Osteoarthritis and Cartilage 29 (2021) 423–432

Osteoarthritis and Cartilage

Machine learning augmented near-infrared spectroscopy: *In vivo* follow-up of cartilage defects

J.K. Sarin † ‡ *, N.C.R. te Moller §, A. Mohammadi †, M. Prakash ‖, J. Torniainen † ‡, H. Brommer §, E. Nippolainen †, R. Shaikh †, J.T.A. Mäkelä †, R.K. Korhonen †, P.R. van Weeren § ¶, I.O. Afara †, J. Töyräs † ‡ #

† *Department of Applied Physics, University of Eastern Finland, Kuopio, Finland*
‡ *Diagnostic Imaging Center, Kuopio University Hospital, Kuopio, Finland*
§ *Department of Clinical Sciences, Faculty of Veterinary Medicine, Utrecht University, Utrecht, the Netherlands*
‖ *A.I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, Kuopio, Finland*
¶ *Regenerative Medicine Utrecht, Utrecht, the Netherlands*
# *School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, Australia*

ARTICLE INFO

Article history:
Received 2 July 2020
Accepted 11 December 2020

Keywords:
Osteoarthritis
Near-infrared spectroscopy
Machine learning
Convolutional neural network
Disease progression

SUMMARY

Objective: To assess the potential of near-infrared spectroscopy (NIRS) for *in vivo* arthroscopic monitoring of cartilage defects.
Method: Sharp and blunt cartilage grooves were induced in the radiocarpal and intercarpal joints of Shetland ponies and monitored at baseline (0 weeks) and at three follow-up timepoints (11, 23, and 39 weeks) by measuring near-infrared spectra *in vivo* at and around the grooves. The animals were sacrificed after 39 weeks and the joints were harvested. Spectra were reacquired *ex vivo* to ensure reliability of *in vivo* measurements and for reference analyses. Additionally, cartilage thickness and instantaneous modulus were determined via computed tomography and mechanical testing, respectively. The relationship between the *ex vivo* spectra and cartilage reference properties was determined using convolutional neural network.
Results: In an independent test set, the trained networks yielded significant correlations for cartilage thickness ($\rho = 0.473$) and instantaneous modulus ($\rho = 0.498$). These networks were used to predict the reference properties at baseline and at follow-up time points. In the radiocarpal joint, cartilage thickness increased significantly with both groove types after baseline and remained swollen. Additionally, at 39 weeks, a significant difference was observed in cartilage thickness between controls and sharp grooves. For the instantaneous modulus, a significant decrease was observed with both groove types in the radiocarpal joint from baseline to 23 and 39 weeks.
Conclusion: NIRS combined with machine learning enabled determination of cartilage properties *in vivo*, thereby providing longitudinal evaluation of post-intervention injury development. Additionally, radiocarpal joints were found more vulnerable to cartilage degeneration after damage than intercarpal joints.

Introduction

Articular cartilage is an aneural and avascular tissue, making early detection of damage challenging. Especially in young people, traumatic and immobilizing joint injuries, which can lead to post-traumatic osteoarthritis (PTOA), are common. Although advanced OA has been extensively characterized, insights on the early stages of the PTOA require further research[1]. To mitigate the progression of degeneration, early detection of traumatic cartilage injuries is

* Address correspondence and reprint requests to: J.K. Sarin, Department of Applied Physics, University of Eastern Finland, Kuopio, Finland. Tel.: 358-407677216.
E-mail addresses: jaakko.sarin@uef.fi, jaakko.sarin@gmail.com (J.K. Sarin), n.c.r.temoller@uu.nl (N.C.R. te Moller), ali.mohammadi@uef.fi (A. Mohammadi), mithilesh.prakash@uef.fi (M. Prakash), jari.torniainen@uef.fi (J. Torniainen), h.brommer@uu.nl (H. Brommer), ervin.nippolainen@uef.fi (E. Nippolainen), rubina.shaikh@uef.fi (R. Shaikh), janne.makela@uef.fi (J.T.A. Mäkelä), rami.korhonen@uef.fi (R.K. Korhonen), r.vanweeren@uu.nl (P.R. van Weeren), isaac.afara@uef.fi (I.O. Afara), j.toyras@uq.edu.au (J. Töyräs).

https://doi.org/10.1016/j.joca.2020.12.007

essential[2]. The current diagnostic measures (i.e., clinical examination, radiography, ultrasonography, and magnetic resonance imaging) are unable to detect initial cartilage damage[3,4]. Early cartilage damage may be detected during an arthroscopic procedure, e.g., when treating meniscal or ligamental tears[5]. The state-of-the-art in the arthroscopic evaluation of cartilage integrity, however, is far from optimal. The current gold-standard relies on visual inspection and manual probing of the tissue[6]; both of which are highly subjective and poorly repeatable[7]. Therefore, objective and reliable measures are required to improve the quality of cartilage diagnostics[8].

To date, several intra-articular diagnostic techniques have been suggested to replace or augment traditional arthroscopic evaluation[9,10]. Two promising optical methods, optical coherence tomography (OCT) and near-infrared spectroscopy (NIRS)[9,10], utilize non-ionizing NIR light to evaluate cartilage integrity non-destructively. NIRS is a widely applied vibrational spectroscopic technique that measures NIR absorbance in biological tissues, i.e., an optical quantity that provides an indicative measure of the tissue's biomolecular composition. In comparison to other spectral regions, such as the mid-infrared region used in Fourier transform infrared spectroscopy, NIR light penetrates deeper into biological tissues (~5 mm) and has less stringent requirements for sample preparation, making it an attractive option for whole tissue characterization[11].

Adaptation of NIRS, however, requires an extensive library of spectral measurements and target properties values (i.e., calibration data) prior to application. In joint diagnostics, common target properties include compositional, structural, and functional properties of cartilage, together describing the overall integrity of the tissue. This calibration data can be used to construct a statistical model for predicting the target properties of independent samples from their NIRS measurements. The gold-standard statistical approach in chemometric applications, such as NIRS evaluation of wheat quality[12] and soil[13], is partial least squares regression (PLSR), which has also been utilized to predict cartilage properties[14]. Recently, machine learning techniques, such as convolutional neural networks (CNNs), have been suggested as a replacement for conventional regression techniques, such as principal component regression (PCR) and PLSR, due to their superior performance[15]. While CNN has been extensively used in image analysis, such as object classification, its applications in spectroscopy, especially joint diagnostics, are still sparse[16–18]. Most recent studies have demonstrated the potential of NIRS for *ex vivo* arthroscopic evaluation of equine and human joint tissues by adapting CNN and PC analysis (PCA), respectively. However, no previous study has demonstrated prediction of cartilage properties from NIR spectra acquired *in vivo*, or utilized the technique in follow-up monitoring[10,17,19].

We hypothesize that NIRS combined with machine learning can be utilized for *in vivo* longitudinal monitoring of changes in cartilage properties (i.e., cartilage thickness and biomechanical properties) during injury progression. The hypothesis is tested by monitoring the progression of different cartilage groove-injury models in the carpal joints of Shetland ponies.

Materials and methods

Blunt and sharp grooves were inflicted via arthrotomy by a European board-certified equine surgeon (dipl. ECVS) on the dorsoproximal surface of the intermediate carpal bone (radiocarpal joint) and at the radial facet of the third carpal bone (intercarpal joint) of a randomized (left or right) front leg of Shetland ponies (female, $N = 9$, age $= 6.8 \pm 2.6$ years) as previously described (Fig. 1)[20]. The number of ponies was determined with a power analysis (power 0.90 and $P < 0.05$) based on the results of a pilot study[21] and previous studies[22–24]. None of the ponies showed lameness preoperatively. For each pony, the sham-operated contralateral joint was used as a control. NIRS measurements were performed on the grooved cartilage surfaces during the initial

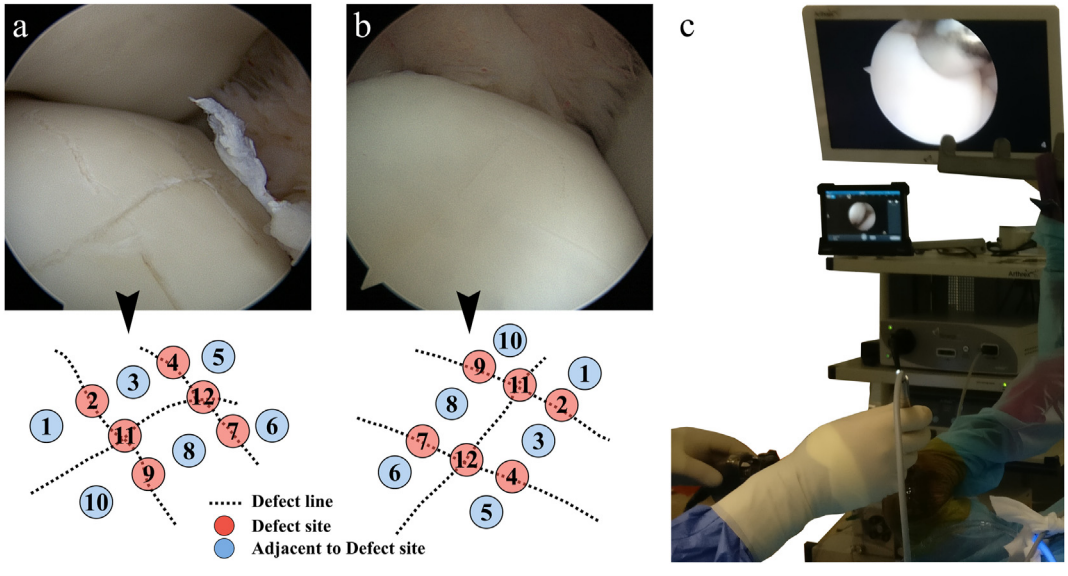

Fig. 1 Arthroscopic photos of induced blunt (a, left joint) and sharp (b, right joint) grooves at the dorsoproximal surface of the intermediate carpal bone from different ponies, and the respective measurement locations, along with a photo captured during the arthroscopic *in vivo* spectral acquisition with a NIRS probe while using an arthroscope for navigation (c).

surgery (baseline, 0 weeks) and arthroscopically at the three follow-up time points (11, 23, and 39 weeks, Fig. 1(c)), whereas the control joints were only measured at baseline and 39 weeks to limit the time of anaesthesia (due to high risk) and as no changes were expected on their cartilage properties. After 39 weeks, the ponies were humanely euthanized, and the joints were stored at −20°C for further analysis. The study was authorized by the Utrecht University Animal Experiments Committee (AVD108002015307, Utrecht, The Netherlands) in compliance with the Dutch Act on Animal Experimentation.

In vivo arthroscopic measurements

During the arthroscopic procedure, a conventional arthroscope (4 mm, 30° inclination, Synergy HD3 system, Arthrex, Naples, FL, USA) was used for navigating and aligning the NIRS probe perpendicular to the cartilage surface [Fig. 1(c)] using standard portals as described by McIlwraith *et al.*[25] at follow-ups. Baseline measurements (0 weeks) were performed during the arthrotomy. Ringer's solution (Fresenius, Bad Homburg v.d.H., Germany) containing sodium chloride (8.6 g/l), potassium chloride (0.3 g/l), and calcium chloride (0.33 g/l) was used for joint distension. *In vivo* NIR spectra were acquired with instrumentation, including two spectrometers and a light source (AvaSpec-ULS2048L, λ = 0.4–1.1 μm, resolution = 0.58 nm, AvaSpec-NIR256-2.5-HSC, λ = 1.0–2.5 μm, resolution = 6.4 nm, and AvaLight-HAL-(S)-Mini, λ = 0.36–2.5 μm, Avantes BV, Apeldoorn, The Netherlands), and a custom arthroscopic fibre optic probe (Avantes BV)[10]. NIRS measurements were conducted at 12 locations, on and adjacent to grooves, in each joint (Fig. 1). At each location, 15 spectra (each spectrum consists of ten coadded scans) were recorded with a total acquisition time of 2.7 s. The total number of *in vivo* spectra from baseline to the final follow-up time point were 5017, 2951, 2917, and 5404, respectively, after exclusion of measurements with instrumentation-related errors. Several locations could not always be assessed due to anatomical constraints.

Ex vivo cartilage thickness, NIRS, and biomechanical measurements

After extraction of osteochondral samples, cartilage thickness was determined from micro-computed tomography images (Fig. 2(a)–(b), 90 kV tube voltage, 40 × 40 × 40 μm^3 voxel size, Quantum FX, Perkin Elmer, Waltham, MA, USA) by using a custom-made Matlab-function (R2018b, Mathworks, Natick, MA) for

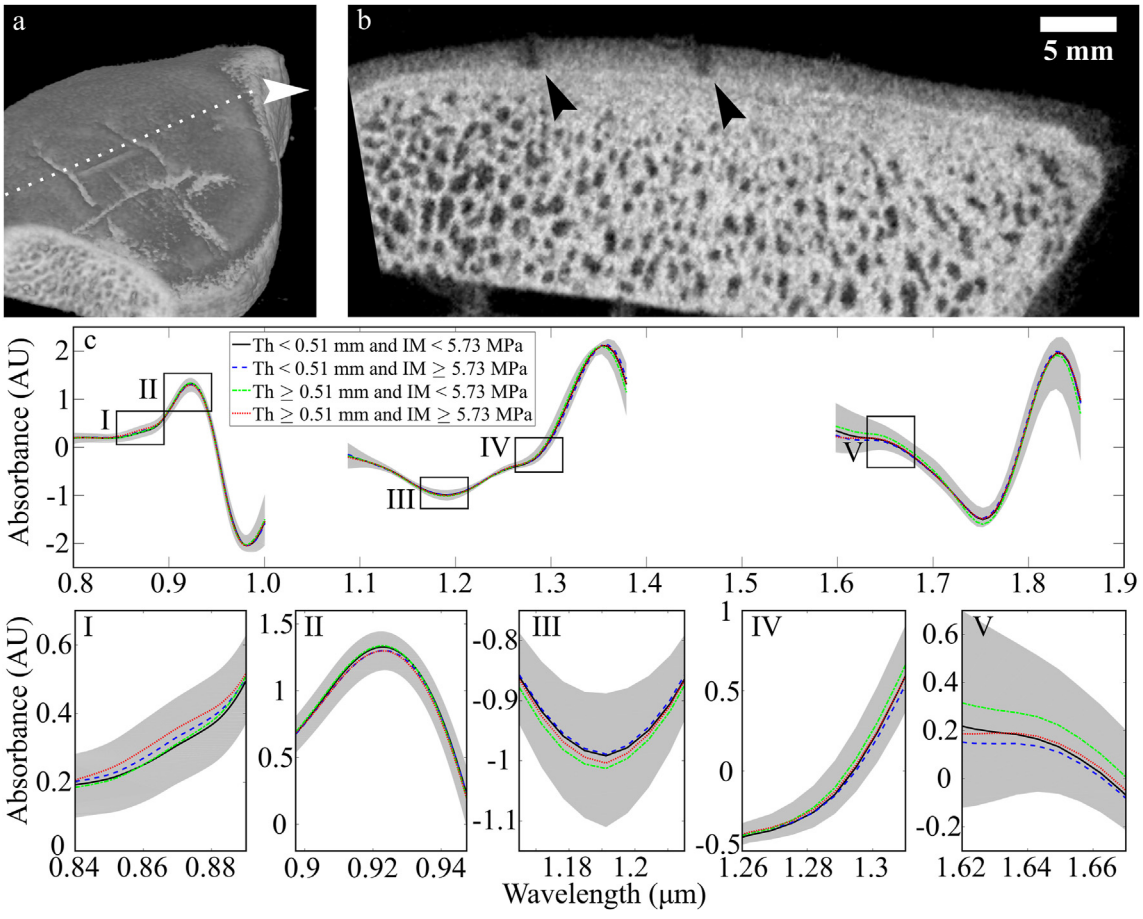

Fig. 2 A 3D rendering of micro-computed tomography images (a) and a cross-sectional image (b, acquisition location indicated with a dotted line in Fig. 2(a)) from the same sample as presented in Fig. 1(a). The average *ex vivo* spectra, including standard deviation (grey), with a division to 4 groups based on lower or higher cartilage thickness (Th) and instantaneous modulus (IM) values compared to their median (c) with additional subfigures on several spectral regions (I–V).

Osteoarthritis and Cartilage

locations adjacent to grooves (i.e., locations 1, 3, 5, 6, 8, and 10, Fig. 1(a)), locations of the control joints, and locations of kissing sites (i.e., the opposing cartilage surfaces in contact with the grooved sites). The measured cartilage thickness was required to customize the biomechanical testing protocol for each measurement location.

NIR spectra (n = 5456) were also acquired *ex vivo* from the osteochondral samples [Fig. 2(c)] with acquisition settings and system identical to those used during *in vivo* arthroscopy. *Ex vivo* measurements were performed at room temperature with the sample submerged in a phosphate-buffered saline bath. Additionally, six locations of each joint were measured at the kissing site.

A commercial mechanical tester (Mach-1 v500css, Biomomentum Inc., Laval, Quebec, Canada) was used to determine the instantaneous modulus (IM) for locations adjacent to defects semi-automatically using a spherical indenter of 0.5 mm in diameter as described previously[20]. Briefly, the mechanical protocol consisted of a preload of 0.1 N, followed by a single 15% strain indentation step (velocity = 100% strain/s). The IM was determined from the peak stress/strain ratio according to Hayes *et al.* by assuming a Poisson's ratio of 0.5[26]. No change in cartilage biomechanical properties was expected to occur due to the freeze–thaw cycles[27].

Outlier detection and data preparation

The relationship between NIR spectra and cartilage reference properties (i.e., cartilage thickness and IM) was investigated using CNNs. First, data from the spectral regions 0.4–0.8 μm (light from the conventional arthroscope), 1.4–1.6 μm (water saturation), and 1.85–2.5 μm (water saturation) were excluded and, thus, the analysis was limited to data in the spectral regions 0.8–1.4 μm and 1.6–1.85 μm[10,17]. Although the spectral region of 2.0–2.5 μm has been previously attributed to cartilage components[28], utilization of this region would have required additional coaddition of spectra to achieve high enough signal-to-noise ratio, thereby resulting in an unreasonably long acquisition time for *in vivo* measurements. The spectra were first smoothed and scatter-corrected using standard normal variate (SNV). In addition, first and second-order derivative pre-processing were computed. Pre-processing was performed using 3rd-degree Savitzky–Golay algorithm with window sizes of 149 (86 nm) and 23 (148 nm) for the different spectrometer outputs.

While CNNs have reasonable tolerance against noisy data, removing bad NIR spectra from the calibration dataset (i.e., outlier rejection) can greatly improve the prediction accuracy. Outlier detection was performed by calculating the first three principal scores of the pre-processed *ex vivo* spectra with PCA and plotting them in a 3D-space (Fig. 3)[29]. Unlike during *in vivo* NIRS measurements, contact between the probe and cartilage surface could be ensured during the *ex vivo* acquisition. Based on the scatter plot, abnormal spectra that were separated from the main group were manually rejected. The main group was enclosed in a volume by adopting Delaunay triangulation. The spectra acquired *in vivo* were then projected to the same PCA space and the spectra with their response outside the volume were deemed outliers. Outlier detection was performed separately for the three spectral regions, and outlier detection in any of the spectral regions resulted in the exclusion of the whole spectrum.

Prior to machine learning analysis, the *ex vivo* spectra were standardized (i.e., rescaling each spectral variable to a mean of 0 and standard deviation of 1) using the StandardScaler class of the sklearn package in Python to account for the different order of magnitude in the pre-processed spectra. This scaling was then applied to the *in vivo* spectra. Due to the skewed distribution of the target properties, logarithm and square root transformations were applied to cartilage thickness and IM, respectively, during CNN training. This was followed by normalization between 0 and 1 using the MinMaxScaler class of the sklearn package. These steps have been shown to improve the performance and training stability of CNNs[30,31].

Machine learning

Neural networks have an input layer (data input), hidden layers (data representation), and an output layer (prediction). Each hidden layer consists of several neurons that include an activation

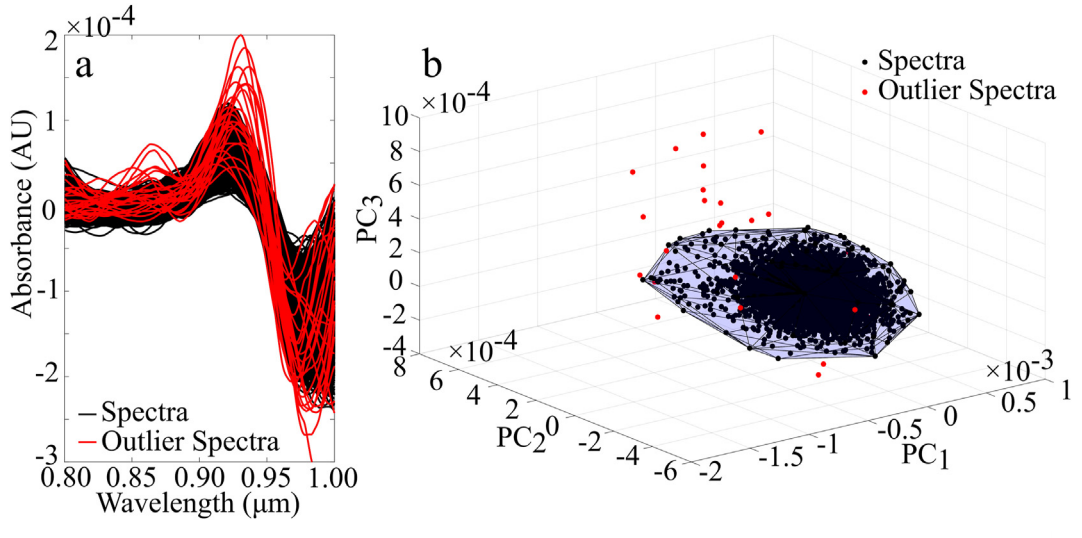

Fig. 3 Pre-processed 2nd-derivative *ex vivo* spectra from the spectral region of 0.8–1.0 μm (a) and their corresponding principal component (PC) scores with the 3D-volume (b). The red lines and points present the outliers.

Osteoarthritis and Cartilage

Model	Timepoints	Treatment groups	Fixed factors	Interactions	Random effect
Model 1	Baseline, 39 weeks	Control blunt, Blunt-grooved, Control sharp, Sharp -grooved	Treatment group, time point, joint, location*	Treatment group, time point, joint	Pony
Model 2	Baseline, 11 weeks, 23 weeks, 39 weeks	Blunt-grooved, Sharp-grooved	Treatment group, time point, joint, location	Treatment group, time point, joint	Pony

Table I Specifications of statistical models. Time point was used as a categorical value. Joint was defined as radiocarpal or intercarpal and location as dorsal or palmar of the central groove. *For analysis of cartilage thickness at the grooved locations (i.e., locations 2, 4, 7, 9, 11, and 12), the location was excluded as a fixed factor as it had no significant effect and AIC improved without this factor

Osteoarthritis and Cartilage

function, such as rectified linear unit (ReLU) and swish[32,33], and a set of weights and biases that are optimized through training with a large set of calibration data. A challenge with conventional multivariate regression techniques, as well as machine learning methods, is that they are prone to overfitting, i.e., a model performs well on the calibration data but poorly on new data. Generally, data splitting into independent subgroups can minimize this limitation. With CNNs, the ability of the network to generalise well to new data can be further ensured by regularization techniques (i.e., Lasso and Ridge regression) and batch normalization[34].

In this study, the data from a spectral region of 0.8–1.0 μm and a concatenation of 1.05–1.4 and 1.6–1.85 μm regions were fed into the network separately (i.e., branches 1 and 2, respectively) due to the resolution difference of the spectrometers (Supp. Fig. 1)[35]. As a result, the resolution difference could be accounted for by separately tuning the widths of the convolutional filters. The network included the two branches, each having three subsequent combinations of a 1D-convolution layer (filters = 128) with swish activation and L2-regularization (i.e., Ridge regression)[33], a batch normalisation layer[34], and a max pooling layer (pool = 2, strides = 2). The outputs of the third max pooling layers were flattened, concatenated, and input to a fully connected dense layer (neurons = 128) with rectified linear unit (ReLU) activation and L2-regularization[32]. This was followed by a dropout layer (=0.50) and a linear dense layer. For branches 1 and 2, the initial kernel sizes of convolutional layers were 40 and 10, respectively, with the kernel size halved at each layer. Glorot uniform (kernel) and zero (bias) weight initialization were used at each layer. To minimize the

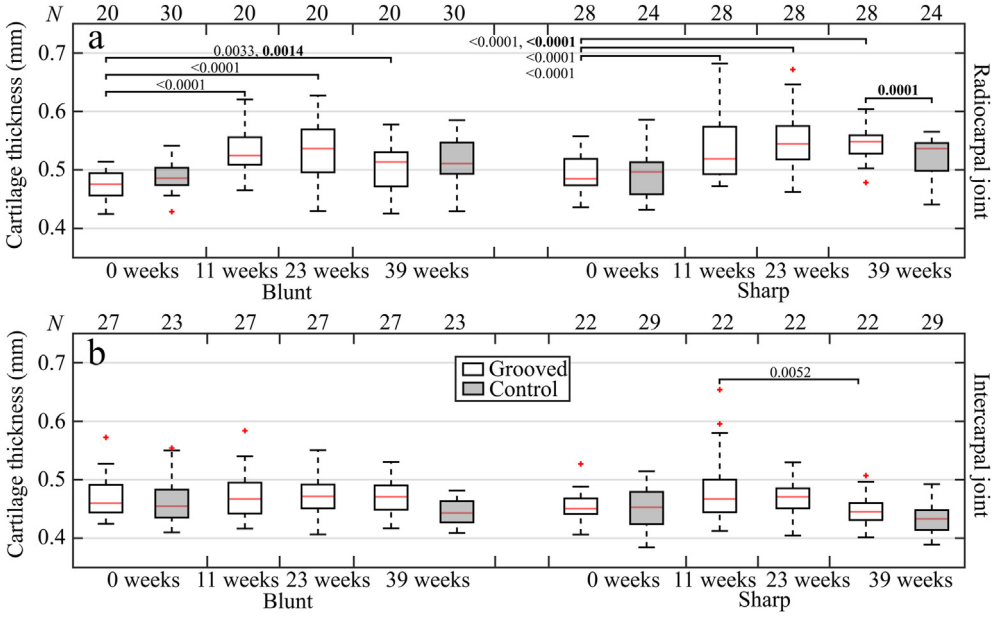

Fig. 4 Boxplots of grooved (white) and control (grey) cartilage with median (red line), quartiles (25% and 75%), and outliers (red crosses) for cartilage thickness predicted from *in vivo* spectra of the independent test set at different time points. Locations measured adjacent to the grooves for radiocarpal (a) and intercarpal (b) joints are presented separately. Significant differences ($P < 0.01$, Table I) are presented with their P-values based on Model 1 (bold) and Model 2. The number of locations (N) is presented above each bar. Only locations with successful measurements at all four (grooved) or two (control) time points were included.

Osteoarthritis and Cartilage

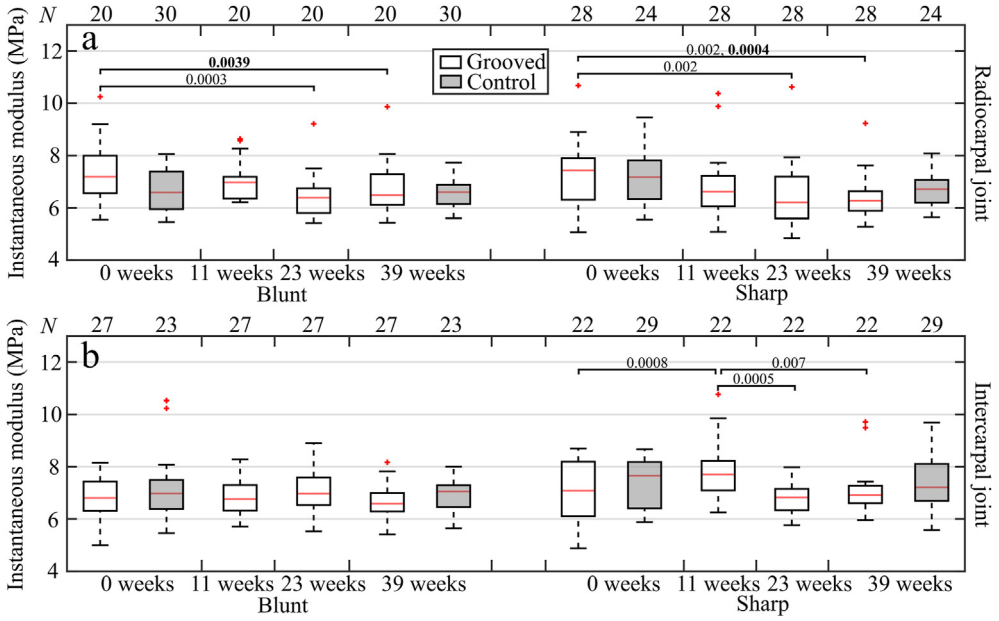

Fig. 5

Boxplots of grooved (white) and control (grey) cartilage with median (red line), quartiles (25% and 75%), and outliers (red crosses) for instantaneous modulus predicted from *in vivo* spectra of the independent test set with instantaneous modulus at different time points. Locations measured adjacent to grooves for radiocarpal (a) and intercarpal (b) joints are presented separately. Significant differences ($P < 0.01$, Table I) are presented with their *P*-values based on Model 1 (bold) and Model 2. The number of locations (*N*) is presented above each bar. Only locations with successful measurements at all four (grooved) or two (control) time points were included.

Osteoarthritis and Cartilage

chance of overfitting, callbacks and EarlyStopping were applied to reduce the learning rate, if the mean squared error (MSE) of the validation group did not decrease for 30 epochs, and to halt training, if no decrease in error was observed for 60 epochs, respectively. Network training was performed using the Adam optimizer in Keras.

The networks were optimized using 4-fold cross-validation with 8 ponies and further evaluated with the data from the remaining pony (i.e., independent test set). This process was repeated 9 times with each pony used once as the independent test set. The spectra acquired *ex vivo* and arthroscopically at 39 weeks were included in the network training to account for possible differences between *in vivo* and *ex vivo* measurements.

Statistics

Network performance was evaluated by comparing the measured and predicted reference values using the non-parametric Spearman's rank correlation (due to non-normal distribution of reference parameters), and by estimating the root mean square error (RMSE) of calibration, validation, and independent test groups. In addition, normalized RMSE (NRMSE, i.e., RMSE divided by the range of the reference variable) was also calculated. The final network architecture was selected based on the average performance (i.e., smallest RMSE) of the validation group. For prediction and visualization of temporal trends of cartilage reference properties (Figs. 4 and 5), only the independent predictions were used. For statistical analysis, RStudio (version 1.1.463) was used. Statistical differences in cartilage thickness and IM between time points and treatment groups were investigated using two linear mixed effect models with 'pony' as the random effect (nlme package, version 3.1–137[36]), followed by pair-wise comparisons of estimated means with false discovery rate correction (Table I). Using this approach, dependencies within animals were considered. Both cartilage thickness and IM data were normalized using logarithmic transformation. Model estimates were based on restricted maximum likelihood estimators. The limit of statistical significance was set to $P < 0.01$ as a more conservative limit was considered necessary based on 95% confidence intervals.

Results

In qualitative evaluations, the average *ex vivo* spectra, divided based on median reference values, revealed spectral data in the regions of 0.84–0.88 μm and 0.9–0.94 μm to be more indicative of IM, whereas differences in cartilage thickness were better observed in the region of 1.18–1.66 μm (Fig. 2).

In total, 0.82% of the *ex vivo* spectra ($n = 45$) were rejected as outliers. From the *in vivo* spectral measurements: 23.0% of spectra at baseline, 15.6% of spectra at 11 weeks, 4.5% of spectra at 23 weeks, and 11.0% of spectra at 39 weeks were rejected by the outlier detection algorithm. Outliers were more common in the spectral region of 0.8–1.0 μm ($n = 1600$) compared to the spectral regions of 1.05–1.4 μm ($n = 470$) and 1.6–1.85 μm ($n = 520$). Investigation of the raw outlying spectra revealed higher absorbances in all three regions (0.02 AU, 0.08 AU, and 0.05 AU, respectively). In addition, frequency-domain analysis of the rejected spectra revealed more high-frequency components than in the accepted spectra.

Spectral data pre-processed with second derivative resulted in the best performing CNNs. In general, more accurate predictions

were obtained with cartilage thickness than IM (Table II), arguably due to the light pathlength effect[37]. Thus, to account for this effect, cartilage thickness was added as an additional input parameter when training the networks for estimation of IM. This improved the networks' prediction accuracy (Table II – IM 2).

Prediction of temporal variation in the target properties revealed similar findings at and adjacent to the grooves; thus, only predictions adjacent to the grooves are presented here (Figs. 4–5, Table III). Predictions at the grooves are included in the supplementary material (Supp. Figs. 2 and 3, Supp. Table I). The minimum effect sizes, determined with a power of 0.90 and a *P*-value of 0.01, for cartilage thickness and IM were 22.2 μm (average 41.4 μm) and 0.65 MPa (average 0.78 MPa), respectively. At 11 weeks, cartilage thickness had significantly increased in both blunt and sharp grooves in the radiocarpal joint and remained significantly higher throughout the experiment (Fig. 4(a), Suppl. Fig. 2(a)). In the intercarpal joint, the only significant difference in cartilage thickness was observed in sharp grooves, showing a decrease from 11 weeks to 39 weeks (Fig. 4(b), Suppl. Fig. 2(b)). In the radiocarpal joint, IM was significantly lower at 23 and 39 weeks when compared to the baseline in both groove types (Fig. 5(a), Suppl. Fig. 3(a)), whereas, in the intercarpal joint, a significant increase (from baseline to 11 weeks) and decrease (from 11 weeks to 39 weeks) were only observed with sharp grooves (Fig. 5(b), Suppl. Fig. 3(b)).

Discussion

In this study, NIRS was utilized for *in vivo* monitoring of longitudinal changes in cartilage thickness and IM after the infliction of sharp and blunt grooves. As hypothesized, machine learning, based on CNNs, was able to predict cartilage properties from its NIR spectra (Table II), as well as to estimate these properties at earlier follow-up time points (Figs. 4 and 5, Suppl. Figs 2-3). To our knowledge, no study has quantitatively evaluated cartilage properties at sequential *in vivo* time points after traumatic injury. Previous studies have merely focused on post-mortem analysis[22,38,39]. Therefore, arthroscopic NIRS represents great potential for *in vivo* evaluation of cartilage integrity, as well as for *in vivo* studies focusing on regenerative medicine, which would benefit from quantitative longitudinal monitoring to better identify promising treatments.

Cartilage damage is often initiated by mechanical wear or traumatic injuries to the joint, eventually leading to irreversible loss of cartilage and deterioration of mechanical performance[40]. In this study, the cartilage groove injuries compromised the integrity of the cartilage macromolecular framework (i.e., collagen), arguably leading to a decrease in the aggregation of proteoglycans, aggrecan concentration, and the length of glycosaminoglycan chains and, thus, increased matrix water content[40]. The initial swelling of the cartilage matrix was likely due to increased water content, which was observed between baseline and 11 weeks in the radiocarpal joint (Fig. 4(a), Suppl. Fig. 2(a)). After 11 weeks, cartilage thickness of blunt grooves presented a more downwards trend compared to sharp grooves. We believe that the initial cartilage loss in the blunt grooves resulted in decreased fluid pressurization and increases tissue strains around the grooves[41,42], leading to additional collagen damage and compromised function. The IM, which is mainly regulated by the collagen network and fluid pressurization[43], did not change during the first 11 weeks in the radiocarpal joint. However, at the later time points, a systematic decrease in the IM was observed, presumably as a result of progressive collagen damage. This observation is supported by Mastbergen *et al.*[22] and Marijnissen *et al.*[38] who demonstrated progressive collagen damage in ovine fetlock joints at 15 and 37 weeks, and in canine knee joints at 20 and 40 weeks with groove model, respectively.

Estimation of native cartilage thickness with NIRS has previously been demonstrated by Afara *et al.*[37,44], McGoverin *et al.*[45], Prakash *et al.*[19], and Sarin *et al.*[14,17]. However, specimens, spectral regions, number of samples, and statistical methods vary substantially between the studies. Both McGoverin *et al.* and Prakash *et al.* reported similar validation accuracy for *in vitro* measurements of human tissue with R^2 = 64% and NRMSE = 15.3%, and ρ = 0.83 and NRMSE = 14%, respectively, at spectral regions comparable to this study[18,39]. In addition, Prakash *et al.*[19] reported the performance (ρ = 0.52 and NRMSE = 25%) of the independent arthroscopic *ex vivo* test group. While the validation performance in this study was slightly weaker (ρ = 0.52, NRMSE = 21.5%), the performance of the independent test group was similar to the values reported by Prakash *et al.*[19] The difference is most likely caused by the variance in the cartilage conditions between the studies, i.e., in this study, the reference measurements were only available for visually healthy cartilage (i.e., not from groove locations), whereas varying degrees of cartilage degeneration were reported by both McGoverin *et al.* and Prakash *et al.* with modified Mankin scores between 2 and 12 and ICRS grades between 0 and 4, respectively[46,47]. The results of the current study are also in line with our earlier study[17], where slightly lower errors (NRMSE = 17.2%) were achieved for *in vitro* estimation of equine cartilage thickness.

Parameter	Mean	95% CI	Range	Statistics	Calibration	Validation	Test
Thickness (mm)	0.507	0.489–0.526	0.22–0.93	Spearman	0.740	0.524	0.473
				RMSE	0.124	0.153	0.155
				NRMSE	17.5%	21.5%	21.8%
Instantaneous modulus (MPa)	7.08	6.69–7.48	1.29–15.16	Spearman	0.768	0.432	0.332
				RMSE	2.07	2.65	2.76
				NRMSE	14.9%	19.1%	19.9%
Instantaneous modulus 2				Spearman	0.784	0.594	0.498
	_	–	–	RMSE	1.94	2.43	2.43
				NRMSE	14.0%	17.5%	17.5%

Table II Cartilage reference properties and statistics of network performance. For instantaneous modulus 2, cartilage thickness was included as an additional predictor to the CNN

			0 weeks		11 weeks	23 weeks	39 weeks	
			Exp.	Control	Exp.	Exp.	Exp.	Control
Cartilage thickness (μm)	Radiocarpal	Blunt (Fig. 4a)	**471 (448, 494)**; 471 (446, 497)	**482 (459, 506)**	533 (504, 564)	532 (503, 562)	**502 (478, 527)**; 503 (476, 532)	**506 (482, 531)**
		Sharp (Fig. 4a)	**495 (474, 517)**; 495 (471, 520)	**483 (463, 505)**	536 (510, 563)	551 (525, 579)	**542 (519, 566)**; 542 (516, 569)	**503 (481,525)**
	Intercarpal	Blunt (Fig. 4b)	**469 (449, 490)**; 469 (447, 493)	**472 (452, 493)**	472 (449, 496)	478 (455, 502)	**465 (446, 486)**; 465 (443, 489)	**463 (443, 484)**
		Sharp (Fig. 4b)	**456 (435, 479)**; 456 (432, 482)	**449 (427, 471)**	482 (456, 509)	471 (446, 498)	**448 (427, 471)**; 449 (425, 474)	**430 (409, 451)**
Instantaneous modulus (MPa)	Radiocarpal	Blunt (Fig. 5a)	**7.58 (6.91, 8.32)**; 7.58 (6.87, 8.37)	**6.81 (6.21, 7.48)**	7.08 (6.40, 7.84)	6.54 (5.92, 7.23)	**6.83 (6.21, 7.50)**; 6.84 (6.19, 7.56)	**6.58 (6.00, 7.22)**
		Sharp (Fig. 5a)	**7.06 (6.50, 7.67)**; 7.06 (6.46, 7.71)	**7.12 (6.55, 7.73)**	6.70 (6.13, 7.32)	6.34 (5.81, 6.93)	**6.30 (5.80, 6.85)**; 6.30 (5.77, 6.88)	**6.86 (6.23, 7.46)**
	Intercarpal	Blunt (Fig. 5b)	**6.77 (6.23, 7.36)**; 6.77 (6.19, 7.39)	**6.92 (6.36, 7.52)**	6.82 (6.24, 7.45)	7.03 (6.44, 7.68)	**6.65 (6.12, 7.23)**; 6.65 (6.09, 7.26)	**6.85 (6.30, 7.44)**
		Sharp (Fig. 5b)	**6.86 (6.26, 7.53)**; 6.86 (6.22, 7.57)	**7.29 (6.64, 8.01)**	7.81 (7.08, 8.62)	6.79 (6.16, 7.49)	**7.03 (6.41, 7.72)**; 7.04 (6.38, 7.77)	**7.23 (6.59, 7.93)**

Table III The estimated mean (95% confidence intervals) for cartilage thickness and instantaneous modulus per treatment group, joint and time point derived from Model 1 (bold) and Model 2 for locations adjacent to defects

Osteoarthritis and Cartilage

The relationship between IM and NIR spectra has previously been presented in only two studies[14,19]. The *in vitro* study by Sarin *et al.* involving equines used a narrower spectral region of 0.73–0.95 μm and reported poor calibration accuracy (R^2 = 41.8%, RMSE = 3.01 MPa) compared to this study (ρ = 0.784, RMSE = 1.94 MPa). The aforementioned spectral region has higher penetration depth into cartilage compared to wavelengths above 1.0 μm and, thus, the data has included some contributions of subchondral bone[11]. Prakash *et al.* used a spectral region of 0.7–1.85 μm with cadavers and reported a high correlation in independent testing (ρ = 0.84) and a similar error (NRMSE = 19%) compared to this study (ρ = 0.498, NRMSE = 17.5%). The difference in correlations is most probably due to varying degrees of cartilage degeneration in the study of Prakash *et al.* as described earlier. Also, compared to equine, human cartilage is thicker[48]; as a result, the spectra is less influenced by the underlying subchondral bone. In future studies, the utilization of spectral region of 2.0–2.5 μm could be useful for the prediction of the IM due to its attribution to cartilage collagen[28,45]; albeit, only with hardware-related optimization needed to achieve high enough SNR.

The main limitation of the present study is that cartilage reference parameters could only be measured at the end of the study and not at the follow-up time points. Currently, cartilage thickness cannot be reliably estimated with routine arthroscopic tools. Likewise, arthroscopic evaluation of cartilage biomechanical competence is challenging, as the only clinically available method is the highly subjective manual probing of the tissue[8]. Another limitation was the narrow joint cavities of radio- and intercarpal joints in Shetland ponies, resulting in a relatively high number of outliers due to poor contact between the NIRS probe and the cartilage surface. Spectra measured with poor contact were, however, successfully identified and removed by the outlier detection analysis. Furthermore, the relatively higher number of *in vivo* spectral outliers at baseline (arthrotomy) and the first follow-up (arthroscopy) demonstrated that the reliability of arthroscopic NIRS was improved by optimizing the incision location when dealing with narrow joint spaces. In addition, due to limitations imposed by the biomechanical testing system and the great number of biomechanical measurements required, dynamic and equilibrium moduli were not determined.

In conclusion, arthroscopic NIRS combined with machine learning enabled *in vivo* monitoring of cartilage properties in the equine carpal joint. Therefore, this technique has great potential for *in vivo* evaluation of cartilage integrity, as well as for *in vivo* follow-up of new regenerative therapies. In future studies, the trained CNN can be directly applied *in situ* during similar interventions. Additionally, this work provided valuable information on the clinical application of arthroscopic NIRS, laying the foundation for *in vivo* application during arthroscopies of human joints.

Author contributions

Conception and design of the study: Sarin, JK.; te Moller, NCR.; van Weeren, PR.; Korhonen, RK; Töyräs, J. **Acquisition of the data:** Sarin, JK.; te Moller, NCR.; Mohammadi, A.; Prakash, M.; Brommer, H.; Nippolainen, E.; Shaikh, R.; **Analysis and interpretation of data:** Sarin, JK.; te Moller, NCR.; Mohammadi, A.; Torniainen, J.; Mäkelä, JTA.; Korhonen, RK.; Afara, IO., Töyräs, J.

All authors contributed to the drafting or revising the article, and approved the final submitted version.

Conflict of interest

None declared.

Acknowledgements

The Doctoral Programme in Science, Technology and Computing (SCITECO) of University of Eastern Finland, Kuopio University Hospital (VTR Projects 5041750 and 5041744, PY210 Clinical Neurophysiology), the Academy of Finland (Projects 267551, 315820, 316258, and 324529), the Orion Research Foundation sr, the Finnish Foundation of Technology Promotion, The MIRACLE Project-Horizon 2020 Research and Innovation Programme-H2020-ICT-2017-1 (grant agreement No. 780598), and the NWO Graduate Programme Grant (022.005.018) financially supported this study.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.joca.2020.12.007.

References

1. Favero M, Ramonda R, Goldring MB, Goldring SR, Punzi L. Early knee osteoarthritis: 1. RMD Open 2015;1(Suppl 1), e000062, https://doi.org/10.1136/rmdopen-2015-000062.
2. Stiebel M, Miller LE, Block JE. Post-traumatic knee osteoarthritis in the young patient: therapeutic dilemmas and emerging technologies. Open Access J Sports Med 2014;5:73–9, https://doi.org/10.2147/OAJSM.S61865.
3. von Engelhardt LV, Lahner M, Klussmann A, Bouillon B, Dàvid A, Haage P, *et al*. Arthroscopy vs. MRI for a detailed assessment of cartilage disease in osteoarthritis: diagnostic value of MRI in clinical practice. BMC Muscoskel Disord 2010;11(1):75, https://doi.org/10.1186/1471-2474-11-75.
4. Friemert B, Oberländer Y, Schwarz W, Häberle HJ, Bähren W, Gerngross H, *et al*. Diagnosis of chondral lesions of the knee joint: can MRI replace arthroscopy? A prospective study. Knee Surg Sports Traumatol Arthrosc 2004;12(1):58–64, https://doi.org/10.1007/s00167-003-0393-4.
5. Orlando Júnior N, de Souza Leão MG, de Oliveira NHC. Diagnosis of knee injuries: comparison of the physical examination and magnetic resonance imaging with the findings from arthroscopy. Rev Bras Ortop (English Ed.) 2015;50(6):712–9, https://doi.org/10.1016/j.rboe.2015.10.007.
6. Spahn G, Klinger HM, Hofmann GO, Gunter Spahn, Klinger HM, Hofmann Gunther O. How valid is the arthroscopic diagnosis of cartilage lesions? Results of an opinion survey among highly experienced arthroscopic surgeons. Arch Orthop Trauma Surg 2009;129(8):1117–21, https://doi.org/10.1007/s00402-009-0868-y.
7. Spahn G, Klinger HM, Baums M, Pinkepank U, Hofmann GO. Reliability in arthroscopic grading of cartilage lesions: results of a prospective blinded study for evaluation of inter-observer reliability. Arch Orthop Trauma Surg 2011;131(3):377–81, https://doi.org/10.1007/s00402-011-1259-8.
8. Spahn G, Klinger HM, Hofmann GO. How valid is the arthroscopic diagnosis of cartilage lesions? Results of an opinion survey among highly experienced arthroscopic surgeons. Arch Orthop Trauma Surg 2009;129(8):1117–21, https://doi.org/10.1007/s00402-009-0868-y.
9. Li X, Martin S, Pitris C, Ghanta R, Stamper DL, Harman M, *et al*. High-resolution optical coherence tomographic imaging of osteoarthritic cartilage during open knee surgery. Arthritis Res Ther 2005;7(2):318–23, https://doi.org/10.1186/ar1491.
10. Sarin JK, te Moller NCR, Mancini IAD, Brommer H, Visser J, Malda J, *et al*. Arthroscopic near infrared spectroscopy enables simultaneous quantitative evaluation of articular cartilage and subchondral bone in vivo. Sci Rep 2018;8(1):13409, https://doi.org/10.1038/s41598-018-31670-5.
11. Padalkar MV, Pleshko N. Wavelength-dependent penetration depth of near infrared radiation into cartilage. Analyst 2015;140(7):2093–100, https://doi.org/10.1039/c4an01987c.
12. Mutlu AC, Boyaci IH, Genis HE, Ozturk R, Basaran-Akgul N, Sanal T, *et al*. Prediction of wheat quality parameters using near-infrared spectroscopy and artificial neural networks. Eur Food Res Tech 2011;233(2):267–74, https://doi.org/10.1007/s00217-011-1515-8.
13. Goldshleger N, Chudnovsky A, Ben-Dor E. Using reflectance spectroscopy and artificial neural network to assess water infiltration rate into the soil profile. Appl Environ Soil Sci 2012;2012:1–9, https://doi.org/10.1155/2012/439567.
14. Sarin JK, Amissah M, Brommer H, Argüelles D, Töyräs J, Afara IO. Near infrared spectroscopic mapping of functional properties of equine articular cartilage. Ann Biomed Eng 2016;44(11):3335–45, https://doi.org/10.1007/s10439-016-1659-6.
15. Cui C, Fearn T. Modern practical convolutional neural networks for multivariate regression: applications to NIR calibration. Chemometr Intell Lab Syst 2018;182:9–20, https://doi.org/10.1016/j.chemolab.2018.07.008.
16. Acquarelli J, van Laarhoven T, Gerretzen J, Tran TN, Buydens LMC, Marchiori E. Convolutional neural networks for vibrational spectroscopic data analysis. Anal Chim Acta 2017;954:22–31, https://doi.org/10.1016/j.aca.2016.12.010.
17. Sarin JK, Nykänen O, Tiitu V, Mancini IAD, Brommer H, Visser J, *et al*. Arthroscopic determination of cartilage proteoglycan content and collagen network structure with near-infrared spectroscopy. Ann Biomed Eng 2019;47(8):1815–26, https://doi.org/10.1007/s10439-019-02280-7.
18. Afara IO, Sarin JK, Ojanen S, Finnilä MAJ, Herzog W, Saarakkala S, *et al*. Machine learning classification of articular cartilage integrity using near infrared spectroscopy. Cell Mol Bioeng 2020:1, https://doi.org/10.1007/s12195-020-00612-5.
19. Prakash M, Joukainen A, Torniainen J, Honkanen MKM, Rieppo L, Afara IO, *et al*. Near-infrared spectroscopy enables quantitative evaluation of human cartilage biomechanical properties during arthroscopy. Osteoarthritis Cartilage 2019;27(8):1235–43, https://doi.org/10.1016/j.joca.2019.04.008.
20. te Moller NCR, Mohammadi A, Plomp S, Bragança FMS, Beukers M, Pouran B, *et al*. Structural, compositional, and functional effects of blunt and sharp cartilage damage on the joint: a 9-month equine groove model study. J Orthop Res 2020:1–13, https://doi.org/10.1002/jor.24971.
21. te Moller NCR. Development of an equine carpal groove model to study early changes in osteoarthritis - a pilot study. Osteoarthritis Cartilage 2018;26(2018):S132–3, https://doi.org/10.1016/j.joca.2018.02.288.
22. Mastbergen SC, Pollmeier M, Fischer L, Vianen ME, Lafeber FPJG. The groove model of osteoarthritis applied to the ovine fetlock joint. Osteoarthritis Cartilage 2008;16(8):919–28, https://doi.org/10.1016/j.joca.2007.11.010.
23. de Visser HM, Weinans H, Coeleveld K, van Rijen MHP, Lafeber FPJG, Mastbergen SC. Groove model of tibia-femoral osteoarthritis in the rat. J Orthop Res 2017;35(3):496–505, https://doi.org/10.1002/jor.23299.
24. Maninchedda U, Lepage OM, Gangl M, Hilairet S, Remandet B, Meot F, *et al*. Development of an equine groove model to induce metacarpophalangeal osteoarthritis: a pilot study on 6 horses. PLoS One 2015;10(2):1–18, https://doi.org/10.1371/journal.pone.0115089.
25. McIlwraith CW, Nixon AJ, Wright IM. Diagnostic and Surgical Arthroscopy in the Horse. 4th edn 2015:45–110.
26. Hayes WC, Keer LM, Herrmann G, Mockros LF. A mathematical analysis for indentation tests of articular cartilage. J Biomech 1972;5(5):541–51, https://doi.org/10.1016/0021-9290(72)90010-3.
27. Szarko M, Muldrew K, Bertram JE. Freeze-thaw treatment effects on the dynamic mechanical properties of articular cartilage. BMC Muscoskel Disord 2010;11(1):231, https://doi.org/10.1186/1471-2474-11-231.
28. Palukuru UP, McGoverin CM, Pleshko N. Assessment of hyaline cartilage matrix composition using near infrared spectroscopy.

Matrix Biol 2014;38:3–11, https://doi.org/10.1016/j.matbio.2014.07.007.
29. Burns DA, Ciurczak EW. Handbook of near-infrared analysis, 3rd ed. Anal Bioanal Chem 2009;393(5):1387–9.
30. Shanker MS, Hu MY, Hung MS. Effect of data standardization on neural network training. Omega 1996;24(4):385–97, https://doi.org/10.1016/0305-0483(96)00010-2.
31. Kim D. Normalization methods for input and output vectors in Backpropagation neural networks. Int J Comput Math 1999;71(1–2):161–71, https://doi.org/10.1080/00207169908804800.
32. Nair V, Hinton GE. Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning (ICML). Association for Computing Machinery; 2010:807–14. 10.1.1.165.6419.
33. Ramachandran P, Zoph B, Le QV. Searching for activation functions 2017. Available at: http://arxiv.org/abs/1710.05941. Accessed November 14, 2019.
34. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: 32nd Int Conf Mach Learn ICML 2015 2015;vol. 1:448–56.
35. LeNail A. NN-SVG: Publication-ready neural network architecture schematics. J Open Source Softw 2019;4(33):747, https://doi.org/10.21105/joss.00747.
36. Pinheiro J, Bates D, DebRoy S, Sarkar D. Linear and nonlinear mixed Effects models (nlme). Packag R Softw Stat Comput CRAN Repos 2011. 0–21. Available at: https://cran.r-project.org/package=nlme. Accessed June 22, 2020.
37. Afara IO, Singh S, Oloyede A. Application of near infrared (NIR) spectroscopy for determining the thickness of articular cartilage. Med Eng Phys 2013;35(1):88–95, https://doi.org/10.1016/j.medengphy.2012.04.003.
38. Marijnissen ACA, Van Roermund PM, Verzijl N, Tekoppele JM, Bijlsma JWJ, Lafeber FPJG. Steady progression of osteoarthritic features in the canine groove model. Osteoarthritis Cartilage 2002;10(4):282–9, https://doi.org/10.1053/joca.2001.0507.
39. Afara IO, Prasadam I, Arabshahi Z, Xiao Y, Oloyede A. Monitoring osteoarthritis progression using near infrared (NIR) spectroscopy. Sci Rep 2017;7(1):11463, https://doi.org/10.1038/s41598-017-11844-3.
40. Buckwalter JA, Mankin HJ. Articular cartilage: degeneration and osteoarthritis, repair, regeneration, and transplantation. Instr Course Lect 1998;47:487–504.
41. Venäläinen MS, Mononen ME, Salo J, Räsänen LP, Jurvelin JS, Töyräs J, *et al*. Quantitative evaluation of the mechanical risks caused by focal cartilage defects in the knee. Sci Rep 2016;6(1):37538, https://doi.org/10.1038/srep37538.
42. Dabiri Y, Li L. Focal cartilage defect compromises fluid-pressure dependent load support in the knee joint. Int J Numer Method Biomed Eng 2015;31(6), https://doi.org/10.1002/cnm.2713.
43. Korhonen RK, Laasanen MS, Töyräs J, Lappalainen R, Helminen HJ, Jurvelin JS. Fibril reinforced poroelastic model predicts specifically mechanical behavior of normal, proteoglycan depleted and collagen degraded articular cartilage. J Biomech 2003;36(9):1373–9, https://doi.org/10.1016/S0021-9290(03)00069-1.
44. Afara IO, Hauta-Kasari M, Jurvelin JS, Oloyede A, Töyräs J. Optical absorption spectra of human articular cartilage correlate with biomechanical properties, histological score and biochemical composition. Physiol Meas 2015;36(9):1913–28, https://doi.org/10.1088/0967-3334/36/9/1913.
45. McGoverin CM, Lewis K, Yang X, Bostrom MPG, Pleshko N. The contribution of bone and cartilage to the near-infrared spectrum of osteochondral tissue. Appl Spectrosc 2014;68(10):1168–75, https://doi.org/10.1366/13-07327.
46. Rutgers M, van Pelt MJP, Dhert WJA, Creemers LB, Saris DBF. Evaluation of histological scoring systems for tissue-engineered, repaired and osteoarthritic cartilage. Osteoarthritis Cartilage 2010;18(1):12–23, https://doi.org/10.1016/j.joca.2009.08.009.
47. Brittberg M, Winalski CS. Evaluation of cartilage injuries and repair. J Bone Joint Surg Am 2003;(Suppl 2):58–69. 85-A Suppl.
48. Malda J, de Grauw JC, Benders KEM, Kik MJL, van de Lest CHA, Creemers LB, *et al*. Of mice, men and elephants: the relation between articular cartilage thickness and body mass. Orgel JPRO, ed. PLoS One 2013;8(2), e57683, https://doi.org/10.1371/journal.pone.0057683.