Having in mind the interpretation of Relief's quality evaluations as the portion of the explained concept we can present the results from Table 1 in a more visually appealing graphical form. We normalized the sum of quality evaluations to 1 and presented the first column (0 weeks) in Figure 1. This Figure, as estimated by experts, is easy to comprehend and provides important information.
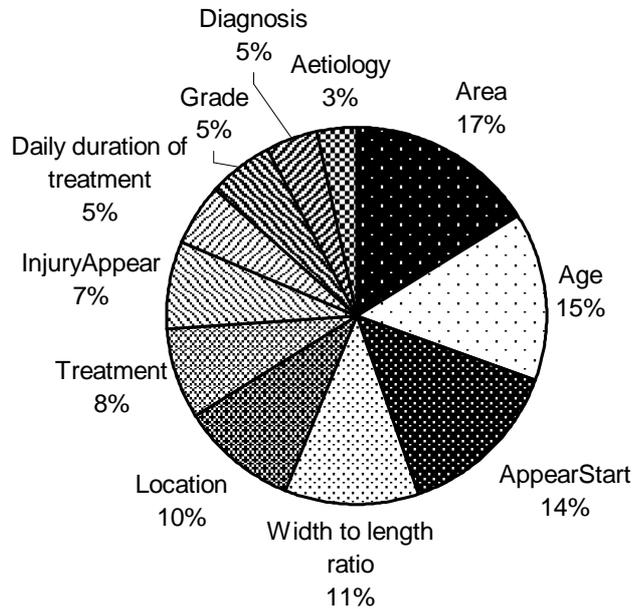
Figure 1: The normalized quality of attributes at the beginning of the treatment assigned by RReliefF.

4.2 Regression trees

We built regression and decision trees with the improved CORE learning system [21]. We used linear equations in the leaves of the tree to model the wound healing rate (we also tried k-NN, median value, mean value, and kernel regression) and stopping rule of minimal five cases in a leaf. We pruned the trees in order to obtain smaller and more comprehensible trees and to get better generalization.

Since the sample size (n=300) was moderate and there are quite some attributes with missing data, we could not afford a separate testing set of sufficient size. We rather used the 10-fold cross-validation as an error estimation method which is known to be the most acceptable estimation method in such cases [9]. An error of the regression trees was measured with the relative squared error (relative error, RE) [2]. The relative error is always nonnegative and should be less than 1 for models of reasonable quality. Trees with the relative error close to 0 produce excellent prediction of the wound healing rate and trees with the relative error around 1 or even greater than 1 produce poor prediction. Some authors use proportion of the variance explained by the regression tree as a measure of the er-

13

ror. It is calculated as (1 - RE). However, this terminology is not quite appropriate [2].

Figure 2 summarizes error of the prediction for the learned trees. The relative error of the prediction at the beginning of the treatment was greater than one, i.e., the resulting regression trees were not usable. By adding the model estimate of the wound healing rate after one week of follow-up we reduced the relative squared error to 0.64. After two weeks the relative error dropped to 0.35 and after three weeks to 0.18. With six weeks of follow-up it was slowly approaching 0.06 (94% of variance was explained by the tree).
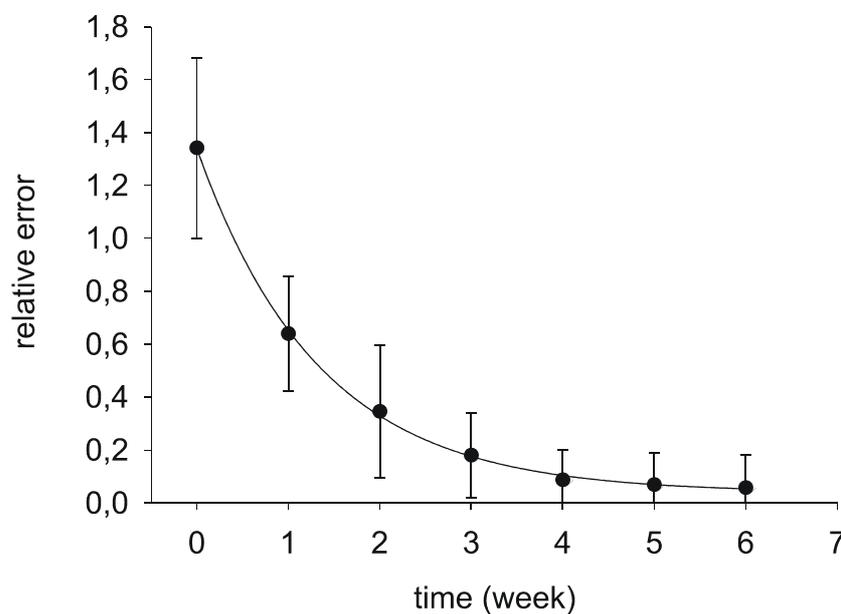

Figure 2: Relative error of regression trees for the wound healing rate prediction as a function of observation time.

Regression trees are more useful than classification trees because the wound healing rate was estimated as continuous variable and can be directly observed in the tree. The minimal follow-up period is two weeks. After five weeks the predicted wound healing rate is equal to the healing rate estimated by the model. However, the predicted wound healing rate with shorter period of follow-up depends also on the wound, patient and treatment attributes. The regression tree built after two weeks of follow-up is presented in Figure 3. The type of the treatment

is indirectly included in the regression trees as daily duration of the treatment, which was zero in case of conservatively treated wounds. Important prognostic attributes seem to be the wound area, grade, shape (width to length ratio), patients age, elapsed time from the spinal cord injury to the wound appearance and elapsed time from the wound appearance to the beginning of the treatment.
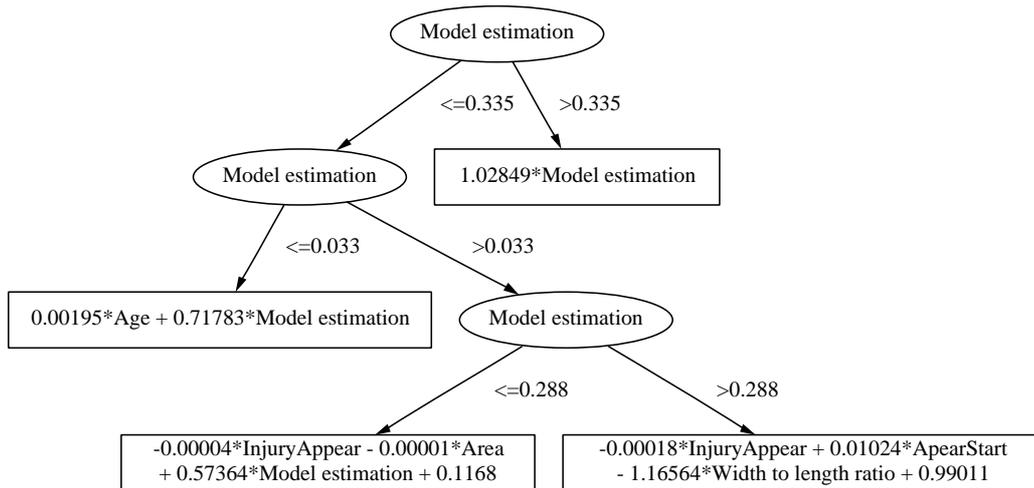

Figure 3: The regression tree with linear equations in the leaves for prediction of the wound healing rate after two weeks of treatment.

4.3 Classification trees

Sometimes we do not need the exact quantity of the wound healing rate, so we decided to divide it into four categories (classes) according to Table 2. We performed the analysis on such classification problem.

We built classification trees with ReliefF as attribute estimation measure. We classified with the majority class as the prediction model in the leaves of the tree (we also tried k-NN) and a stopping rule of at least five cases in a leaf. Pruning was applied in order to obtain smaller and more comprehensible trees and to get better generalization. Again, we used the 10-fold cross-validation as an error estimation method. The error of the trees was measured as a proportion of the misclassified test samples.

15

Table 2: Dividing 300 wound cases into four classes according to their wound healing rate.

class	condition (mm/day)	no. of cases	%
NON HEALING WOUNDS	$\Theta \leq 0.095$	77	0.257
SLOW HEALING WOUNDS	$0.095 < \Theta \leq 0.180$	77	0.257
MEDIUM HEALING WOUNDS	$0.180 < \Theta \leq 0.300$	67	0.223
FAST HEALING WOUNDS	$\Theta > 0.300$	79	0.263

At the beginning of the wound treatment only the initial wound, patient and treatment data are available. The resulting classification error was 70%, which is not much below classification with a priori the most probable class (74%). By adding the model estimate of the wound healing rate after one week of follow-up to the data set, we decreased the classification error to 59%. With data available for two weeks the classification error was 38% and with three weeks 20%. Afterwards it is slowly approaching 10% with six weeks of follow-up. Figure 4 summarizes the results. In the trees built after two weeks of follow-up only the model estimate of the wound healing rate can be found in the tree nodes.

We found out that accurate prediction of the wound healing rate is possible with data available for at least three weeks of follow-up. Therefore, with classification trees we also managed to shorten the time of follow-up for one week compared to the mathematical model. Only rough estimate of the wound healing rate is possible after two weeks.

5 Discussion

We estimated the prognostic power of the wound, patient and treatment attributes with RReliefF algorithm. The results revealed that the initial wound area, followed by the patients' age and time from the wound appearance to the beginning of treatment, are the most important prognostic attributes. Important prognostic attributes are also the wound shape (width to length ratio), location, and type of the treatment. The experts agree with these findings because they understand the meaning of these quality evaluations.

The dynamics of wound healing can be accurately predicted after at least four weeks of the wound healing process follow-up. Therefore for accurate wound
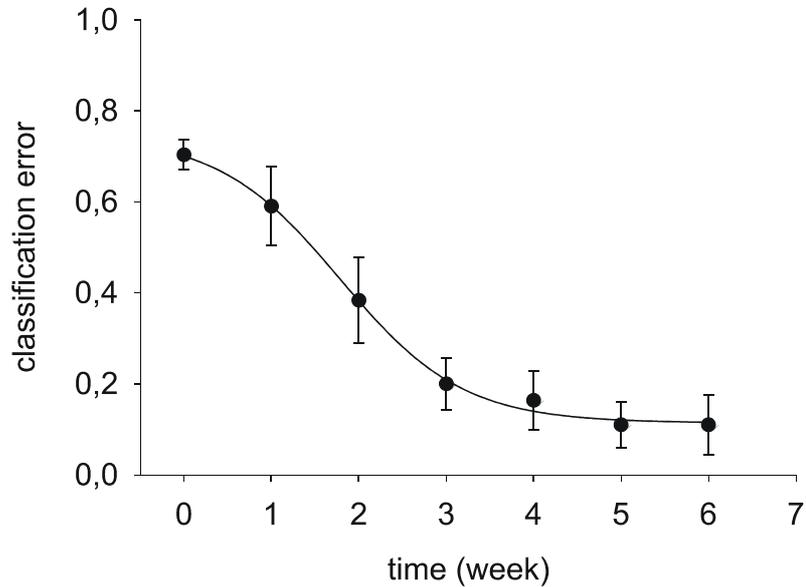
Figure 4: The classification error of the classification trees for the wound healing rate prediction as a function of observation time.

healing rate prediction, wounds should be followed for at least four weeks. In clinical practice the wound healing rate or the time to complete wound closure should be estimated as soon as possible to select a proper treatment and thus to improve the patient care. Prediction of the wound healing rate from the initial wound, patient and treatment data is not possible. The best prognostic factors are weekly follow-up measurements of wound area. We determined that the minimal follow-up period is two weeks. After three weeks we were able to predict the wound healing rate with 80% classification accuracy (using the classification trees), and explain 82% of the variance with the regression trees. The best results were obtained using regression trees with linear equations in the leaves. In the literature also other wound and patient attributes were reported to have prognostic value. By considering those additional attributes which, unfortunately were not measured in our study, our prediction might be even more accurate.

The presented regression trees in combination with the mathematical model of the wound healing process dynamics present a basis of an expert system for the chronic wound healing rate prediction. If the wound healing rate is known,

the provided information can help to formulate the appropriate management decisions, reduce the cost, and orient resources towards individuals with poor prognosis.

Acknowledgements

References

[1] J. A. Birke, A. Novick, C. A. Patout, and W. C. Coleman. Healing rates of plantar ulcers in leprosy and diabetes. *Leprosy Rev.*, 63:365–374, 1992.

[2] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and regression trees*. Wadsworth Inc., Belmont, California, 1984.

[3] C. E. Brodley. Automatic selection of split criterion during tree growing based on node location. In A. Prieditis and S. Russell, editors, *Machine Learning: Proceedings of the Twelfth International Conference (ICML'95)*, pages 73–80. Morgan Kaufmann, San Francisco, 1995.

[4] D. Cukjati, R. Karba, S. Reberšek, and D. Miklavčič. Modeling of chronic wound healing dynamics. *Med. Biol. Eng. Comput.*, 38:339–347, 2000.

[5] D. Cukjati, S. Reberšek, and D. Miklavčič. A reliable method of determining wound healing rate. *Med. Biol. Eng. Comput.*, 39(2):263–271, 2001.

[6] D. Cukjati, M. Robnik-Šikonja, S. Reberšek, I. Kononenko, and D. Miklavčič. Prognostic factors in the prediction of chronic wound healing by electrical stimulation. *Med. Biol. Eng. Comput.*, 39:542–550, 2001.

[7] A. Dalaka, B. Kompare, M. Robnik-Šikonja, and S. Sgardelis. Modeling the effects of environmental conditions on apparent photosynthesis of Stipa bromoides by machine learning tools. *Ecological Modelling*, 129:245–257, 2000.

[8] T. G. Dietterich. Machine learning research: Four current directions. *AI Magazine*, 18(4):97–136, 1997.

[9] T. G. Dietterich. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, 10(7):1895–1924, 1998.

[10] J. A. Feedar and L. C. Kloth. Conservative management of chronic wounds. In L. Kloth and J. McCulloch, editors, *Wound Healing: Alternatives in Management*, pages 135–172. F. A. Davis Co., Philadelphia, 1990.

[11] A. Jerčinović, R. Karba, L. Vodovnik, A. Stefanovska, P. Krošelj, R. Turk, I. Džidić, H. Benko, and R. Šavrin. Low frequency pulsed current and pressure ulcer healing. *IEEE Trans. Rehab. Eng.*, 2(4):225–233, 1994.

[12] M. Johnson. Using cluster analysis to develop a healing typology in vascular ulcers. *J. Vasc. Nurs.*, 15:45–49, 1997.

[13] R. Karba, L. Vodovnik, M. Prešern-Štrukelj, and M. Klešnik. Promoted healing of chronic wounds due to electrical stimulation. *Wounds*, 3(1):16–23, 1991.

[14] R. Karba, D. Šemrov, L. Vodovnik, H. Benko, and R. Šavrin. DC electrical stimulation for chronic wound healing enhancement. Part 1. Clinical study and determination of electrical field distribution in the numerical wound model. *Bioelectrochem. Bioenerget.*, 43:265–270, 1997.

[15] K. Kira and L. A. Rendell. A practical approach to feature selection. In D. Sleeman and P. Edwards, editors, *Machine Learning: Proceedings of International Conference (ICML'92)*, pages 249–256. Morgan Kaufmann, San Francisco, 1992.

[16] I. Kononenko, E. Šimec, and M. Robnik-Šikonja. Overcoming the myopia of inductive learning algorithms with RELIEFF. *Applied Intelligence*, 7: 39–55, 1997.

[17] D. J. Lubinsky. Increasing the performance and consistency of classification trees by using the accuracy criterion at the leaves. In A. Prieditis and S. Russell, editors, *Machine Learning: Proceedings of the Twelfth International Conference (ICML'95)*, pages 371–377. Morgan Kaufmann, San Francisco, 1995.

[18] I. R. Lyman, J. H. Tenery, and R. P. Basson. Corelation between decrease in bacterial load and rate of wound healing. *Surg. Gynecol. Obstet.*, 130(4): 616–620, 1970.

[19] T. M. Mitchell. *Machine Learning*. McGraw-Hill, New York, 1997.

[20] J. R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Francisco, 1993.

[21] M. Robnik Šikonja. CORE - a system that predicts continuous variables. In F. Solina and B. Zajc, editors, *Proceedings of Electrotehnical and Computer Science Conference (ERK'97)*, pages B145–148. Slovene section of IEEE, Ljubljana, 1997.

[22] M. Robnik-Šikonja, D. Cukjati, and I. Kononenko. Evaluation of prognostic factors and prediction of chronic wound healing rate by machine learning tools. In S. Quaglini, P. Barahona, and S. Andreassen, editors, *Proceedings of Artificial Intelligence in Medicine Europe, AIME 2001*, pages 77–87. Springer, Berlin, 2001.

[23] M. Robnik-Šikonja and I. Kononenko. Comprehensible interpretation of Relief's estimates. In C. E. Brodley and A. Danyluk, editors, *Machine Learning: Proceedings of the Eighteenth International Conference on Machine Learning (ICML'2001)*, pages 433–440. Morgan Kaufmann, San Francisco, 2001.

[24] M. Robnik-Šikonja and I. Kononenko. Theoretical and empirical analysis of ReliefF and RReliefF. *Machine Learning Journal*, 53:23–69, 2003.

[25] D. J. Shea. Pressure sores classification and management. *Clin. Orthop. Rel. Res.*, 112:89–100, 1975.

[26] A. I. Skene, J. M. Smith, C. J. Doré, A. Charlett, and J. D. Lewis. Venous leg ulcers: a prognostic index to predict time to healing. *Brit. Med. J.*, 305: 1119–1121, 1992.

[27] I. H. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann, San Francisco, 1999.