Chapter 6

An Expert System Based on Computer Vision and Statistical Modelling to Support the Analysis of Collagen Degradation

Yaroslava Robles-Bykbaev, Salvador Naya,
Silvia Díaz Prado, Daniel Calle-López,
Vladimir Robles-Bykbaev, Luis Garzón-Muñóz,
Clara Sanjurjo Rodríguez and Javier Tarrío Saavedra

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.72982

Abstract

The poly(DL-lactide-co-glycolide) (PDLGA) copolymers have been specifically designed and performed as biomaterials, taking into account their biodegradability and biocompatibility properties. One of the applications of statistical degradation models in material engineering is the estimation of the materials degradation level and reliability. In some reliability studies, as the present case, it is possible to measure physical degradation (mass loss, water absorbance, pH) depending on time. To this aim, we propose an expert system able to provide support in collagen degradation analysis through computer vision methods and statistical modelling techniques. On this base, the researchers can determine which statistical model describes in a better way the biomaterial behaviour. The expert system was trained and evaluated with a corpus of 63 images (2D photographs obtained by electron microscopy) of human mesenchymal stem cells (CMMh-3A6) cultivated in a laboratory experiment lasting 44 days. The collagen type-1 sponges were arranged in 3 groups of 21 samples (each image was obtained in intervals of 72 hours).

Keywords: computer vision, collagen degradation, statistical modelling, long short-term neural networks

1. Introduction

The statistical analysis (almost classical) of data collected through techniques like segmentation of images of biomaterials focuses their attention on a descriptive analysis and an implication

analysis or quasi-implication analysis. This approach can cause the loss of the study of possible influences of variables relative to the experiment of cell seed over biomaterials like type I collagen.

However, there are alternatives to classical statistical analysis based on classification methods, descriptive statistics, implication statistics and quasi-implication statistics, among others. One of them is the hybrid methodology based on the application of neural networks, image segmentation and statistical modelling of the probable relation of variables that affect biomaterials degradation like type I collagen in the mesenchymal stromal cell culture over those biomaterials. This hybrid system becomes a robust system with high complexity and low computational cost. On the one hand, it allows a reliable analysis of experimental data relative to the seeding of those cells, that is, to establish if there is any possible relationship between the medium of cell culture and the degradation of the biomaterial where the named cell lineage is seeded. On the other hand, the system enables the making of decisions taking into account into the acquired data after the application of the analytical model.

Meanwhile, image segmentation allows the improvement of images to use them later and the following data collection to analyse this data statistically. While neural networks are capable of improving data prediction [1], the statistical modelling allows identifying and explaining possible relationships among variables (predictor ones) that could influence in the degradation of type I collagen as regards time (in vitro one). Therefore, the hybrid system encompasses several data analysis systems. In addition to the cluster analysis, the system includes an alternative statistic for improving the method to examine the experimental results of the cell culture over biomaterials.

The growing requirement for making new materials compatible with life implies not only their design and tests, but it also involves the statistical study of the relationship between the type of biomaterial and the cells seeded in it. This analysis is necessary due to the following experimental phases that depend upon it. Some examples of this kind of studies are:

- Statistical analysis devoted to determining the biomaterials degradation. This kind of studies allows a better identification of the effects of the variables under investigation like the type of biomaterial, the cell group, the culture medium, the time of cell growth and the degradation of the biomaterial as regards time [2–5].
- Chen et al. developed a numeric model taking into account the stochastic hydrolysis and the transportation of mass to simulate the biomaterials degradation process and their erosion [2].
- Hoque et al. have modelled the loss of mass using an exponential expression. They made it, supposing that water diffusion and water hydrolysis are the leading causes of the degradation processes of the biomaterials under analysis [3].
- To our purpose, we applied statistical learning tools (field related to the interrelation between statistics and informatics) relative to complex data for modelling tends of degradation and reliability regarding the studied materials [6].

The above research shows up the need for modelling statistically not only the mechanical, physical or rheological phenomes of biomaterials like type I collagen, as well as the requirement

to model the degradation degree for this biomaterial in cell cultures. But also, it is necessary to statistically model the cell growth and cell distinction. All these studies establish robust methods to analyse these degradation processes thanks to the contribution with additional information about the variables effects. The analysis of these variables is usually unknown with the application of just descriptive statistical methods.

The current proposal focuses their attention on a smart model that combines segmentation and analysis of images to get databases, which can be analysed later with a statistical model. Additionally, the intelligent system applies neural networks to the previously obtained data for improving the capacity of prediction processes. Namely, the smart system is a methodological proposal that allows predicting and understanding the behaviour of experimental data of biological populations.

2. Baseline methodologies for the system development

The objective of this proposal is to determine the relationship among the degradation of type I collagen where we seeded mesenchymal stromal cells. This deterioration was conditioned by the time (period) of study we made the observation.

The experiment was done through the statistical modelling of the type I collagen degradation degree. Additionally; we did a previous segmentation analysis (particle identification and particle detachment) of images acquired by an optical microscope and coloured with haematoxylin-eosin techniques.

This intelligent system allows the improvement of making decisions and conclusions because its methodology is more precise thanks to the system that applies data arithmetic analysis. Thus, decision-making is made according to a robust statistical analysis of the relationship between the type I collagen degradation and the presence of possible influent variables in that degradation.

As is depicted in **Figure 1**, below we describe all methodologies used to make up our proposal.

2.1. Generalised linear models (GLM)

Regarding the statistical modelling of experiments relative to the cell culture, that is, modelling the biological behaviour, we applied GLM because this kind of models allows some degree of flexibility for this type of data. Nelder and Wedderburn used GLM for the first time in 1972. These models let variables to follow an exponential probability distribution and not just a normal distribution [7].

Concerning the summary of the GLM function, this last one does not produce a p-value to the model nor an R^2. The maximum verisimilitude estimation is the base for estimation and inference with GLM, even though, maximisation of probability requires an iterative method for the least-squares approach [8].

GLM capacity to adjust the mean of the data μ, instead of the data, is the base to choose these models. In a GLM context, a reasonable approach would be to select among models considering their capacity to maximise log-likelihood $l(\beta; \mu)$ instead of $l(\beta; y)$. But, to apply this focusing,

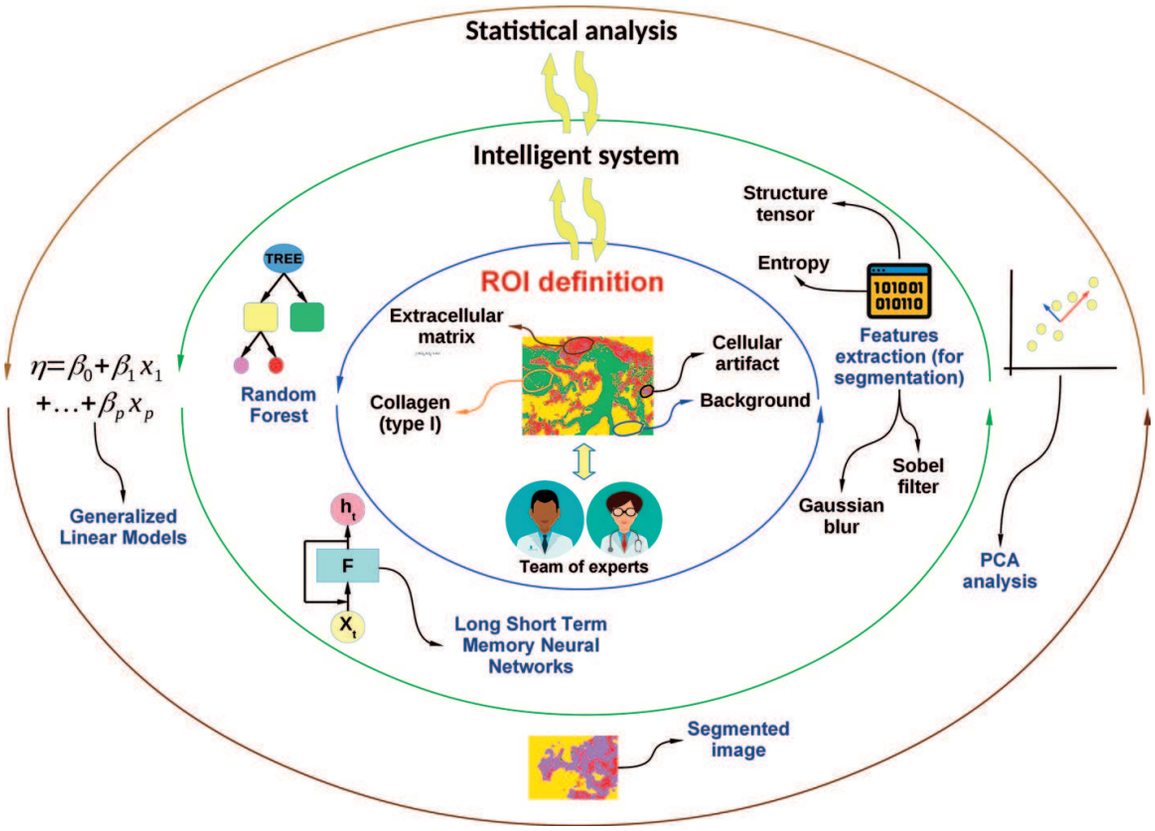

Figure 1. General architecture of the proposed approach.

it is necessary to estimate l (β; μ) first. Likewise, it is essential to select the model with the lowest value of Akaike information criterion (AIC) [8].

In GLM models, the linear predictor is

$$n = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p \quad (1)$$

where

h(μ) = η

h = link function

η = nonadditive lineal predictor

2.1.1. Predictive values

Predictive values communicate what the value of result would be expected according to the observed pattern between the co-variables and the outcome. At least three different values are essential for us, which can be calculated by regression adjustment:

1. Values adjusted to representative or particular values of X
2. Values adjusted to the mean of X
3. Values adjusted to the arithmetic mean [8]

http://dx.doi.org/10.5772/intechopen.72982

2.2. Computer vision in decision-making for cell differentiation

Nowadays, the computer vision allows us determining the cells' biological behaviour such as the growth and cell differentiation as well as the biomaterials degradation (collagen type I). In this line, the image segmentation is an important methodology used to achieve this objective: "a central problem in many studies, and often considered as the cornerstone of image analysis, is its segmentation" [9]. That is why "the type and quality of the acquired images influence the success of cell segmentation (identification and separation of objects)" [10].

In the same line, although segmentation seems a process with a certain degree of complexity and "although the segmentation is conceptually simple, it lacks generality and, therefore, cannot be implemented reliably and effortlessly in all cell lines, modalities of image and densities of cells without pre-processing images" [11].

The limitations such as the specific needs related to the research, the type of objects to be treated in the image, the objectives pursued by the research and the restricted knowledge of the technician in charge of the segmentation process lead to the need to create specialised proposals for the treatment and image analysis. The absence of a universal image segmentation procedure is no surprise; however, it is now possible to analyse the 2D images of the behaviour of stem cells in vivo, such as sequential growth and differentiation with time using various techniques [11]. The most commonly used segmentation and image processing techniques include colour threshold, region growth, edge detection and Markov random fields (MRF) [12].

2.2.1. Random forest classifier

Using the extracted descriptors from the images, it is possible to apply any classifier to perform the image segmentation by pixels (a division of the image into different segments or groups of pixels that share certain characteristics). In this research, we have used a classifier that allows operating on modular attributes avoiding the overfitting of certain classes and has an optimised computational cost.

In recent years, decision forests have established themselves as one of the most promising techniques in machine learning, computer vision and medical image analysis [13].

The random forests operate by constructing several decision trees (predictive processes that map observations on an article to conclusions about the objective value of the article) in the training phase, to then result in class fashion (by its nature as a classifier) for each tree.

In order to train the classifier, it is necessary to define attributes, so it is necessary to extract the following information for each region of interest (ROI) in the training images:

- Structure tensor

A structure tensor is a matrix representation of the image partial derivatives defined as the second-order symmetric positive matrix **J**:

$$J = \begin{bmatrix} < f_x, f_x > w & < f_x, f_y > w \\ < f_x, f_y > w & < f_y, f_y > w \end{bmatrix} \tag{2}$$

where f_x and f_y are the images of the partial spatial derivatives, ∂f/∂x and ∂f/∂y, respectively [14].

From this matrix, all major and minor eigenvalues are separated for each pixel and channel in the image:

$$T(v) = \lambda v \tag{3}$$

- Entropy

Draws a circle of radius **r** around each pixel, obtains the histogram of that separated circle as fragments of binarised image and then calculates the entropy as $\sum_{pin\ histogram} - p * log_2(p)$ for each particle, where **p** is the probability of each chunk in the histogram of each channel of the image, in both RGB and HSB.

- Gaussian blur

In order to obtain the features related with the Gaussian blur, we perform circumvolutions with a Gaussian function to smooth; for this we define the following values:

- σ represents the decay radius exp(−0.5)~61%. For example, the standard deviation σ of the Gaussian.
- Scale units represents that the value of σ is not in pixels but in units defined by the scale of the image size.

Then, for the process of extracting attributes, we perform **n** individual circumvolutions with Gaussian nuclei with **n** normal variations of **σ**. The larger the radius, the more unfocused the image will be until it reaches the point where the pixels are homogeneous [15]:

$$\boldsymbol{\sigma}_{min}, 2\,\boldsymbol{\sigma}_{min}, 4\,\boldsymbol{\sigma}_{min}, \ldots, 2^{n-1}\,\boldsymbol{\sigma}_{min} \tag{4}$$

where $2^{n-1}\,\boldsymbol{\sigma}_{min} \leq \boldsymbol{\sigma}_{max}$.

It should be noted that for all convolution operations, the pixels that are outside the image are assigned the value of the pixel corresponding to the nearest edge. This gives more weight to the pixels at the edge of the image with respect to the central ones and greater weight to the pixels of the corners than to the non-corners [16].

- Sobel filter

The Sobel operator, sometimes called the "Sobel-Feldman operator" or "Sobel filter", is used in image processing and computer vision, particularly in edge detection algorithms where images are generated with sharp edges. This operator makes a measurement of the spatial gradient of an image in order to highlight areas with high spatial frequency that corresponds to the edges. The numerical analysis shows that for certain kinds of surfaces, an even better estimate can be obtained by using the average weights of three such central differences [17].

For the extraction of attributes related to this filter, an approximation of the gradient of the intensity of each pixel in the image is calculated. Prior to the application of the filter, Gaussian blurs are applied varying the value of σ.

Based on the identification of the components of the training images, the intelligent system will be in charge of receiving the pertinent attributes to said classes, to later classify each pixel of all the images of the corpus. The correct identification of these classes allows generating probability maps [15] for each object class of CMMh3A6 cells and extracting attributes such as the area for further analysis. With the probability maps assembled, we proceeded to segment the images and extract several interesting features (physical features) such as area, mean, ratios, etc.

2.3. Long short-term memory neural networks as forecasting support tools

In addition to the statistical modelling techniques that were applied to model the level of degradation of type I collagen, a long short-term memory neural network (LSTM NN) was implemented. A network of this type is characterised by being able to learn long-term dependencies, an aspect that makes them an ideal strategy to carry out prediction processes based on previously viewed values. Traditional recurrent neural networks (RNNs) work with predetermined time lags in order to learn the processing of temporal sequences. This aspect makes it inappropriate to use an RNN for the problem described in this chapter, since the time periods in which the laboratory samples that are taken could be variable. When using an NTS LTSM, we have two important advantages over traditional RNNs: (i) it is feasible to take a long number of samples to train the system, and (ii) the optimal time window size can be variable [18].

Figure 2 presents the general architecture of the LSTM NN used. The basic unit of this network is the block of memory that contains one or more memory cells and a pair of adaptive, multiplicative gather units which gate input and output to all cells in the block. Memory blocks allow cells to share the same gates in order to reduce the number of adaptive parameters. Each memory cell has a linear unit called Constant Error Carousel (CEC) connected to it. This unit allows that when there is no new input or error signals sent to the cell, the local value of the error (CEC) remains constant [18, 19].

In this line, we have used in this research a LSTM NN with the aim of forecasting intermediate values of the collagen type I degradation. Originally, the laboratory samples were taken each day (21 samples), whereas with the neural network, we can generate around 180 projection values (for every variable such as area, perimeter, diameter, eccentricity or roundness, mean intensity, centroid (x, y), skew and kurtosis, with respect to time). To this aim, the neural network was trained with the 21 original samples, and posteriorly, it predicted the rest of the values according to intervals of 0.25* day (6 hours).

2.4. Principal component analysis (PCA)

During the last decades, some techniques such as the Fourier-transform infrared spectroscopy (FTIR) has shown their high potential carrying out some genetic studies as well as providing support as a complementary tool for immunohistochemical methods. The great development that experimented several techniques of molecular biology aimed at the study of cell differentiation has shown that implementing alternative approaches to perform the analysis of an important set of data that can be obtained from in vivo or in vitro tests is necessary [20].

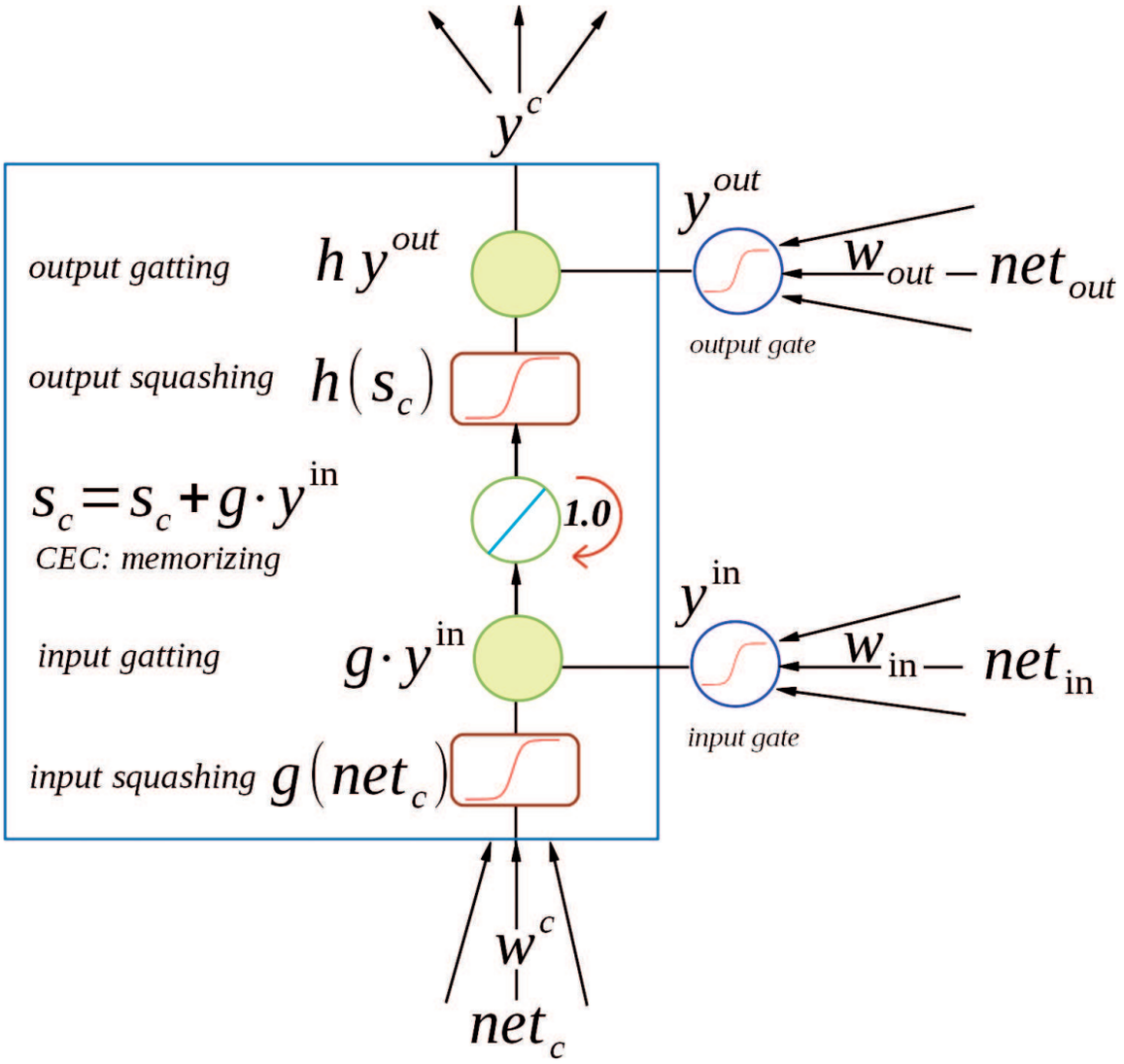

Figure 2. General architecture of a LSTM NN [19].

The principal component analysis (PCA) is a statistical method that has been widely used in several types of research of different scientific areas. Given a set data obtained experimentally, this method allows selecting the most representative variables and, consequently, reducing the dimensionality. For example, nowadays this method is used to perform different tasks such as X-ray fluorescence image analysis [21, 22] to identify objects; classify and extract features from gastric cancer images [23]; determine complex interrelations between patients, diseases and the best treatments for lung cancer [24]; or analyse sets of data obtained from brain magnetic resonance imaging (MRI) [25].

3. Experiment: Materials and samples getting

Biomaterials are considered like mechanically, functionally and physiologically acceptable products used to replicate the function of living tissues in biological systems securely. These

products are implanted temporarily or permanently inside a body to try the restoration of the existent effect and, in some cases, tissue regeneration [26].

We cultivated human mesenchymal stem cells 3A6 (CMMh-3A6); it means a lineage of immortalised mesenchymal stem cells, which were given by the Department of Medical Research & Education and Orthopaedics & Traumatology, Veterans General Hospital, Taipei, Taiwan.

We cultivated 4'200.000 CMMh-3A6 cells from x + passages. We made two changes in culture medium and one subcell culture (passage2 or overseeding) per week. These changes were made due to the observed plaques by the inverted microscope; we saw a great confluence (90%) in a relatively short period (approximately between 3 and 4 days).

The biomaterial type I collagen (trade house) was obtained by using a biopsy punch and cutting the biomaterial in shape of 8 mm diameter discs. We arrange Col I sponges in 3 groups of 21 samples each of them. For technical purposes, we call group #1, #2 and #3 like CCO, CCT and CO, respectively.

The 21 samples of group No. 1 (CCO) were type I collagen, CMMh-3A6 cells and commercial osteogenic cell culture medium: hMSC Osteogenic Differentiation BulletKit™ Medium (Lonza, España). The group No. 2 (CCT) had 21 samples formed by CMMh-3A6 cells, type I collagen and cell culture medium DMEM (Dulbecco's modified essential medium). These cells had 1 g/L D-glucose and pyruvate (Gibco, Estados Unidos), 5% glutamax (Gibco) and 10% foetal bovine serum (Gibco). This cell culture medium was seeded over type I collagen biomaterial. Finally, the group N° 3 (CO) had only 21 samples with type I collagen and commercial osteogenic cell culture medium—hMSC Osteogenic Differentiation BulletKit™ Medium (Lonza, España)—but without cells. Therefore, the last group was the group of control.

The experiment lasted 44 days in total, and it was under conditions to replicate human organism (culture oven): pH = 7.4, temperature = 37°C and 5% of CO2. The samples were sent to histomorphology to embed them in paraffin. After that, paraffin was removed from the samples (de-paraffinisation, de-waxing). Later, we stain the samples with haematoxylin and eosin. Finally, we took 2D photographs with electronic microscopy. The result was 60 photos, which were segmented through machine learning algorithms for each group of analysis with a set of binary features. The classification takes into account according to the following characteristics: type I collagen, extracellular matrix, image artefacts and background.

That process requires extraction of attributes to train the system, characteristics that are the base for the learning process. Then, we extract a set of binary images (one per each segmentable attribute), and from these pictures, we get relevant information for detecting these characteristics in any photo.

4. Results

We segmented the images of the three experimental groups (group #1 CCO, group #2 CCT and group #3 CO). The group of control was group #3 because it had only type I collagen and commercial osteogenic cell culture medium without cells: hMSC Osteogenic Differentiation

BulletKit™ Medium (Lonza, España). Groups #2 and #1 contained not only cell culture medium (osteogenic and non-osteogenic, respectively), but also they had mesenchymal stromal cells. Besides, each group had 21 samples. Then, the results we got from the image segmentation with and without the use of a neural network and after the application of generalised linear models (GLM) to the database are:

4.1. Generalised linear models (GLM) without the neural network

In this section, we show up the results we got from the statistical modelling of the data group that was acquired from the image segmentation and before the application of a neural network.

GLM modelling: collagen degradation as regards time + cell culture group—time effects and cell group in the degradation of type I collagen.

With this model, where $\eta = \mu_x = \beta_0 + \beta_1 x_1 (Time) + \beta_2 x_2 (group)$, μ_x follows a normal distribution. The variance proportion explained in the model (residual deviance) was apparently small (4.0879e-05), and the AIC was 1454.2.

To test $H0{:}\beta_0 = 0$, we use $z = 2.049$ ($p - value = 2.70e-08$). Consequently, the cell culture group, as regards time, seems to have a meaningful impact on the probability of the type I collagen degradation after time goes (i.e., once the model includes that variable "glm.without.network"). Namely, that model has the best p-value for time and the group and the smallest values for AIC and residual deviance compared to the other proposed models (see **Table 1**).

Next, we display the graphic diagnosis of the model (see **Figure 3**):

The normal QQ-plot shows normalised standard waste. Waste for predicted values is in the left panel. This waste presents a tendency to the mean; therefore, the error independence condition is fulfilled. It means the lower left panel exhibits that collagen degradation (pixels) has a tendency.

The right panel displays how the waste values of the model "glm.without.network" adjust to the regression line of the model, and there are atypical values. Probably, these atypical results belong to the error allowed in this type of experiment, as it is complicated to control shifts that occur due to the intrinsic activity of the cell samples under culture.

To determine the waste normality, we applied the Shapiro–Wilk test, where:

- HO: this sample comes from a normal distribution.
- H1: this sample does not come from a normal distribution.

Coefficients	P-value
(Intercept)	<2e-16***
Time	2.70e-08***
Group	0.0453*

Table 1. Statistical significance of the time and group co-variables in the results of the model without neural network.

http://dx.doi.org/10.5772/intechopen.72982

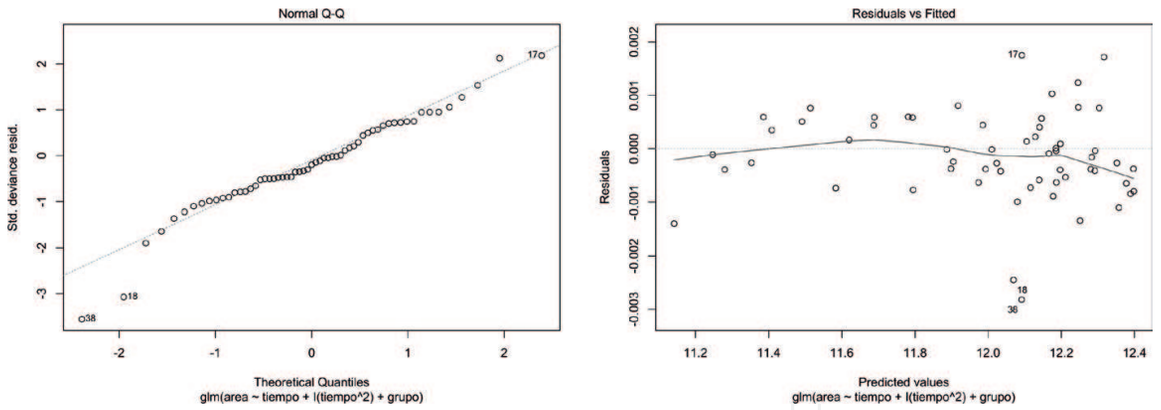

Figure 3. Images of primary control (diagnosis) to the model "glm.Without.Network" of the type I collagen degradation as regards time and cell culture group.

As the obtained p-value (0.0006396) is less than 0.05, we cannot deny that the distribution is normal (**Figure 4**).

Violet tone, as result of haematoxylin and eosin stain, depicts how collagen gradually degrades by time. This colour shows up according to the pixel intensity (from 0 to 255 values) or to the initial amount of collagen. (See **Figure 3**, these images display how the collagen has

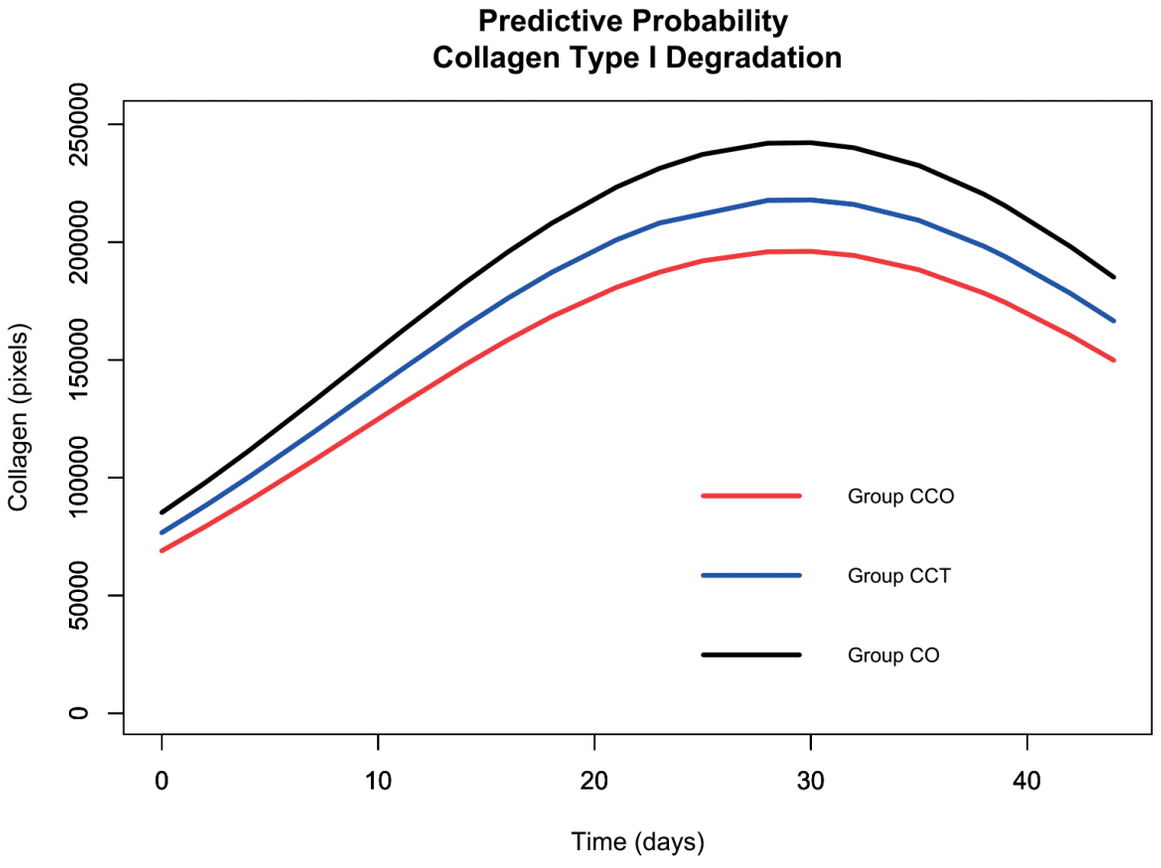

Figure 4. Adjustment of type I collagen degradation according to time and the group through the use of the model "glm.Without.Network" without the neural network. The image displays the prediction of this model for collagen degradation as regards time and according to each group of cell culture (CCO, CCT and CO).

more colour at the beginning and less intensity at the end. Also, the graphics show how the part of the extracellular matrix is more stained than the beginning.) As white is predominant, then the value of pixels is close to 255 (this is the maximum colour for pixels: white). In such manner, this variance of intensity from violet hue to palest hue can be understood as indirect degradation of collagen in pixels as regards time and influenced by it. Then, this figure of the model predictions illustrates, in a particular manner, how this model adjusts better to type I collagen degradation as regards time and the cell culture group.

However, the fall of the curve, in the graphic, points out that after time, the cell activity produces an extracellular matrix (biologic cell activity). This matrix has a colouration darker than the collagen during the degradation process; therefore, it tends to the initial values. To understand this process, see **Figure 5** that depicts the three groups of the sample cells in culture (CCO, CCT and CO) and their colour change. This figure presents the cells at the beginning (T0) and the end (T40) of the experiment and exhibits how the intensity of colour diminishes in the collagen but increases in the extracellular matrix.

To calculate how the probability of collagen degradation changes as regards time and the group, we computed the odds ratio for time (1.074290e+00), the odds ratio for the group (1.111437e+00) and the corresponding intervals of confidence. As the interval of confidence is from 1.051247 to 1.097838e+00 and odds ratio for time is 1.074290e+00, then this value is inside the range. It means if the group variable is included in the model "glm.without.network", then the probabilities of collagen degradation, regarding the time, will increase by 11.6% (0.111).

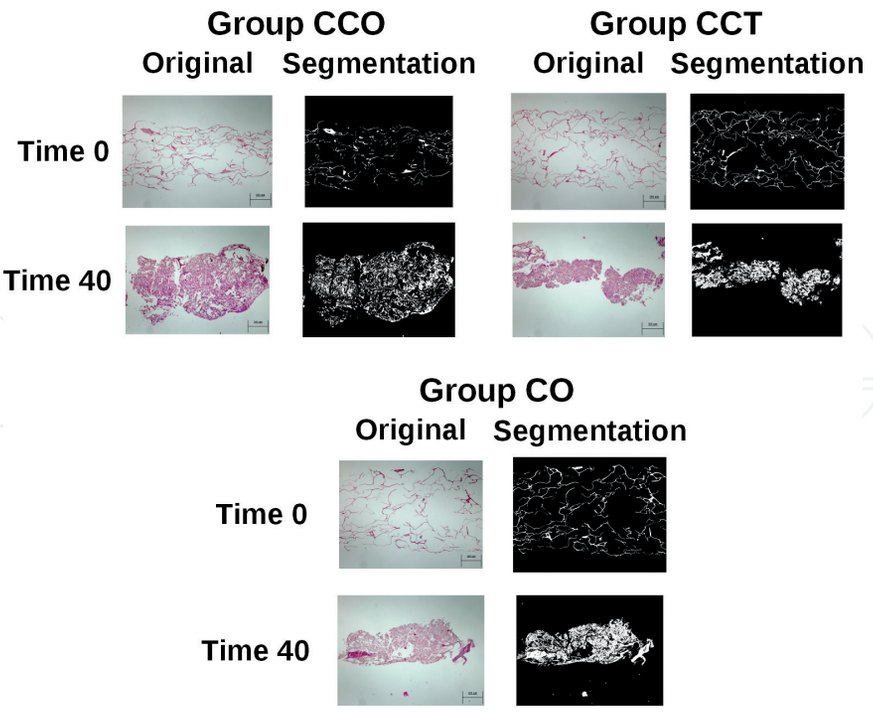

Figure 5. The difference of the colour of type I collagen between time 0 and 40 for the three cell culture groups is shown. Notice how the intensity of haematoxylin-eosin for collagen diminishes, while the tone for extracellular matrix begins to increase. The scale of all images is 300 μm.

Therefore, when we include the group in the model "glm.without.network", the time of collagen degradation is associated with an increase of 11.1% in the mean of probabilities for the collagen degradation.

4.2. Generalised linear models (GLM) with the neural network

GLM modelling: collagen degradation as regards time + a group of cell culture—time effects and cell group in the degradation of type I collagen.

With this model, where $\eta = \mu_x = \beta_0 + \beta_1 x_1 (Time) + \beta_2 x_2 (group)$, μ_x follows a normal distribution. The variance proportion explained in the model (residual deviance) was apparently small (0.00006353), and the AIC was 12,200.

To test $H0{:}\beta_2 = 0$, we use $z = 38.035$ $(p - value = 2e - 16)$. Consequently, the cell culture group, as regards time, seems to have a meaningful impact on the probability of type I collagen degradation after time goes (i.e., once the model includes that variable "glm.with.network"). Namely, that model has the best p-value for time and the group, and the smallest values for AIC and residual deviance compared to the other proposed models.

Then, the following image shows the graphic diagnosis of the model (see **Figure 6**).

The normal QQ-plot shows normalised standard waste. Waste for predicted values is in the left panel. This waste presents a tendency to the mean; therefore, the error independence condition is fulfilled. Thus, collagen degradation (pixels) has a tendency.

The right panel displays how the waste values of the model "glm.with.network" adjust to the regression line of the model, and there are atypical values. Probably, these atypical results, same as the model without red, could be due to the error allowed in this type the experiment. As we explained previously, it is complicated to control shifts that occur due to the intrinsic activity of the cell samples under culture.

To determine the waste normality, we applied the Shapiro–Wilk test, where:

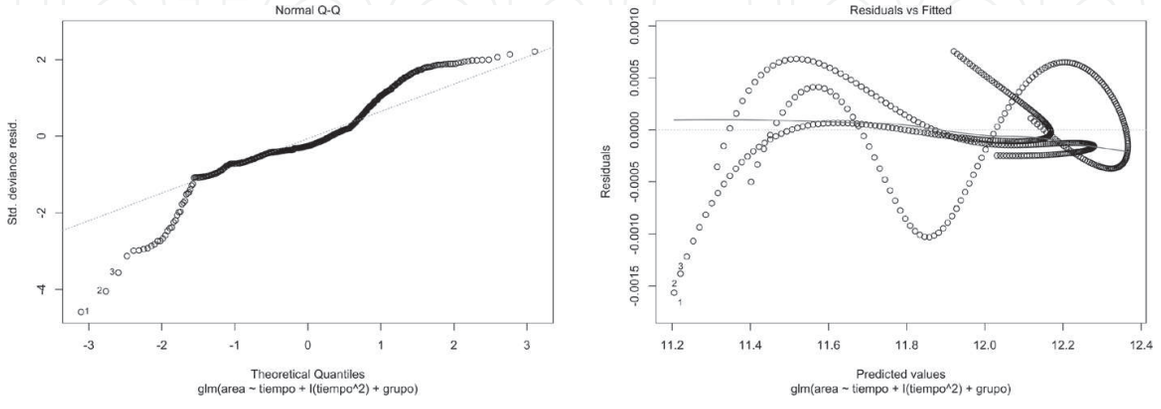

Figure 6. Graphics of primary control (diagnosis) to the model "glm.With.Network" for type I collagen degradation as regards time and cell culture group.

- HO: this sample comes from a normal distribution.
- H1: this sample does not come from a normal distribution.

As the obtained p-value (3.414e-14) is less than 0.05, we cannot deny that the distribution is normal.

Figure 7 shows how the model with the neural network, as **Figure 3** does with the model without the neural network, represents the prediction of collagen degradation as regards time and the group. **Figure 7** depicts a softer behaviour about the recovering of violet tone that belongs to the colour of extracellular matrix that is produced by cell culture in their last days as we explained with **Figure 4**. It means the colour shows up according to the intensity of the values of pixels (0–255) or to the initial value of collagen. Then, the violet tone, as a result of haematoxylin and eosin stain, shows how collagen gradually degrades by time (see **Figure 7**; these images display how the collagen has more colour at the beginning and less intensity at the end. Also, the graphics show how the part of the extracellular matrix is more stained than the beginning). As white is predominant, then the value of pixels is close to 255 (this is the maximum colour for pixels: white). In such manner, this variance of intensity from violet hue to palest hue can be understood as indirect degradation of collagen in pixels as regards time and influenced by it. Then, this figure of the model predictions illustrates, in a particular manner, how this model adjusts better to type I collagen degradation as regards time and the cell culture group.

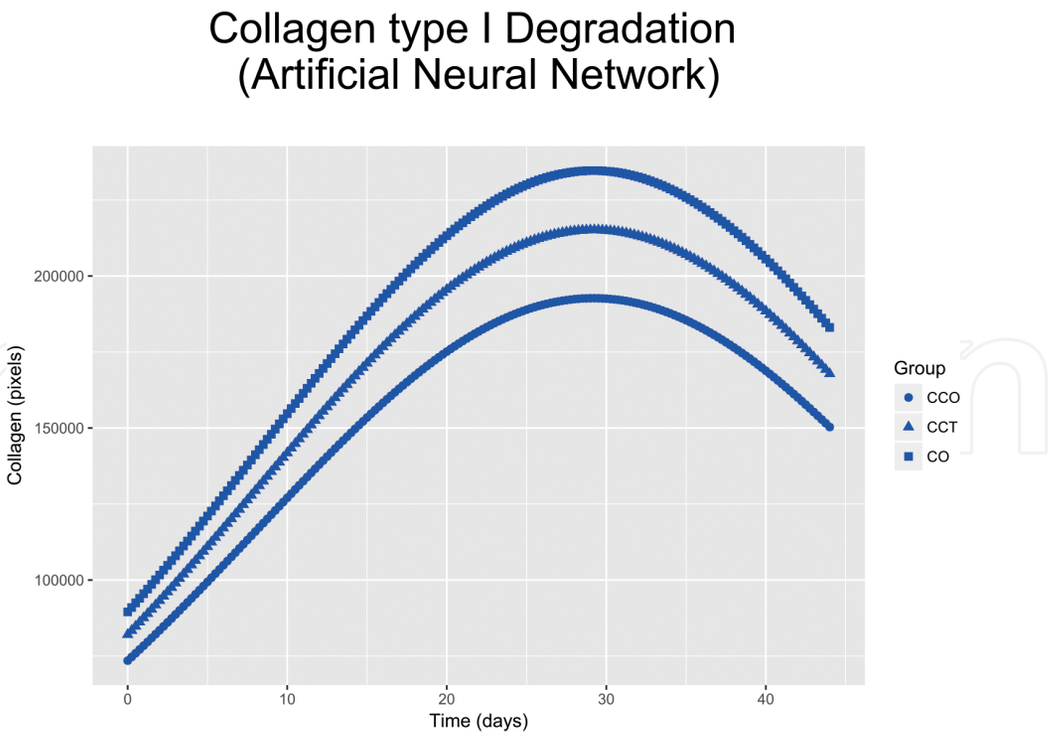

Figure 7. Adjustment of type I collagen degradation according to time and the group through the use of model "glm. With.Network" with the neural network. The image displays the prediction of this model for collagen degradation as regards time and in function of each cell culture group (CCO, CCT and CO).

Nonetheless, as we stated before, the biological behaviour of the cell culture is the same. This conduct means the fallen of the curve, in the graphic, shows up that after time and the cell activity produces an extracellular matrix (biologic cell activity). This matrix has a colouration darker than the collagen during the degradation process; therefore, it tends to the initial values. To understand this process, see **Figure 7** that depicts the three groups of sample cells in culture (CCO, CCT and CO) and their colour change. This figure presents the cells at the beginning (T0) and the end (T40) of the experiment and exhibits how the intensity of colour diminishes in the collagen but increases in the extracellular matrix.

To calculate how the probability of collagen degradation changes as regards time and the group, we computed the odds ratio for time (1.068284e+00), the odds ratio for the group (1.116472e+00) and the corresponding intervals of confidence. As the interval of confidence is from 1.064654e+00 to 1.071926e+00 and odds ratio for time is 1.074290e+00, then this value is inside the range. The odds ratio for the group is 1.116472e+00, value that is also inside the interval of confidence (1.085327e+00, 1.148511e+00). It means if the group variable is included in the model "glm.with.network", then the probabilities of collagen degradation, regarding the time, will increase by 11.6% (0.111).

Hence, when we include the group in the model "glm.with.network", the time of collagen degradation is associated with an increase of 11.6% in the mean of probabilities for the collagen degradation.

4.3. PCA applied to a set of images acquired by means of optical microscopy

In this section, we will describe the strategy followed with the aim of determining the relations among variables or descriptors obtained through optical microscopy from the calibrated images. For our study, we have worked with the following variables (descriptors): area, perimeter, diameter, eccentricity or roundness, mean intensity, centroid (x,y), skew and kurtosis. Each of the aforementioned variables can be related to physical variables in the statistical models. In our case, with the support of the PCA method, we want to determine how the area varies with respect to time, considering that several groups of study with specific characteristics exist.

In order to establish points of comparison for the descriptors, we used on each group of images the machine learning approach described in Section 2. The mathematical backgrounds as well as the details of the method followed (PCA) are described in several researches published in the last years [27–30].

In this analysis, we have established three groups of study, where the collagen is present in each of them. Through the image analysis and machine learning approach are followed, we were able to define regions of interest and extract the corresponding descriptor for each value of time. In this way, the information matrix of each group of images is defined as follows:

a. Cells + collagen + osteogenic culture medium (CCO).

b. Cells + collagen + non-osteogenic medium (CCT).

c. Collagen + osteogenic culture medium, (CO), has no cells (control group).

With the aim of applying the PCA analysis, we started from the hypothesis of time dependence (0–40 days) in which determining the area variation is possible. The area "variation" in which determining aspects such as presence of extracellular matrix due to the collagen degradation, the change of pixel intensity as time goes by or the geometric shape that adopts the group in study according to time is possible is part of the analysis proposal as well as of the statistical model to interpret the images according to the group and time. Likewise, as a complementary part, we propose that groups in the analysis have a relevant weight in the collagen degradation.

The following analysis shows how the area classified according to the groups CO, CCT and CCO changes. We have used labels with the following structure: **XX_T**, where **XX** represents the group and **T** the time. For example, the label **CO_14** is the label for group **CO** and for the day (time variable) **14**.

From the principal components extracted in the CO group, we selected the three most representative that have a cumulative variance of 80.60%. In **Figure 8**, the weights of the components selected are represented, and it is possible to see that the variables of the right side have a positive correlation. However, the variables standard deviation (StdDeev) and Round have a negative weight (negative correlation). In this line, it is possible to establish that a positive correlation between time and the descriptors of the right side of the figure exists. This means that area grows when the variables of the right side grow.

Figure 9 shows a graphic dispersion (biplot) of the data obtained for the groups CO, CCO and CCT in the 40 days of experimentation (two principal components). Likewise, it is possible to see small groups far from the centre.

Figure 8 presents the groups under study and the corresponding descriptors. In the same way, in **Figures 9** and **10**, it is possible to see how the first two components explain the 68.4% of data. The standard deviation is in the region of control groups during the first 3 days. On the other hand, all the variables placed on the right side of the principal component are positively correlated. In the same way, the groups that can be observed are those of control CO_T, with values greater than 30 days, CCO with values greater than 20 days and CCT with values greater than 15 days.

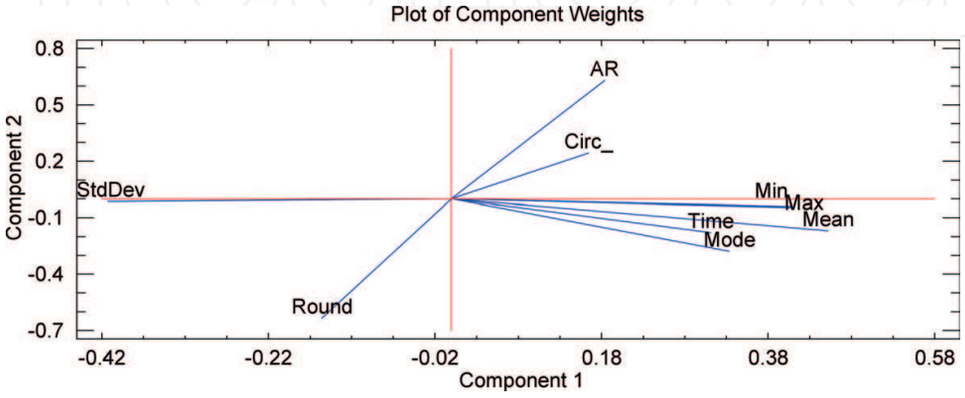

Figure 8. Plot of component weights obtained for CO control group.

http://dx.doi.org/10.5772/intechopen.72982

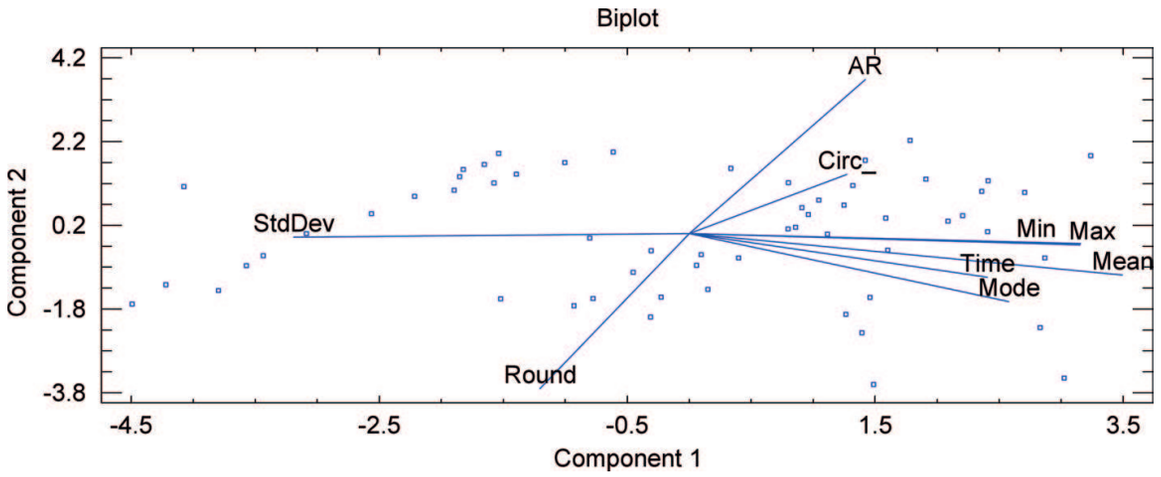

Figure 9. Biplot for descriptors and dispersion values.

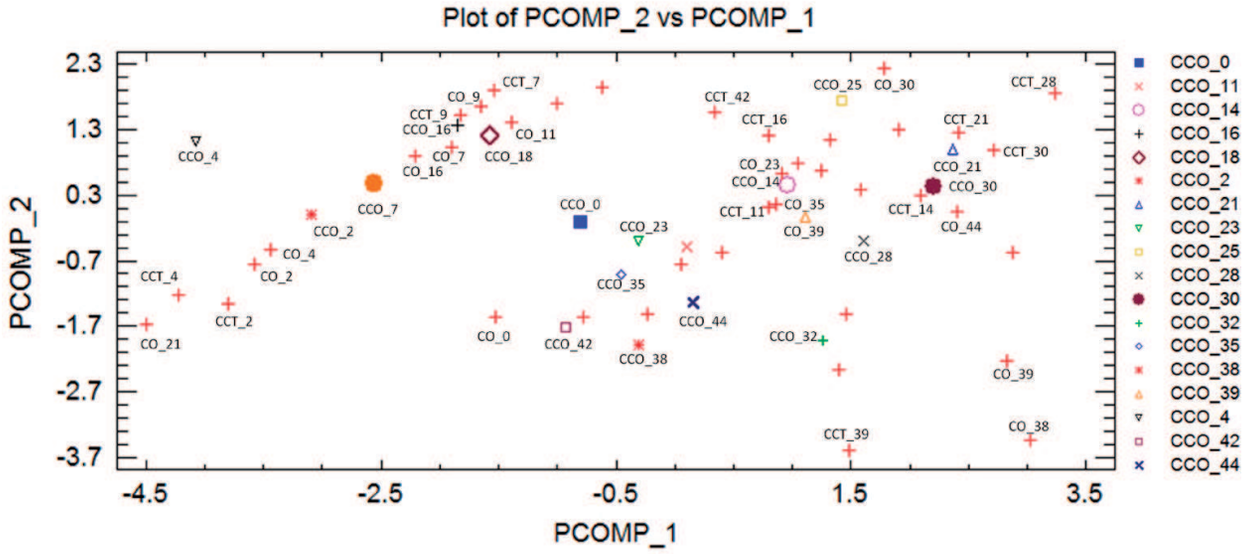

Figure 10. Plot of principal components enhances CO, CCO and CCT groups.

5. Conclusions

GLM models allow greater flexibility in the statistical modelling. Namely, we can observe how neural networks support the improvement of p-values both for group co-variables like for the time. Neural networks also enhance probabilities of collagen degradation (see **Tables 2–5**) as it changes from 11.1% when a neural network is not applied to 11.6%. In the same manner, waste of the model is presumably smaller 4.0879e-05 than 0.00006353 without the neural network. It is necessary to point out that the AIC value is slightly sacrificed when the neural network is applied. However, this loss is part of this implementation; and despite this, in a general view, the use of a neural network has allowed a better adjust of the model.

Variables	IC	
	2.5%	**97.5%**
(Intercept)	47839.657556	8.052878e+04
Time	1.051247	1.097838e+00
Group	1.004596	1.229641e+00

Table 2. Confidence intervals to the parameters of the model "glm.Without.Network".

Coefficients	P-value
(Intercept)	<2e-16***
Time	<2e-16***
Group	1.09e-13*

Table 3. Statistical significance of the time and group co-variables in the results of the model with neural network.

Variables	IC	
	2.5%	**97.5%**
(Intercept)	7.115130e+04	7.591534e+04
Time	1.064654e+00	1.071926e+00
Group	1.085327e+00	1.148511e+00

Table 4. Confidence ranges to the parameters of the model "glm.With.Network".

Coefficients	GLM without neural network		GLM with neural network	
	P-value	**Odds ratio**	**P-value**	**Odds ratio**
(Intercept)	<2e-16	6.206826e+04	<2e-16	7.349473e+04
Time	2.70e-08	1.074290e+00	<2e-16	1.068284e+00
Group	0.0453	1.111437e+00	1.09e-13	1.116472e+00
AIC	0.00006353		4.0879e-05	
Residual deviance	0 12,200		1454.2	

Table 5. Comparison for goodness of fit of parameters between GLM with and without the application of a neural network.

The proposal that was established about the dependence of the area with respect to time, the study group and the descriptors obtained through image analysis will help to establish a mathematical model to explain the variation of the area. In the method used, we observed groupings that allow us to interpret similarities and highly positive or negative correlations with respect to time and the study group. An important aspect to take into account when

deciding what type of neural network and the types of filters to use can both define the thresholds in the images and eliminate information that is not representative.

On the other hand, the LSTM neural network allows predicting a presumptive value of the level of collagen degradation for those instants of time for which information is not available (features, measurements, etc.). All this is feasible since the neural network has a recurrent structure and a short-term memory, whereby can infer better what are the possible values that will have to increase the time. Similarly, we have observed in this work that neural networks of this type can work with a long number of samples to train the system and also support optimal time window size variables.

Acknowledgements

This project was partially funded by the "Ministerio de Educación y Ciencia MTM2014-59543-P"; also by the "Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación del Ecuador (SENESCYT)"; by the "Catedra UNESCO UPS Cuenca-Ecuador, Grupo de Investigación en Inteligencia Artificial y Tecnologías de Asistencia (GIATA) Cuenca-Ecuador"; and by the "Grupo de Investigación en Materiales (GiMaT) Universidad Politéncia Salesiana, Cuenca-Ecuador".

Author details

Yaroslava Robles-Bykbaev[1,2], Salvador Naya[3], Silvia Díaz Prado[1], Daniel Calle-López[4], Vladimir Robles-Bykbaev[4]*, Luis Garzón-Muñóz[5], Clara Sanjurjo Rodríguez[1] and Javier Tarrío Saavedra[3]

*Address all correspondence to: vrobles@ups.edu.ec

1 Grupo de Investigación en Terapia Celular y Medicina Regenerativa (TCMR), Departamento de Medicina, PROTERM, MODES, Universidade da Coruña, España

2 GI-IATa, Universidad Politécnica Salesiana, Cuenca, Ecuador

3 Departamento de Matemáticas, Grupo MODES, Universidad de la Coruña, Ferrol, La Coruña, Spain

4 GI-IATa, Cátedra UNESCO Tecnologías de apoyo para la Inclusión Educativa, Universidad Politécnica Salesiana, Cuenca, Ecuador

5 GI-MAT, Grupo de Investigación en Nuevos Materiales y Procesos de Transformación, Universidad Politécnica Salesiana, Cuenca, Ecuador

References

[1] Santana J. Predicción de series temporales con redes neuronales: una aplicación a la inflación colombiana. Revista Colombiana de Estadística. 2006;**29**(1):77-92

[2] Chen Y, Zhou S, Li Q. Mathematical modeling of degradation for bulk-erosive polymers: Applications in tissue engineering scaffolds and drug delivery systems. Acta Biomaterialia. 2011;**7**(3):1140-1149. DOI: 10.1016/j.actbio.2010.09.038

[3] Hoque ME, Yong LC, Ian P. Mathematical modeling on degradation of 3d tissue engineering scaffold materials. Regenerative Research. 2012;**1**(1):58-59

[4] Pitt CG, Zhong-wei G. Modification of the rates of chain cleavage of poly (ϵ-caprolactone) and related polyesters in the solid state. Journal of Controlled Release. 1987;**4**(4):283-292. DOI: 10.1016/0168-3659(87)90020-4

[5] Sandino C, Planell JA, Lacroix D. A finite element study of mechanical stimuli in scaffolds for bone tissue engineering. Journal of Biomechanics. 2008;**41**(5):1005-1014. DOI: 10.1016/j.jbiomech.2007.12.011

[6] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: Data mining, inference, and prediction. Biometrics. 2002

[7] Bocanegra G, Domínguez J. Modelos lineales generalizadas en el contexto de diseño robusto. In: Instituto Nacional de Estadística, Geografía e Informática, editors. Memorias XX Foro Nacional de Estadística. México:2006

[8] Wood S. Generalized Additive Models: An Introduction with R. CRC Press; 2006

[9] Meijering E. Cell segmentation: 50 years down the road [life sciences]. IEEE Signal Processing Magazine. 2012;**29**(5):140-145. DOI: 10.1109/MSP.2012.2204190

[10] Kasprowicz R, Suman R, O'Toole P. Characterising live cell behaviour: Traditional label-free and quantitative phase imaging approaches. The International Journal of Biochemistry & Cell Biology. 2017;**84**:89-95. DOI: 10.1016/j.biocel.2017.01.004

[11] Alanazi H, Canul AJ, Garman A, Quimby J, Vasdekis AE. Robust microbial cell segmentation by optical-phase thresholding with minimal processing requirements. Cytometry Part A. 2017;**91**(5):443-449. DOI: 10.1002/cyto.a.23099

[12] Grys BT, Lo DS, Sahin N, Kraus OZ, Morris Q, Boone C, Andrews BJ. Machine learning and computer vision approaches for phenotypic profiling. Journal of Cell Biology. 2016. jcb-201610026. DOI: 10.1083/jcb.201610026

[13] Criminisi A, Shotton J, Konukoglu E. Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. Foundations and Trends® in Computer Graphics and Vision. 2012;**7**(2-3):81-227. DOI: 10.1561/0600000035

[14] Budde MD, Frank JA. Examining brain microstructure using structure tensor analysis of histological sections. NeuroImage. 2012;**63**(1):1-10. DOI: 10.1016/j.neuroimage.2012.06.042

[15] Arganda-Carreras I, Kaynig V, Rueden C, Eliceiri KW, Schindelin J, Cardona A, Sebastian Seung H. Trainable Weka Segmentation: A machine learning tool for microscopy pixel classification. Bioinformatics. 2017;btx180. DOI: 10.1093/bioinformatics/btx180

[16] Schindelin J, Rueden CT, Hiner MC, Eliceiri KW. The ImageJ ecosystem: An open platform for biomedical image analysis. Molecular Reproduction and Development. 2015;**82**(7-8):518-529. DOI: 10.1002/mrd.22489

[17] Sobel I. An isotropic 3×3 image gradient operator. Machine vision for three-dimensional scenes. 1990:376-379

[18] Ma X, Tao Z, Wang Y, Yu H, Wang Y. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transportation Research Part C: Emerging Technologies. 2015;**54**:187-197. DOI: 10.1016/j.trc.2015.03.014

[19] Gers F. Long Short-Term Memory in Recurrent Neural Networks. Unpublished PhD dissertation. Lausanne, Switzerland: Ecole Polytechnique Fédérale de Lausanne; 2001

[20] Cao J, Ng ES, McNaughton D, Stanley EG, Elefanty AG, Tobin MJ, Heraud P. The characterisation of pluripotent and multipotent stem cells using Fourier transform infrared microspectroscopy. International Journal of Molecular Sciences. 2013;**14**(9):17453-17456. DOI: 10.3390/ijms140917453

[21] Aida S, Matsuno T, Hasegawa T, Tsuji K. Application of principal component analysis for improvement of X-ray fluorescence images obtained by polycapillary-based micro-XRF technique. Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms. 2017. DOI: 10.1016/j.nimb.2017.03.123

[22] Egan CK, Jacques SDM, Cernik RJ. Multivariate analysis of hyperspectral hard X-ray images. X-Ray Spectrometry. 2013;**42**(3):151-157. DOI: 10.1002/xrs.2448

[23] Gan L, Lv W, Zhang X, Meng X. Improved PCA+ LDA applies to gastric cancer image classification process. Physics Procedia. 2012;**24**:1689-1695. DOI: 10.1016/j.phpro.2012.02.249

[24] Juma K, He M, Zhao Y. Lung cancer detection and analysis using data mining techniques, principal component analysis and artificial neural network. American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS). 2016;**26**(3):254-265

[25] Smith SM, Hyvärinen A, Varoquaux G, Miller KL, Beckmann CF. Group-PCA for very large fMRI datasets. NeuroImage. 2014;**101**:738-749. DOI: 10.1016/j.neuroimage.2014.07.051

[26] Ballester A, Sueiro-Fernández J, editors. Biomateriales y Sustitutos Óseos en Traumatología y Cirugía Ortopédica. 1st ed. Cádiz: Universidad de Cádiz; 2011

[27] Hervé A, Williams LJ. Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics. 2010;**2**(4):433-459

[28] Ghosh A, Barman S. Application of Euclidean distance measurement and principal component analysis for gene identification. Gene. 2016;**583**:112-120

[29] Godoy JL, Vega JR, Marchetti JL. Relationships between PCA and PLS-regression. Chemometrics and Intelligent Laboratory Systems. 2014;**130**:182-191

[30] Han Y, Feng X-C, Baciu G. Variational and PCA based natural image segmentation. Pattern Recognition. 2013;**46**:1971-1984

IntechOpen