TechRxiv
Powered by IEEE

Quantitative ultrasound classification of healthy and chemically degraded ex-vivo cartilage

This paper was downloaded from TechRxiv (https://www.techrxiv.org).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

05-12-2023 / 07-12-2023

CITATION

Sorriento, Angela (2023). Quantitative ultrasound classification of healthy and chemically degraded ex-vivo cartilage. TechRxiv. Preprint. https://doi.org/10.36227/techrxiv.24747252.v1

DOI

10.36227/techrxiv.24747252.v1

SUPPLEMENTARY MATERIAL

Quantitative ultrasound classification of healthy and chemically degraded ex-vivo cartilage

Angela Sorriento[1,2], Lorena Guachi-Guachi[1,2], Enrico Lenzi [3] , Paolo Dolzani [3], Gina Lisignoli [3], Sajedeh Kerdegari[4,5], Gaetano Valenza[6], Claudio Canale[4], Andrea Cafarelli[1,2], Leonardo Ricotti[1,2]

[1] The BioRobotics Institute, Scuola Superiore Sant'Anna, 56127 Pisa, Italy. [2] Department of Excellence in Robotics & AI, Scuola Superiore Sant'Anna, 56127 Pisa, Italy. [3] IRCCS Istituto Ortopedico Rizzoli, SC Laboratorio di Immunoreumatologia e Rigenerazione Tissutale, 40136 Bologna, Italy. [4]Physics Department, University of Genoa, via Dodecaneso 33, 16146 Genoa, Italy. [5]Nanoscopy, Istituto Italiano di Tecnologia, Via Enrico Melen, 83 Edificio B, 16152 Genova, Italy [6]Bioengineerring and Robotics Research Centre E Piaggio, University of Pisa, 56122 Pisa, Italy.[6]Department of Information Engineering, University of Pisa, 56123 Pisa, Italy.

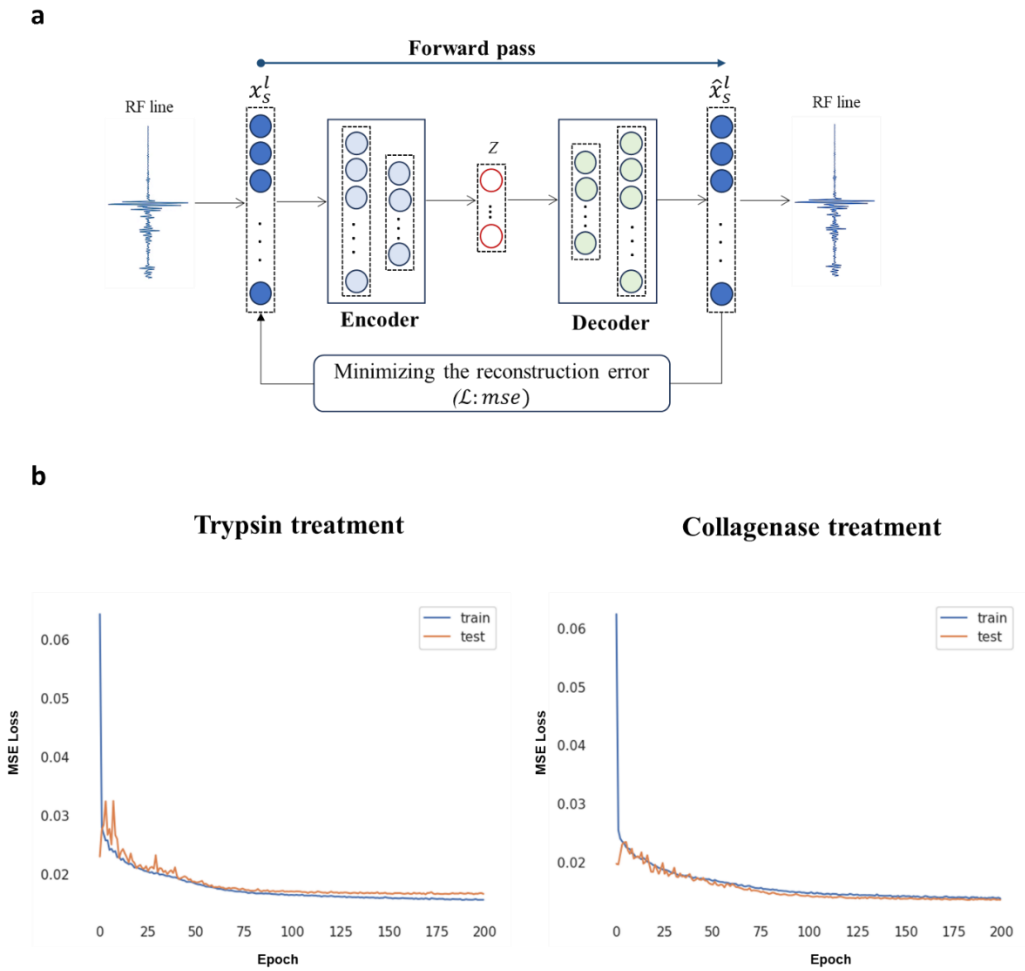

Figure S 1. a) Architecture of the autoencoder network used in our study to extract five compressed features during the forward pass. This network was trained to take an RF line as the input, and its reconstructed RF line while minimizing the reconstruction error, as the output. b) Learning curves of the autoencoder networks showing the convergence of both training and testing sets. For the trypsin treatment group, the convergence value were 0.0156 and 0.0166 at epoch 200 for training and testing sets, respectively. For the collagenase treatment the convergence values were 0.0137 and 0.135 at epoch 200 for training and testing sets, respectively.

	Trypsin treatment			Collagenase treatment		
	0h	**2h**	**4h**	**0h**	**6h**	**24h**
ApEn	0.1716 ± 0.1378	0.1746 ± 0.1253	0.1416 ± 0.0878	0.1604 ± 0.0990	0.1333 ± 0.0677	0.1282 ± 0.0604
SampEn	0.0572 ± 0.0644	0.0580 ± 0.0478	0.0457 ± 0.0347	0.0525 ± 0.0400	0.0416 ± 0.0281	0.0360 ± 0.0235
Spectral Entropy	0.7283 ± 0.0500	0.7292 ± 0.0477	0.7344 ± 0.0490	0.7302 ± 0.0430	0.7323 ± 0.0530	0.7335 ± 0.0431
Mean crossing	137 ± 12	132 ± 10	138 ± 13	133 ± 9	135 ± 10	136 ± 7
Katz fractal dimension	2.0074 ± 0.3799	1.9906 ± 0.3597	1.9347 ± 0.3320	2.0036 ± 0.2961	1.9453 ± 0.2200	1.9537 ± 0.2779
50th percentile	-0.1136 ± 0.1021	-0.1070 ± 0.1012	-0.0859 ± 0.0885	-0.1065 ± 0.0747	-0.0898 ± 0.0566	-0.0939 ± 0.0665
Std	0.1364 ± 0.0380	0.1338 ± 0.0353	0.1292 ± 0.0337	0.1344 ± 0.0293	0.1330 ± 0.0295	0.1308 ± 0.0318
Kurtosis	23.4112 ± 17.9032	25.5768 ± 17.5030	26.8749 ± 19.0188	24.3362 ± 15.0227	26.6827 ± 15.4959	26.1147 ± 16.8780
RI_c	323.5 ± 741.0	236.5 ± 634.5	397 ± 496	164.5 ± 438.0	39± 49	-
RI_b	1610.5 ± 1704.5	1342.0 ± 1035.5	2089.5 ± 1716.5	1933.5 ± 1527.0	2047.5 ± 1286.0	2050.5 ± 2093.0
Bone propagation	49 ± 19	45 ± 19	49 ± 17	55 ± 29	86 ± 22	95 ± 11
Cartilage length	80 ± 20	80 ± 21	75 ± 12	99 ± 37	76 ± 30	29 ± 15
Feature F1	-0.9721 ± 5.9951	-0.6431 ± 6.6928	-1.0999 ± 5.4171	2.0019 ± 7.4839	1.7682 ± 7.2840	1.8143 ± 8.1127
Feature F2	0.1561 ± 4.7435	0.0077 ± 7.6535	0.5895 ± 6.9865	1.8887 ± 5.7114	1.9513 ± 7.5532	1.5234 ± 9.6724
Feature F3	-2.4669 ± 5.4874	-1.5755 ± 6.7212	-2.2016 ± 6.5942	-1.6363 ± 8.3789	-0.1226 ± 8.0785	-1.5679 ± 9.8474
Feature F4	2.3230 ± 6.6958	2.0542 ± 7.4033	2.3259 ± 5.6944	0.5599 ± 5.6037	0.0059 ± 7.2781	-1.4252 ± 7.1011
Feature F5	-0.6680 ± 6.5286	-0.9382 ± 7.1222	-2.0297 ± 6.3026	-0.9013 ± 8.8102	-2.3392 ± 10.2088	-1.1991 ± 9.3578

Table S 1 - Median ± interquartile range of QUS features for trypsin and collagenase treatments.

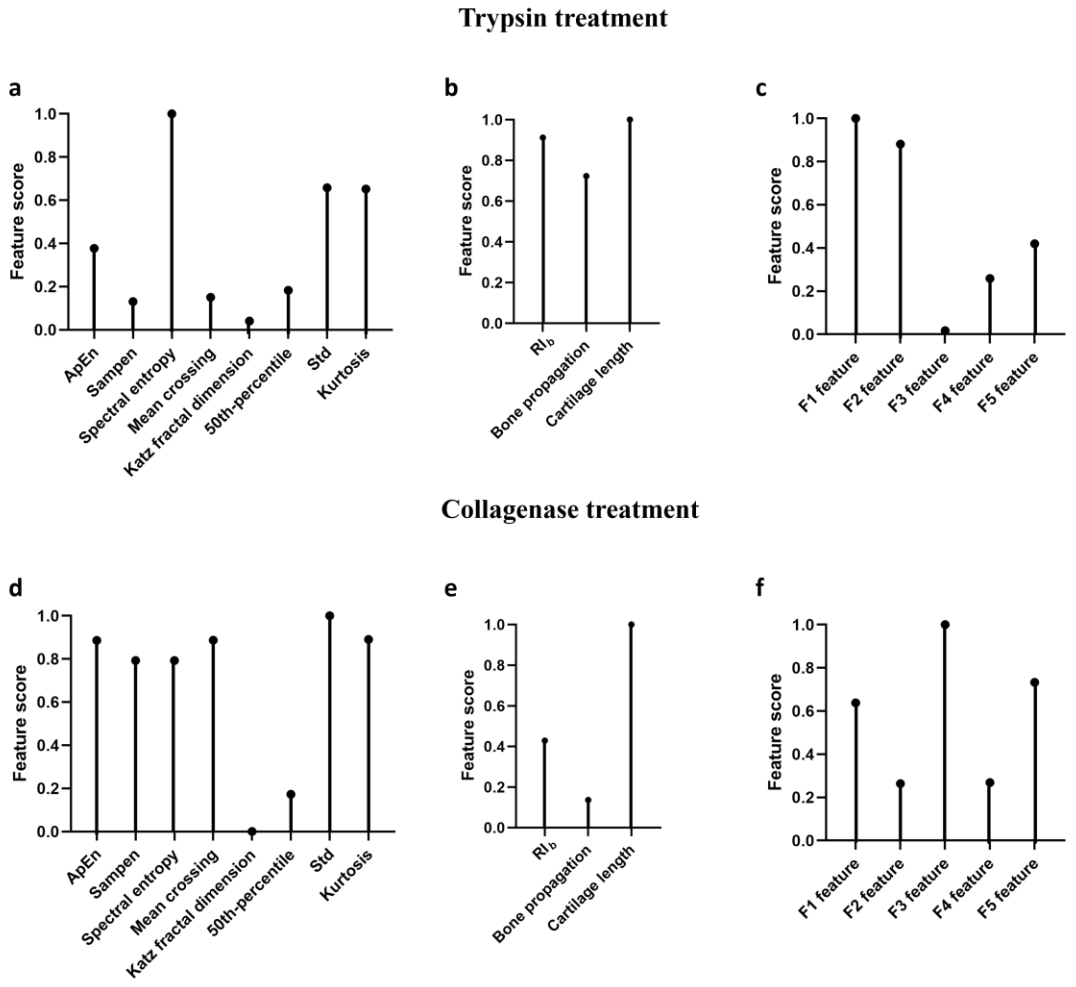

Figure S 2 - Feature importance scores for each metric to distinguish healthy from degenerated cartilage subjected to trypsin and collagenase treatment. Dataset of complexity and irregularity metrics for trypsin (a) and collagenase (d), dataset of cartilage features for trypsin (b) and collagenase (e), dataset of compressed features for trypsin (c) and collagenase (f). Spectral entropy, cartilage length, and Feature F1 significantly contributed to the discrimination between healthy and degraded samples treated with trypsin. Std, cartilage length, and Feature F3 were key contributors to the discrimination between healthy and degraded samples treated with collagenase.

Classifier	Precision [%]		Recall [%]		F1-score [%]		Accuracy [%]	AUC [%]
	Healthy (0h)	Degenerated (2h and 4h)	Healthy (0h)	Degenerated (2h and 4h)	Healthy (0h)	Degenerated (2h and 4h)		
Decision tree	25.00	63.63	20.00	70.00	22.22	66.67	53.33	45.00
XGB Classifier	42.86	75.00	60.00	60.00	50.00	66.67	60.00	60.00
SVM	100.00	71.43	20.00	100.00	33.33	83.33	73.33	60.00
Random forest	66.67	75.00	40.00	90.00	50.00	81.82	73.33	65.00
Logistic regression	100.00	76.92	40.00	100.00	57.14	86.96	80.00	70.00
Ensemble	75.00	81.82	60.00	90.00	66.67	85.71	80.00	75.00

Table S 2 - Overall performance for classification models to distinguish healthy and degenerated samples treated by trypsin. All the models were trained 10 times on the $Training_{MLmodel}$ dataset and then tested on the $Test_{MLmodel}$ dataset to distinguish two and three classes, as their results are shown in Table S 4 and Table S 5.

Classifier	Precision [%]		Recall [%]		F1-score [%]		Accuracy [%]	AUC [%]
	Healthy (0h)	Degenerated (6h and 24h)	Healthy (0h)	Degenerated (6h and 24h)	Healthy (0h)	Degenerated (6h and 24h)		
Decision tree	100.00	76.92	40.00	100.00	57.14	86.96	80.00	70.00
XGB Classifier	75.00	82.82	60.00	90.00	66.67	85.71	80.00	75.00
SVM	0.00	66.67	0.00	100.00	0.00	80.00	66.67	50.00
Random forest	100.00	76.92	40.00	100.00	57.14	86.96	80.00	70.00
Logistic regression	100.00	90.91	80.00	100.00	88.89	95.24	93.33	90.00

Table S 3 - Overall performance for classification models to distinguish healthy and degenerated samples treated by collagenase . All the models were trained 10 times on the $Training_{MLmodel}$ dataset and then tested on the $Test_{MLmodel}$ dataset to distinguish two and three classes, as their results are shown in Table S 6 and Table S 7.

Classifier	Precision [%]	Recall [%]	F1-score [%]	Accuracy [%]
Decision tree	50.90 ± 0.20	50.70 ± 3.30	50.40 ± 1.80	50.70 ± 3.30
XGB Classifier	64.30 ± 0.00	60.00 ± 0.00	61.10 ± 0.00	60.00 ± 0.00
SVM	81.00 ± 0.00	73.30 ± 0.00	66.70 ± 0.00	73.30 ± 0.00
Random forest	49.80 ± 9.70	56.70 ± 9.10	52.40 ± 8.40	56.70 ± 9.10
Logistic regression	84.60 ± 0.00	80.00 ± 0.00	77.00 ± 0.00	80.00 ± 0.00
Ensemble	72.60 ± 11.50	72.00 ± 11.90	72.00 ± 11.60	72.00 ± 11.90

Table S 4 - Mean ± standard deviation of the overall performance for classification methods to distinguish two classes: control (0h) from degenerated cartilage samples by trypsin treatment (2h-4h). All classification methods were trained and evaluated 10 times.

Classifier	Precision [%]	Recall [%]	F1-score [%]	Accuracy [%]
Decision tree	18.90 ± 6.80	19.30 ± 6.30	19.00 ± 6.40	19.30 ± 6.30
XGB Classifier	28.30 ± 0.00	33.30 ± 0.00	30.30 ± 0.00	33.30 ± 0.00
SVM	11.10 ± 0.00	33.30 ± 0.00	16.70 ± 0.00	33.30 ± 0.00
Random forest	37.20 ± 12.90	4.0 ± 7.40	36.90 ± 7.70	44.00 ± 7.40
Logistic regression	68.10 ± 0.00	53.30 ± 0.00	50.20 ± 0.00	53.30 ± 0.00
Ensemble	66.50 ± 2.60	57.30 ± 3.30	55.90 ± 2.50	57.30 ± 3.33

Table S 5 - Mean ± standard deviation of the overall performance for classification methods to distinguish three classes: control (0h), degenerated at 2h, and degenerated at 4h. All classification methods were trained and evaluated 10 times.

Classifier	Precision [%]	Recall [%]	F1-score [%]	Accuracy [%]
Decision tree	84.60 ± 0.00	80.00 ± 0.00	77.00 ± 0.00	80.00 ± 0.00
XGB Classifier	79.50 ± 0.00	80.00 ± 0.00	79.36 ± 0.00	80.00 ± 0.00
SVM	44.40 ± 0.00	66.70 ± 0.00	53.30 ± 0.00	66.70 ± 0.00
Random forest	70.80 ± 5.00	72.00 ± 15.20	65.10 ± 5.00	72.00 ± 5.00
Logistic regression	93.90 ± 0.00	93.30 ± 0.00	93.10 ± 0.00	93.30 ± 0.00

Table S 6 - Mean ± standard deviation of the overall performance for classification methods to distinguish two classes: control (0h) from degenerated cartilage samples by collagenase treatment (6h-24h). All classification methods were trained and evaluated 10 times.

Classifier	Precision [%]	Recall [%]	F1-score [%]	Accuracy [%]
Decision tree	66.70 ± 7.50	59.30 ± 7.00	57.90 ± 6.90	59.30 ± 7.00
XGB Classifier	50.00 ± 0.00	53.30 ± 0.00	47.20 ± 0.00	53.30 ± 0.00
SVM	11.10 ± 0.00	33.30 ± 0.00	16.70 ± 0.00	33.30 ± 0.00
Random forest	62.40 ± 12.60	58.70 ± 7.20	55.30 ± 9.60	58.70 ± 7.20
Logistic regression	79.20 ± 0.00	73.30 ± 0.00	72.90 ± 0.00	73.30 ± 0.00
Ensemble	88.80 ± 1.40	86.00 ± 2.00	84.90 ± 2.60	86.00 ± 2.00

Table S 7 - Mean ± standard deviation of the overall performance for classification methods to distinguish three classes: control (0h), degenerated at 6h, and degenerated at 24h. All classification methods were trained and evaluated 10 times.

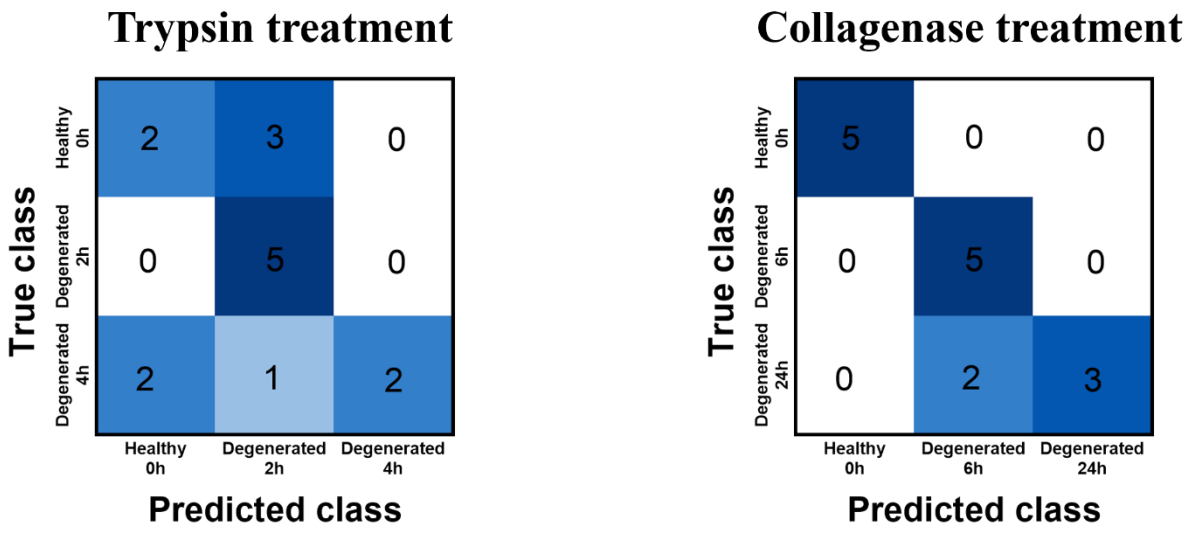

Figure S 3 - Confusion matrices relative to the ensemble methods for trypsin and collagenase treatments at different time points.

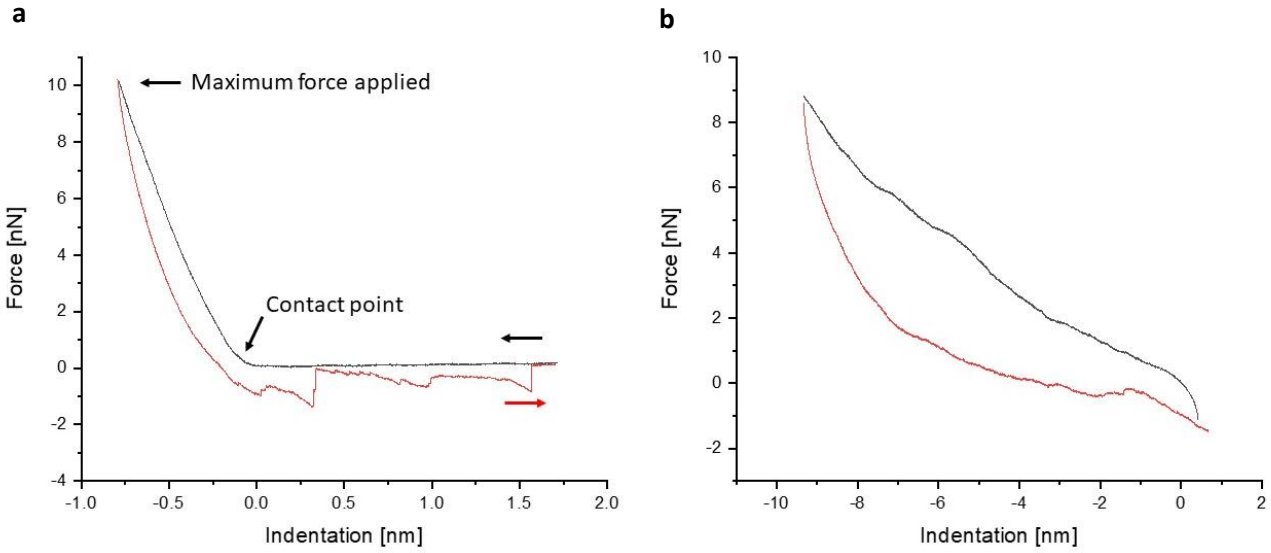

Figure S 4. A typical F-D curve acquired on untreated cartilage (a). The black line is the approaching part of the F-D curve-cycle, where the distance between the AFM tip is decreasing until the contact point. The initial part of the curve, conventionally positioned on the right in the graphical representation, is characterized by a horizontal baseline; here the tip-sample distance is large enough to exclude the presence of any interaction between probe and sample. After reaching the maximum force applied, the tip is retracted to the initial position. The definition of a contact point is fundamental for the application of the Hertz-Sneddon's model. Although applying the maximum displacement range allowed by the AFM system, the tip cannot detach from the surface of most of the collagenase treated samples, as demonstrated by the absence of the horizontal le baseline in the F-D curve of the panel b. The sticky nature of the collagenase-treated cartilages precluded the calculation of the elastic modulus in several samples. For this reason, the Young's modulus of collagenase-treated samples here indicated can be considered as representative of a subpopulation of the samples. However, the experiments demonstrated the extreme softening effect of collagenase-treatment.

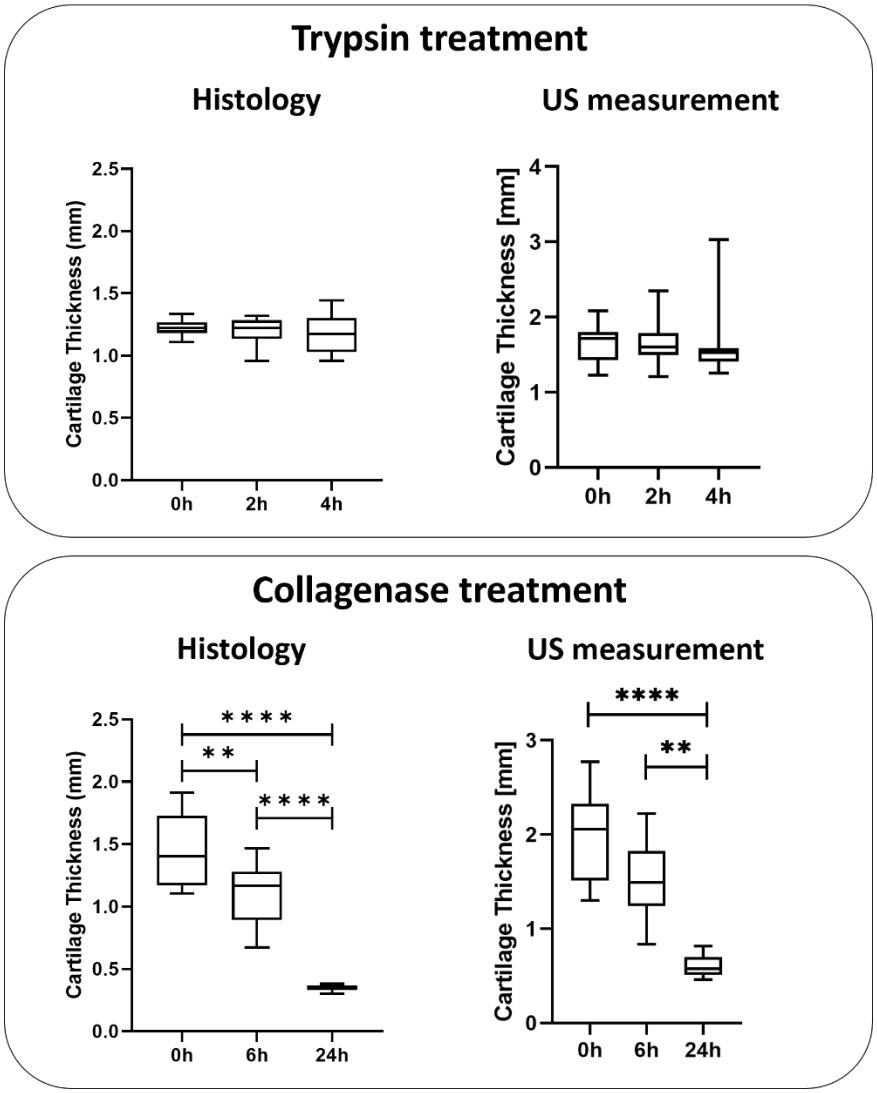

Figure S 5 – Thickness measurement derived from histology and ultrasound analyses for trypsin and collagenase treatment. An average thickness value was calculated for each sample by averaging along the RF lines. Trypsin treatment did not affect the thickness of the samples, while significant differences in the thickness were found after collagenase treatment in both histological and ultrasound analyses. Kruskal-Wallis with Dunn's multiple comparisons test was used for statistical analysis: **** $p < 0.0001$.

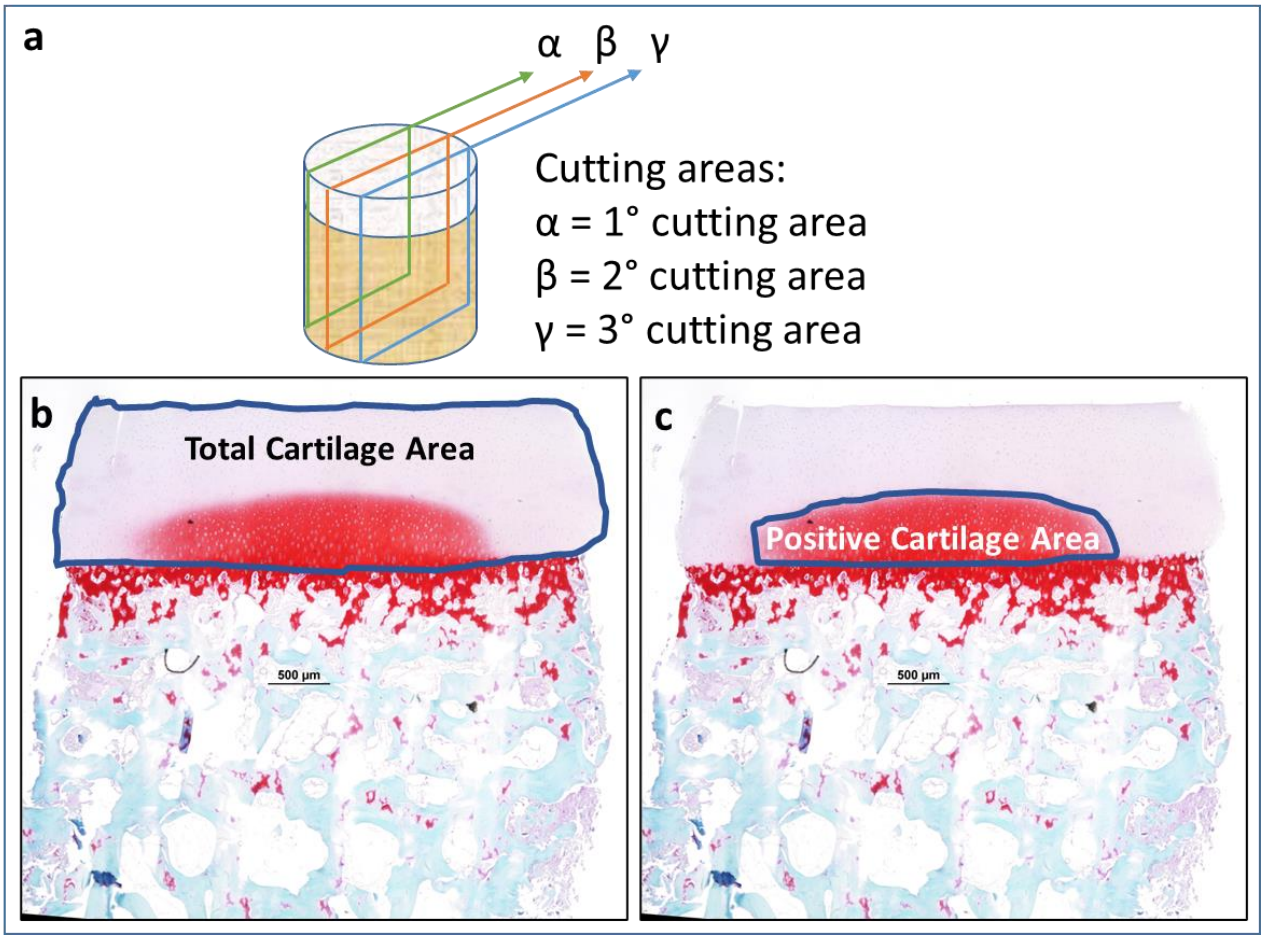

Figure S6: Scheme of the three sections of the osteochondral plug that were cut: α, β and γ (a). Images of Safranin O - Fast Green staining indicating both the total (b) and the positive cartilage (c) areas selected for quantification.

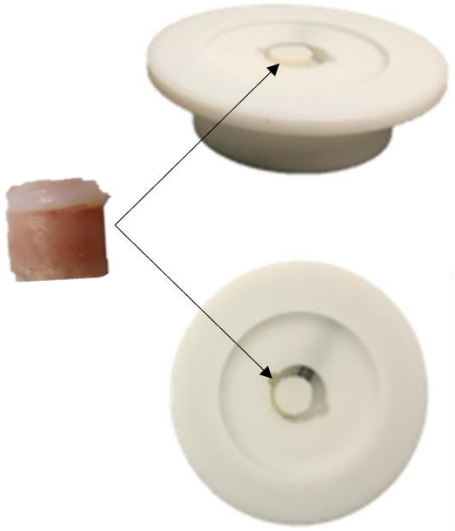

Figure S 7 - As the samples were acquired by punching bovine articular cartilages, they are relatively of a big size for AFM (diameter: 6mm, length: 5/6 mm). Consequently, we designed a custom sample holder made of Teflon to accommodate this size. This holder serves a dual purpose: it allows for securing the sample with a bottom screw and enables the adjustment of the sample's angle to avoid tilting of the sample surface, unsuitable for AFM applications. Furthermore, it features a container to house the liquid buffer and enzymes while the AFM probe can access the chamber.

Appendix Section

a) Quantitative ultrasound parameters

Complexity and irregularity

1. *Approximate entropy (ApEn):* it measures correlation, symmetry, uncertainty, and complexity in a series. Smaller values indicate regular, predictable patterns, while larger values indicate irregularity and unpredictability. *ApEn* is defined by (1):

$$ApEn\ (m, r, N) = \varphi^m(r) - \varphi^{m+1}(r) \quad (1)$$

where N is the time series signal, m is the embedding dimension, r is the tolerance for accepting changes, $\varphi^m(r) = \frac{1}{N-m+1}\sum_{i=1}^{N-m+1} \ln \mathrm{C}_i^m(i)$, and $\mathrm{C}_i^m = \frac{N^m(\mathrm{i})}{N-m+1}$; $1 \le i \le N-m+1$

2. *Sample entropy (Sampen):* it quantifies the negative conditional probability that two similar sequences of m points (A) remain similar at the next point m+1 (B). *Sampen* describes the complexity of time series, and it often is used for diagnosing diseased states in physiological time series. It is defined by (2):

$$Sampen\ (m, r, N) = -\log\frac{A}{B} \quad (2)$$

where N is the length of the original series, m is the length of the sequences to be compared, r is the tolerance value to accept matches.

3. *Spectral entropy:* it characterizes the disturbance present in a signal in the frequency domain. Higher values denote more uniform power spectral distribution, and lower values denote low level of uniformity of the data points. It is defined by (3):

$$Spectral\ entropy = -\sum_{r=1}^{N} P_r \ln P_r \quad (3)$$

where N refers to the length of data, P_r is the proportion of r-th spectrum energy in all spectrum energy, and it can be denoted also as the proportion probability $P_r = \frac{S_r}{\sum_{r=1}^{N} S_r}$. S_r (r = 1, 2, . . . , N) can be seen as energy partition of a time series in the frequency domain obtained by Fast Fourier Transform.

4. *Mean crossing:* it counts the number of times a series crosses its mean value.

5. *Katz fractal dimension:* it is an statistical index of complexity detail in a pattern of a series. It is defined by (4):

$$Katz\ fractal\ dimension = \frac{\log(N)}{\log(N) + \log(d/L)} \quad (4)$$

where N denotes the total length of a series, d is the maximum distance between the initial point to the other points, and L is the length of the waveform. They can be defined as $d = \max(x_i - x_1)$ and $L = \sum_{i=0}^{N-2}(x_{i+1} - x_i)^2$.

6. *50-th percentile:* it defines the value located exactly in the middle of a series. This metric provides information about the central tendency of the data distribution.

7. *Standard deviation(Std):* it measures the distribution of samples around the mean ($\bar{X}$) of a series. Low and high value of std denotes less and more spread of samples around $\bar{X}$ of data, respectively. Std is defined by (5):

$$Std = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{X})^2} \quad (5)$$

where N is the length of the time series signal, x_i is the amplitude value at sampling point i, $\bar{X}$ is the mean value of the series.

8. *Kurtosis:* it measures data concentration in a distribution. Higher Kurtosis denotes a sharper peak and more concentrated values, while lower Kurtosis denotes a flatter distribution with dispersed data points. It helps understand data shape and their characteristics. Kurtosis is defined by (6):

$$Kurtosis = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \bar{X}}{\sigma}\right)^4 \quad (6)$$

where N is the length of the time series signal, x_i is the amplitude value at sampling point i, $\bar{X}$ is the mean value of the series, while σ is its standard deviation of the series.

For both the *ApEn* and *Sampen* metrics, the parameters *m* and *r* were set to fixed values of 2 and 0.2 times the standard deviation of the series, respectively. These parameter values have been suggested for cases of N ≥ 100 [S1].

Cartilage structure

9. *Reflection index (RI).* RI is calculated at the water-cartilage interface (RI_c) and at the cartilage-bone interface (RI_b), as described in the equation (7):

$$RI = \max(RF) - \min(RF) \quad (7)$$

where *RF* indicates the magnitude of the RF signal for each scanning line. The index of the first peak in the RF signal (ind_{peak_start}) was identified on the squared RF signal, using a threshold method with respect to the baseline (*i.e.*, the signal from water). Then, the peak-to-peak value RI_c

was calculated by identifying the maximum and minimum peak value in a window of 20 samples (< 1 mm) starting from ind_{peak_start}; while RI_b was measured by identifying the maximum and minimum peak value in a window of 100 samples (around 2 mm) taking the end of the cartilage interface (RI_c + 30 samples) as a starting point.

10. *Bone propagation.* This metric calculates how much the signal spreads after the second reflection at the bone surface. It is calculated as difference between the index of the last peak in the signal (ind_{peak_end}) and the index of the maximum peak in the signal (ind_{RI_b}), as described in (8):

$$Bone\ propagation = ind_{peak\_end} - ind_{RI_b} \quad (8)$$

11. *Cartilage length.* The cartilage length was calculated in terms of difference between the index of the maximum peak (ind_{RI_b}) and the index of the first peak in the signal (ind_{peak_start}), identified as described in (9).

$$Cartilage\ length = ind_{RI_b} - ind_{peak\_start} \quad (9)$$

12. *Cartilage thickness.* The cartilage thickness was calculated as described in (10).

$$Cartilage\ thickness = c_c \times \frac{TOF}{2} \quad (10)$$

where TOF is the time of flight calculated from the cartilage top surface to the bone surface and c_c is the speed of sound of the cartilage (1610 m/s for the healthy cartilage, 1595 m/s after the trypsin treatment and 1580 m/s after the collagenase treatment [S2]). The TOF was calculated by dividing the *cartilage length* by the sampling frequency of the system (40 MHz).

***Compressed features**: F1, F2, F3, F4 and F5*

These features were extracted by encoding the data from each RF line using the encoder layers of the autoencoder network (see Figure S2a).

To encode an input vector $x \equiv x_s^l \in \mathbb{R}^{300}$, the encoder was mapped linearly to the input with a set of weights $W^1_{\text{encoder}} \in \mathbb{R}^{K_1}$ with K_1 units. Then, a bias vector $b^1_{\text{encoder}} \in \mathbb{R}^{K_1}$ was added and a nonlinear activation function f_{encoder} was applied to produce the outputs of the first layer $h_1 = f_{\text{encoder}}(W^1_{\text{encoder}} \cdot x + b^1_{\text{encoder}}) \in \mathbb{R}^{K_1}$. This output was used to compute the output of the next layer as $h_2 = f_{\text{encoder}}(W^2_{\text{encoder}} \cdot h_1 + b^2_{\text{encoder}}) \in \mathbb{R}^{K_2}$. Then, the compressed features (final representation of the encoder layers) $z = f_{\text{encoder}}(W^3_{\text{encoder}} \cdot h_2 + b^3_{\text{encoder}}) \in \mathbb{R}^{K_3}$ was computed. In our study, z represents the feature vector containing the compressed features F1, F2, F3, F4 and F5.

During the training phase, the input was reconstructed using the decoder layers. To obtain the reconstructed input $\hat{x} \equiv \hat{x}_s^l \in \mathbb{R}$, the decoder transformed the encoded representation z using another set of weights $W^3_{\text{decoder}} \in \mathbb{R}^{K_3}$ as $\hat{h}_2 = f_{\text{decoder}}(W^3_{\text{decoder}} \cdot z + b^3_{\text{decoder}}) \in \mathbb{R}^{K_2}$ until obtaining the final reconstruction $\hat{x} = f_{\text{decoder}}(W^1_{\text{decoder}} \cdot \hat{h}_1 + b^1_{\text{decoder}}) \in \mathbb{R}^{300}$.

Here, W^{l}_{decoder} and b^{l}_{decoder} denoted the weights matrix and the bias for decoder layer l, and f_{decoder} was the activation function for the decoder. We set K_1, K_2, and K_3 to 600, 300, and 5 respectively. Both the f_{encoder} and f_{decoder} were configured as Leaky ReLU [S3], and they followed a batch normalization layer.

Using the constructed training datasets $\{\boldsymbol{X_s^*}\}_{n=1}^{N_{\text{TrypsinTraining}} \times 30}$ and $\{\boldsymbol{X_s^*}\}_{n=1}^{N_{\text{CollagenaseTraining}} \times 30}$, both reconstruction functions $f_{trypsin}: x^* \rightarrow \hat{x}^*$ and $f_{collagenase}: x^* \rightarrow \hat{x}^*$ were learned separately by minimizing the Mean Square Error between the input data x^* and its decoded counterpart $\hat{x}^*$. Where $N_{\text{TrypsinTraining}} = 33$ and $N_{\text{CollagenaseTraining}} = 27$, denoted the number of specimens allocated for training within the trypsin and collagenase groups, respectively. We used the optimizer Adam with a batch size of 8 and a number of epochs of 200 [S4]. Figure S2b illustrates the learning of the autoencoder networks showing reconstruction errors lower than 0.0166.

b) Data preparation for building ML models

Data preprocessing. Whitin the datasets (complexity and irregularity, cartilage features, compressed features), each cartilage specimen was characterized by a feature matrix $F = N_{\text{Features}} \times n_{\text{lines}}$. The values of N_{Features} were 8, 4, and 5 for complexity and irregularity features, cartilage features and compressed features, respectively. While n_{lines} was related to dimension of the ROI used for the analysis ($n_{\text{lines}} = 30$ for ROI_1 and $n_{\text{lines}} = 21$ for ROI_2). To handle the variations in the maximum signal amplitude we implemented a preprocessing normalization step for each RF line ($x \equiv x_s^l \in \mathbb{R}^{300}$). This step was carried out dividing each RF line data by its maximum absolute value. By standardizing all features, they could be seamlessly integrated, allowing subsequent models to learn weights more effectively.

Feature selection. Irrelevant features often pose a challenge during the model training process, and some noise features can even lead the model to deviate from the correct path. To address this issue, a feature selection process was employed to choose a subset of the most informative features that can effectively describe the input data and ensure accurate prediction results [S5].

In our study, after normalizing the extracted features, we applied the LinearSVC algorithm using the L1 norm regularization [S6] as a penalty item. This algorithm enabled the selection of features that exhibited a significant association with cartilage degeneration. To filter out the essential features from each complete dataset, we used a regularization parameter *C = 2*.

For the three datasets derived from the trypsin group, the feature selection step resulted in the selection of the top 33, 35 and 25 most informative features from the original sets of 240 ($n_{\text{lines}}(30) \times N_{\text{Features}}(8)$), 84 ($n_{\text{lines}}(21) \times N_{\text{Features}}(4)$) and 150 ($n_{\text{lines}}(30) \times N_{\text{Features}}(5)$) features, respectively. Similarly, for the three datasets derived from the collagenase treatment, the feature selection step identified the most important 32, 30 and 40 features out of the 240 features. The overall importance scores were calculated by aggregating the feature weights from all RF lines. The importance scores of the seventeen explored metrics are presented in Figure S3.

Feature fusion. In this process, we integrated the three datasets of QUS metrics, considering the samples allocated for training and testing as separate datasets. To do this, for the three pairs of datasets, we first selected the most crucial features based on the scores obtained from the feature selection analysis. Then, we created the $\text{Training}_{\text{MLmodel}}$ and $\text{Test}_{\text{MLmodel}}$ datasets for the trypsin treatment group by concatenating the top 33, 35, and 25 most informative features for each cartilage

sample. Similarly, for the collagenase treatment group, the $\text{Training}_{\text{MLmodel}}$ and $\text{Test}_{\text{MLmodel}}$ datasets were established by concatenating the top 32, 30, and 40 most informative features for each cartilage sample.

REFERENCES

[S1] J. S. Richman and J. R. Moorman, “Physiological time-series analysis using approximate entropy and sample entropy maturity in premature infants Physiological time-series analysis using approximate entropy and sample entropy,” *Americal Journal of Physiology Heart and Circulatory Physiology*, vol. 278, pp. H2039–H2049, 2000.

[S2] S. Z. Wang, Y. P. Huang, S. Saarakkala, and Y. P. Zheng, “Quantitative Assessment of Articular Cartilage with Morphologic, Acoustic and Mechanical Properties Obtained Using High-Frequency Ultrasound,” *Ultrasound Med Biol*, vol. 36, no. 3, pp. 512–527, 2010, doi: 10.1016/j.ultrasmedbio.2009.12.005.

[S3] A. K. Dubey and V. Jain, “Comparative Study of Convolution Neural Network’s Relu and Leaky-Relu Activation Functions,” 2019, pp. 873–880. doi: 10.1007/978-981-13-6772-4_76.

[S4] F. Zou, L. Shen, Z. Jie, W. Zhang, and W. Liu, “A Sufficient Condition for Convergences of Adam and RMSProp,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2019, pp. 11119–11127. doi: 10.1109/CVPR.2019.01138.

[S5] G. Chandrashekar and F. Sahin, “A survey on feature selection methods,” *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, Jan. 2014, doi: 10.1016/j.compeleceng.2013.11.024.

[S6] S. L. Kukreja, J. Löfberg, and M. J. Brenner, “A LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR (LASSO) FOR NONLINEAR SYSTEM IDENTIFICATION,” *IFAC Proceedings Volumes*, vol. 39, no. 1, pp. 814–819, 2006, doi: 10.3182/20060329-3-AU-2901.00128.