

Near-Infrared Spectroscopy Enables Arthroscopic Histologic Grading of Human Knee Articular Cartilage

Jaakko K. Sarin, Ph.D., Mithilesh Prakash, Ph.D., Rubina Shaikh, Ph.D.,
Jari Torniainen, Ph.D., Antti Joukainen, M.D., Ph.D., Heikki Kröger, M.D., Ph.D.,
Isaac O. Afara, Ph.D., and Juha Töyräs, Ph.D.

Purpose: To develop the means to estimate cartilage histologic grades and proteoglycan content in ex vivo arthroscopy using near-infrared spectroscopy (NIRS). **Methods:** In this experimental study, arthroscopic NIR spectral measurements were performed on both knees of 9 human cadavers, followed by osteochondral block extraction and in vitro measurements: reacquisition of spectra and reference measurements (proteoglycan content, and three histologic scores). A hybrid model, combining principal component analysis and linear mixed-effects model (PCA-LME), was trained for each reference to investigate its relationship with in vitro NIR spectra. The performance of the PCA-LME model was validated with ex vivo spectra before and after the exclusion of outlying spectra. Model performance was evaluated based on Spearman rank correlation (ρ) and root-mean-square error (RMSE). **Results:** The PCA-LME models performed well (independent test: average $\rho = 0.668$, RMSE = 0.892, $P < .001$) in the prediction of the reference measurements based on in vitro data. The performance on ex vivo arthroscopic data was poorer but improved substantially after outlier exclusion (independent test: average $\rho = 0.462$ to 0.614, RMSE = 1.078 to 0.950, $P = .019$ to .008). **Conclusions:** NIRS is capable of nondestructive evaluation of cartilage integrity (i.e., histologic scores and proteoglycan content) under similar conditions as in clinical arthroscopy. **Clinical Relevance:** There are clear clinical benefits to the accurate assessment of cartilage lesions in arthroscopy. Visual grading is the current standard of care. However, optical techniques, such as NIRS, may provide a more objective assessment of cartilage damage.

Current diagnostic measures of musculoskeletal disorders involve clinical examination and imaging, e.g., x-ray imaging and magnetic resonance imaging,¹ which lack sensitivity to pinpoint localized defects and their severity (structural damage).² Conditions and injuries requiring medical interventions (e.g., a debridement, or cartilage repair) are treated in minimally invasive arthroscopy, often revealing previously

unobserved focal cartilage defects.³ As current arthroscopic measures of evaluation are qualitative and subjective (poor interobserver reliability),⁴ a quantitative arthroscopic tool could be beneficial for the objective evaluation of chondral lesions.

Histology is the gold-standard measure of cartilage degeneration that is used to assess structural and compositional changes within the tissue.⁵ The most

From the Department of Applied Physics, University of Eastern Finland, Kuopio, Finland (J.K.S., R.S., J.To., I.O.A., J.T.); Department of Medical Physics, Medical Imaging Center, Pirkanmaa Hospital District, Tampere, Finland (J.K.S.); A.I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, Kuopio, Finland (M.P.); Department of Orthopedics, Traumatology and Hand Surgery, Kuopio University Hospital, Kuopio, Finland (A.J., H.K.); School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, Australia (I.O.A., J.T.); and Science Service Center, Kuopio University Hospital, Kuopio, Finland (J.T.).

The authors report the following potential conflicts of interest or sources of funding: M.P. reports grants from the Academy of Finland, during the conduct of the study. R.S. reports grants from the MIRACLE project-Horizon 2020 research and innovation programme H2020-ICT-2017-1; grants from the Finnish Cultural Foundation; and grants from State Research Funding, during the conduct of the study. J.To. reports grants from the MIRACLE

project-Horizon 2020 research and innovation programme H2020-ICT-2017-1, during the conduct of the study. I.O.A. reports grants from the Academy of Finland, during the conduct of the study. J.T. reports grants from the MIRACLE project-Horizon 2020 research and innovation programme H2020-ICT-2017-1, during the conduct of the study. Full ICMJE author disclosure forms are available for this article online, as [supplementary material](#).

Received March 9, 2022; accepted July 1, 2022.

Address correspondence to Jaakko K. Sarin, Ph.D., Pirkanmaa Hospital District Tampere, Finland. E-mail: jaakko.sarin@uef.fi

© 2022 THE AUTHORS. Published by Elsevier Inc. on behalf of the Arthroscopy Association of North America. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

2666-061X/22300

<https://doi.org/10.1016/j.asmr.2022.07.002>

commonly used histologic grading systems are the modified Mankin score (0-13),⁶ Osteoarthritis Research Society International (OARSI) grading (0-6),⁵ and International Cartilage Repair Society (ICRS) grading (0-4).⁷ However, the practical clinical value of histology is limited due to invasive tissue extraction, which can greatly jeopardize cartilage integrity. Therefore, a nondestructive technique capable of providing an evaluation similar to histology would be of great value.

Optical techniques, such as near-infrared spectroscopy (NIRS)⁸ and optical coherence tomography,⁹ are capable of beyond-surface evaluation, making these techniques potentially superior to the current standard of visual evaluation during arthroscopy. The techniques use the nonionizing region of light (i.e., no deleterious effects) and the measurement can be performed in seconds. Previously, NIRS has been successfully applied for tissue diagnostics in the laboratory environment¹⁰⁻¹³ and more recently in a few in vivo applications^{14,15} by using custom sterilizable probes similar to the traditional arthroscopic hook. Studies also have evaluated the histologic properties of cartilage via NIRS with moderate-to-strong correlations, especially the Mankin score, which is the most common reference for tissue integrity in animals (bovine¹⁶ and ovine¹⁷) and humans^{16,18,19} in the laboratory environment. In addition, several NIRS studies have focused on the estimation of cartilage composition, especially proteoglycan (PG) content.^{8,12,20,21}

The clinical in situ application of NIRS requires a pre-trained multivariate model (e.g., chemometrics and neural networks), arising from the overlapping nature of spectral peaks in the NIR range, developed using a library of both spectral and reference measurements.²² Lately, convolutional neural networks (CNNs) have outperformed classical chemometrics methods,²³ such as principal component regression. These approaches, however, are yet to account for the inherent dependencies within the data that can arise, for example, from repeated measures from the same subject^{13,24}—especially with valuable human data. Linear mixed-effects (LME) modeling can resolve this dependency problem and is often paired with a dimensionality reduction technique, such as principal component analysis (PCA),²⁵ due to multicollinearity (i.e., a high number of collinear variables in spectral data).

In addition to the impactful role of the modeling approach, spectral preprocessing has a substantial impact on the signal interpretation and the accuracy of the resulting model.²⁶ Currently, the optimal preprocessing pipelines are based on an expert opinion and, to some extent, trial-and-error. To ease the decision on optimal pipeline, open-source preprocessing pipelines, such as *nippy*,²⁷ have become available to explore vast combinations of different preprocessing operators.

The purpose of this study was to develop the means to estimate cartilage histologic grades and PG content in

ex vivo arthroscopy using NIRS. We hypothesized that NIRS would estimate cartilage lesion severity and PG content during an ex vivo arthroscopy.

Methods

In this experimental study, NIR spectra were collected in both ex vivo arthroscopy and in vitro, and the extracted samples were subjected to extensive reference measurements. Ex vivo spectral measurements from several standardized locations (n = 19 [9 in the femur, 8 in the tibia, and 2 in the patella]) were recorded from both knees of human cadavers (N = 9, age = 68.4 ± 7.5 years) by an experienced orthopaedic surgeon (no living cartilage was assessed).¹³ The inclusion criteria for the donors were that they were scheduled for a medical obduction, had no history of knee surgery (also visually verified), and had no infectious risks. The examinations and sample extraction of this study were performed before the medical obduction and post-haste after postmortem (max 4 days) during which time the donors were stored at 4 to 7°C. In the arthroscopies, the inferior extremity of the cadaver was freely movable on a straight table and stabilized with a lateral post on the femur to allow valgus or varus forces created by the surgeon, to apply the setting of normal knee arthroscopy. Anteromedial and anterolateral 1 cm parapatellar interchangeable portals were created for the conventional arthroscope (4 mm, 30° inclination; Karl Storz GmbH & Co, Tuttlingen, Germany) and the novel NIRS probe, respectively. The knee joint was filled with saline using a hand pump. If necessary, an arthroscopic shaver was used to flush the intra-articular space and resect the liposynovia, thereby clearing the visualization of the areas under examination. The NIRS probe was aligned perpendicular and in contact with the cartilage under the examination based on the visualization of the conventional arthroscope. After the measurements, the condyles of the tibia and femur, and patella were harvested, followed by the extraction of cylindrical osteochondral plugs (diameter = 8 mm, the total number of extracted plugs = 303 after exclusion of 39 plugs due to completely eroded cartilage) with a drill punch machine. The plugs were subjected to in vitro spectral measurements and histologic evaluation. The local research ethics committee (decision number 134/2015, Research Ethics Committee of the Northern Savo Hospital District, Kuopio University Hospital, Kuopio, Finland) approved the study. The followed procedures were by the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000.

Near-Infrared Spectroscopy

The spectra were collected both ex vivo and in vitro (Fig 1) with the hardware, consisting of 2 spectrometers

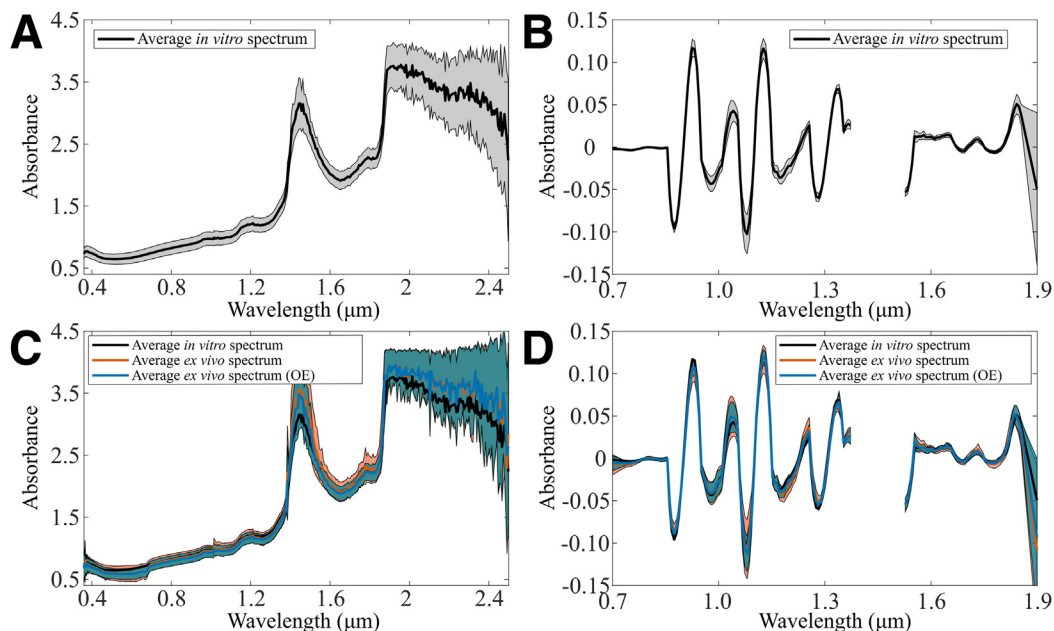


Fig 1. The average (lines) and standard deviation (shaded background) for nonpreprocessed spectra of in vitro (A, C) and ex vivo measurements (C) before and after outlier exclusion (OE). Similarly, preprocessed versions of the spectra are presented (B and D), the optimal preprocessing when predicting ICRS and Mankin score).

(AvaSpec-ULS2048L, $\lambda = 0.35\text{-}1.1\ \mu\text{m}$, $\Delta\lambda = 0.6\ \text{nm}$ and AvaSpec-NIR256-2.5-HSC, $\lambda = 1.0\text{-}2.5\ \mu\text{m}$, $\Delta\lambda = 6.4\ \text{nm}$; Avantes BV, Apeldoorn, Netherlands), a light source (AvaLight-HAL-(S)-Mini, $\lambda = 0.36\text{-}2.5\ \mu\text{m}$, Avantes BV), and an arthroscopic optical probe.²⁸ The custom-made probe resembles the conventional arthroscopic hook and has a total of 114 optical fibers (fiber diameter = $100\ \mu\text{m}$) within the sterilizable stainless-steel housing (outer diameter = $3.25\ \text{mm}$). Thinner fibers (with a smaller minimum bend radius) were used to support the hook-based design.

During the ex vivo spectral measurements, joint cavities were irrigated and distended with saline similarly to routine clinical arthroscopy to enhance the visibility of the articulating surfaces and to ease probe alignment. The probe was aligned perpendicular and in contact with cartilage surface under the guidance of a conventional endoscope ($4\ \text{mm}$, 30° inclination, Karl Storz GmbH & Co.) to prevent spectral saturation from the fluid environment (effective absorber of NIR light). A total of 15 spectra were recorded per location with each spectrum consisting of 10 coadded spectra (acquisition time per location = $2.4\ \text{seconds}$).

The spectral measurements were repeated in vitro on osteochondral plugs extracted from the same ex vivo measurement locations. Contrary to the ex vivo measurements where optimal probe alignment could not be always ensured, in in vitro measurements the sample plugs were fixed into a goniometer (#55-841; Edmund Optics Inc., Barrington, NJ) to achieve reliable contact between the probe and the sample surface.

Spectral Preprocessing

Before preprocessing, data from the spectral region of $0.35\ \text{to}\ 1.10\ \mu\text{m}$ were downsampled to the same resolution as data from the spectral region of $1.0\ \text{to}\ 2.5\ \mu\text{m}$ and combined. An open-source preprocessing module *nippy*²⁷ was used in Python 3.7 to create datasets with different combinations of preprocessing as highlighted by Tornainen et al. The preprocessing options used included smoothing, scatter correction techniques, and trimming. A third-degree Savitzky-Golay filter with 0th (i.e., smoothing), first, and second derivatives with different filter windows ($5\text{-}47$) were evaluated. For scatter correction, standard normal variate and localized standard normal variate (LSNV, windows = $2i$, $i = 1\text{-}6$) were tested. The spectral regions of $0.70\ \text{to}\ 1.90\ \mu\text{m}$, $0.75\ \text{to}\ 1.85\ \mu\text{m}$, $0.70\ \text{to}\ 1.375\ \mu\text{m}$ and $1.525\ \text{to}\ 1.90\ \mu\text{m}$; $0.75\ \text{to}\ 1.375\ \mu\text{m}$ and $1.525\ \text{to}\ 1.85\ \mu\text{m}$ were tested separately. The visible spectral region ($0.35\text{-}0.70\ \mu\text{m}$) was excluded due to the interference originating from the endoscope. The spectral region $1.90\ \text{to}\ 2.50\ \mu\text{m}$ was excluded due to the poor signal-to-noise ratio (high water absorption).

Histology

The osteochondral plugs were halved and one half was decalcified in ethylenediaminetetraacetic acid, cut into $3\text{-}\mu\text{m}$ thick sections ($n = 3$), and stained with Safranin-O (attracted by PGs). Digital densitometry system, consisting of a light microscope (Nikon Microphot-FXA; Nikon Co., Tokyo, Japan) with a monochromatic light source (wavelength $492 \pm 8\ \text{nm}$) and a

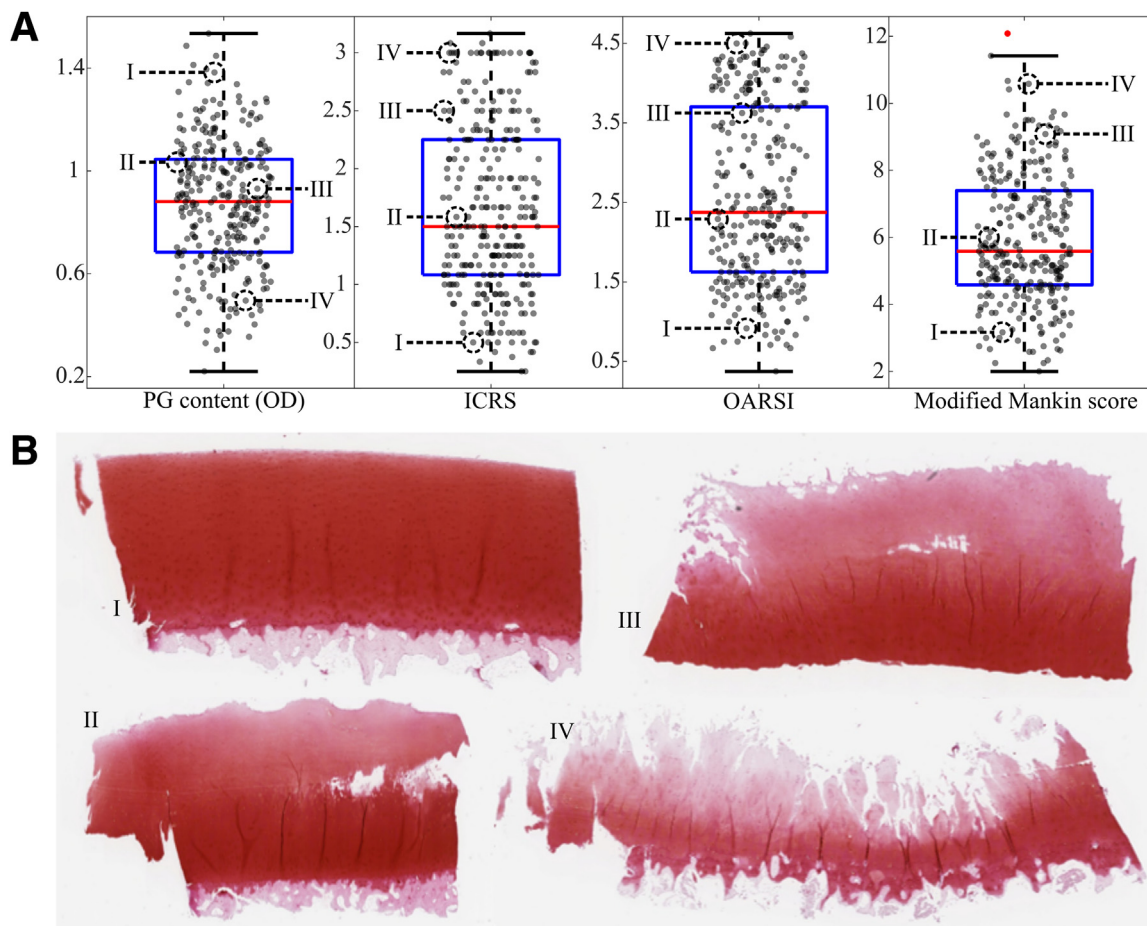


Fig 2. Boxplots along with individual samples (gray) and outliers (red) of reference variables (A). Four representative microscopy images (B, I-IV) are presented with their reference parameter values indicated in subfigure A.

12-bit CCD camera (ORCA-ER; Hamamatsu Photonics K.K., Hamamatsu, Japan), was used to determine sections optical density (OD \sim PG content). The system was calibrated with neutral density filters (0-3.0). The severity of OA was evaluated with 3 histologic grading systems: modified Mankin score,⁶ OARSI grading,⁵ and ICRS grading⁷ (Fig 2). Four independent assessors (M.P., R.S., M.H., and N.H.) scored the sections in a randomized order and the final score for each sample was determined as the average over the 3 sections. The same sections were used for digital densitometry and histologic grading.

Regression Analysis

Before regression analysis, spectral (multivariate) and reference parameter (univariate) outliers were investigated. The *in vitro* spectra were visualized to ensure the absence of hardware-related recording errors. In the preprocessing of *ex vivo* spectra, 10 hardware-related outliers were identified and excluded. For univariate values, the normality of distribution (non-normal for all references) was determined with the one-sample

Kolmogorov–Smirnov test and, thus, any values exceeding 3 median absolute deviations from the reference median were excluded.

PCA was used due to its ability to reduce the high dimensionality and collinearity of datasets, thereby enabling less computationally exhaustive modeling and outlier estimation, as well as reducing the chances of overfitting. PCA scores along with the nested data (i.e., patient, left/right, femur/tibia/patella, and measurement site) were used as inputs for the LME model.

In the modeling, 8 cadavers were assigned as the training set and a single cadaver was assigned as the independent test set. The test set was subsequently changed (9 iterations) until all cadavers were used (also known as nested cross-validation). The *in vitro* models were calibrated and optimized using 10-fold cross-validation with the number of PCA scores limited to 12. The model with the smallest root-mean-square error of cross-validation (RMSECV) was selected to minimize overfitting. The optimal preprocessing pipeline was selected based on the highest median Spearman rank correlation in the independent test set.

Table 1. Performance Metrics as Median (Interquartile Range) for Optimized PCA-LME Model, Classifier, and Ex Vivo Predictions

				PG Content	ICRS	OARSI	Mankin	
In vitro	PCA-LME	Train	ρ	0.739 (0.033)	0.757 (0.012)	0.759 (0.009)	0.734 (0.012)	
			RMSECV	0.185 (0.005)	0.536 (0.014)	0.800 (0.032)	1.381 (0.043)	
			N	8 (2)	8 (2)	7 (1.25)	8 (1.25)	
	Test	ρ	0.583 (0.371)	0.687 (0.253)	0.731 (0.291)	0.669 (0.258)		
		RMSE	0.256 (0.148)	0.628 (0.206)	0.911 (0.282)	1.772 (0.607)		
		SRL	ρ	0.561 (0.420)	0.715 (0.387)	0.671 (0.263)	0.693 (0.252)	
Ex vivo	All	Test	RMSE	0.246 (0.147)	0.602 (0.150)	0.830 (0.368)	1.718 (0.641)	
			ρ	0.421 (0.505)	0.492 (0.274)	0.555 (0.230)	0.379 (0.253)	
			RMSE	0.286 (0.173)	0.825 (0.340)	1.057 (0.201)	2.145 (0.729)	
		Classifier	Train	Accuracy	0.775 (0.027)	0.850 (0.010)	0.850 (0.015)	0.807 (0.009)
				F1-score	0.775 (0.027)	0.859 (0.010)	0.849 (0.014)	0.710 (0.020)
				Test	Accuracy	0.484 (0.042)	0.581 (0.125)	0.538 (0.050)
	SRL	Test	F1-score	0.431 (0.082)	0.533 (0.138)	0.507 (0.021)	0.503 (0.045)	
			OutlierS %	52.7 (9.0)	61.1 (21.6)	41.9 (8.7)	78.9 (4.1)	
			OutlierN %	0 (4.4)	24 (19.6)	0 (15.8)	24 (13.9)	
		SRL	ρ	0.522 (0.481)	0.716 (0.315)	0.630 (0.227)	0.586 (0.383)	
			RMSE	0.274 (0.173)	0.703 (0.247)	0.946 (0.193)	1.876 (0.548)	
			ρ	0.513 (0.481)	0.686 (0.305)	0.599 (0.210)	0.515 (0.368)	
			RMSE	0.274 (0.165)	0.703 (0.256)	0.952 (0.171)	1.876 (0.548)	

Same locations (SRL) are presented to enable better comparison between in vitro and ex vivo performance. OutlierS presents the percentage of spectra excluded by the classifier and OutlierN the percentage of excluded measurement locations (i.e., the location was excluded if all 15 spectra were outliers).

ICRS, International Cartilage Repair Society; OARSI, Osteoarthritis Research Society International; PCA-LME, principal component analysis—linear mixed-effects; PG, proteoglycan; RMSECV, root mean square error of cross-validation.

Arthroscopic Outlier Detection

Due to the relatively narrow joint cavities and limited field of view in the ex vivo measurements, optimal contact between the cartilage surface and probe could not always be ensured. Furthermore, the high water content of cartilage makes the spectral separation between the measurements with good and bad contact especially challenging and, thus, a classifier was trained to identify spectra with non-optimal contact. The performance of classifiers, including fine k-nearest neighbors (kNN), weighted kNN, and support vector machines (SVM), was investigated due to their superior performance in the initial testing. The classifier optimization was performed as follows: the cross-validated PCA-LME model was used to predict the properties based on the ex vivo spectra of the 8 cadavers (same as

in training). If the error between the predicted and reference value was greater than a set threshold ($2 \times \text{RMSECV}$, $3 \times \text{RMSECV}$, or half set as outliers), the label was set to 1 (= outlier), otherwise to 0. These labels along with PCA scores ($N = 12$) of ex vivo spectra were then used to train a 10-fold cross-validated classifier, which was used to classify the remaining independent arthroscopic measurements (one cadaver). The effect of preprocessing pipeline on classifier performance was also investigated. Ultimately, the performance of the same retained locations in both in vitro and ex vivo should be equal—this was used as an indicator (along with classifier accuracy and F1-score) to determine the optimal combination of algorithm, threshold, and preprocessing. The combinations that classified $<10\%$ or $>90\%$ of ex vivo

Table 2. Optimal Preprocessing Pipelines for Different Models With Their Derivative Order (deriv_order) and Window Size (filter_win)

		Scatter Correction	Preprocessing	Spectral Range, μm
PCA-LME	PG content	LSNV_win: 10	deriv_order: 1, filter_win: 23	0.70-1.90
	ICRS	LSNV_win: 6	deriv_order: 2, filter_win: 15	0.70-1.375, 1.525-1.90
	OARSI	LSNV_win: 2	deriv_order: 0, filter_win: 47	0.70-1.375, 1.525-1.90
	Mankin	LSNV_win: 6	deriv_order: 2, filter_win: 15	0.70-1.375, 1.525-1.90
Classifier	PG content	LSNV_win: 6	deriv_order: 0, filter_win: 35	0.70-1.375, 1.525-1.90
	ICRS	LSNV_win: 2	deriv_order: 0, filter_win: 7	0.75-1.375, 1.525-1.85
	OARSI	LSNV_win: 4	deriv_order: 0, filter_win: 11	0.75-1.375, 1.525-1.85
	Mankin	LSNV_win: 6	deriv_order: 0, filter_win: 31	0.70-1.90

ICRS, International Cartilage Repair Society; LSNV, localized standard normal variate; OARSI, Osteoarthritis Research Society International; PCA-LME, principal component analysis—linear mixed-effects; PG, proteoglycan.

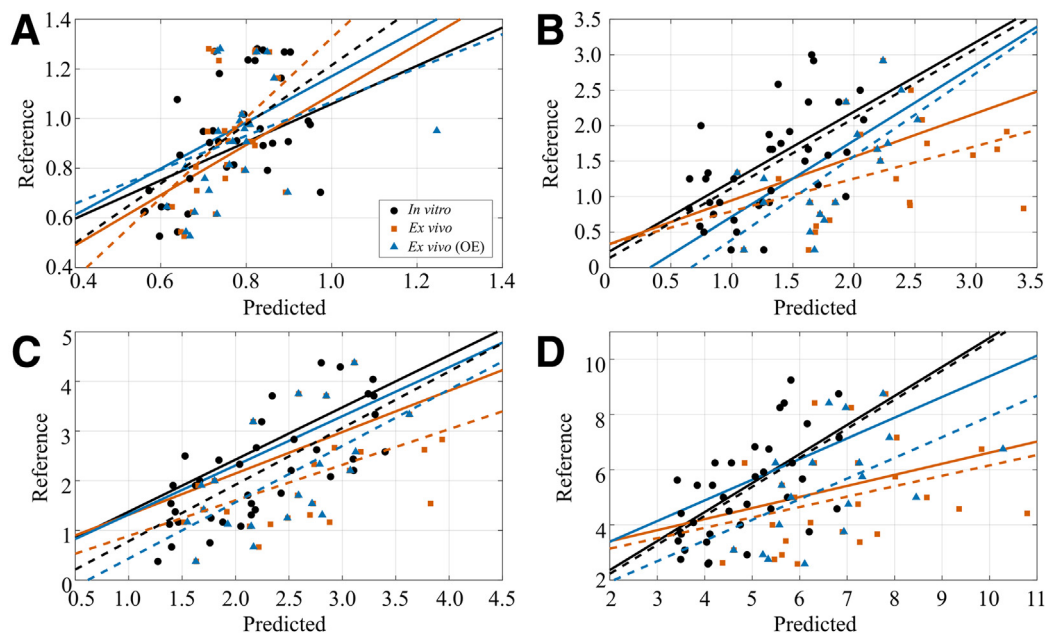


Fig 3. Predicted and reference values for PG content (A), ICRS (B), OARSI (C), and modified Mankin (D) scores are presented for the independent test (1 iteration out of 9), including their linear fits (dashed lines) and the median linear fit of the 9 iterations (solid lines). In vitro, ex vivo, and ex vivo after outlier exclusion (OE) are presented separately. (ICRS, International Cartilage Repair Society; OARSI, Osteoarthritis Research Society International; PG, proteoglycan.)

spectra as outliers were not included as these were not considered realistic.

Statistical Analysis

The PCA-LME model performance was evaluated based on performance in calibration (Spearman rank correlation [ρ], RMSECV) and the independent test (ρ , RMSE). As a statistic, the median was chosen over the average as it is less susceptible to outliers. Model and classifier training were performed in MATLAB (R2020b, MathWorks, Natick, MA). The level of significance was set at $P < .05$. Data of the current study are available from the corresponding author on reasonable request.

Results

The distributions of reference properties were all non-normal with a single outlier detected with the Mankin score (Fig 2). The optimal PCA-LME models had a moderate performance with the histology scores (i.e., ICRS, OARSI, and Mankin) and slightly poorer performance with PG content (Table 1: PCA-LME). The optimal preprocessing pipelines always included scatter correction (i.e., LSNV); furthermore, for the histologic scores, the combination of spectral regions of 0.70 to 1.375 and 1.525 to 1.90 μm was optimal (Table 2). The optimal preprocessing pipelines for ICRS and Mankin scores were identical.

Before outlier classification, the initial performance of the PCA-LME models on ex vivo spectra was assessed (Table 1: All Test). The optimal combination of preprocessing, classifier algorithm, and threshold

substantially improved model performance on the ex vivo performance of the independent test set ($\rho = 0.462$ to 0.614 , $\text{RMSE} = 1.078$ to 0.950 , $P = .019$ to $.008$, Fig 3). To better compare in vitro and ex vivo performance in the test set, a comparison of same retained locations revealed slightly different performance ($\rho = 0.660$, $\text{RMSE} = 0.849$, $P = .001$ vs $\rho = 0.578$, $\text{RMSE} = 0.951$, $P = .018$, respectively). In addition, to estimate model reliability, the prediction error was assessed in 4 classes (i.e., dividing reference ranges to four equally spaced subranges), which revealed the prediction error to be smallest with the 2 middle classes. Although the percentage of outlier spectra (Table 1: OutlierS) was relatively high, the exclusion percentage of measurement locations was substantially lower (Table 1: OutlierN). The variability in the percentage of outliers between reference properties relates to the differences in spectral preprocessing, the performance of the PCA-LME model (RMSECV as a metric for data labeling), and the accuracy of the classifier (prediction reliability).

The optimal preprocessing pipelines for classification systematically included scatter correction (i.e., LSNV) and smoothing (i.e., no derivative preprocessing, Table 2). The outlier thresholds of $3 \times \text{RMSECV}$ and 50% were optimal for the histology scores and PG content, respectively. The optimal classifiers for PG content, ICRS, OARSI, and modified Mankin score were fine-kNN, SVM, weighted kNN, and fine-kNN, respectively. Overall, none of the algorithms performed systematically better than the others and SVM classified more spectra as outliers than the kNN algorithms. We

also investigated classifier performance when using the same preprocessing pipeline as in the modelling; however, the *ex vivo* performance was systematically worse compared with the optimized pipeline ($\rho = 0.452$ to 0.614 , $RMSE = 1.032$ to 0.950 , $P = .027$ to $.008$).

Outlier exclusion decreased the standard deviation of *ex vivo* spectra by 23.3% and also decreased the absorption at $1.4 \mu\text{m}$ (water peak, Fig 1), depicting the exclusion of spectra with saline interference (i.e., more water). Most importantly, the visualization of *ex vivo* spectra before and after outlier exclusion confirmed that extreme spectra were excluded.

Discussion

In this study, we demonstrate that NIRS is capable of estimating cartilage lesion severity and PG content during an *ex vivo* arthroscopy, thus validating the hypothesis. The *in vitro* performance of PCA-LME models was moderate to strong and the *ex vivo* arthroscopic performance was slightly poorer, which was nevertheless substantially improved by excluding outlying spectra. We therefore, believe the technique could provide previously unobtainable diagnostic information during arthroscopic surgery.

NIRS has been previously applied to estimate cartilage histologic scores in animals²⁹ and humans^{16,18,19} *in vitro*. In addition, ICRS scores based on visual evaluation in arthroscopy have been associated with *in vitro*^{19,30} and *in vivo*³¹ NIRS measurements. The human studies^{16,18,19,30,31} have employed varying spectral ranges and analysis techniques with none accounting for the spatial dependency caused by multiple measurements per subject or validating model performance by independent testing. McGoverin et al.¹⁶ reported a similar cross-validation error (PLS: calibration $R^2 = 0.84$, $RMSECV = 1.3$) on modified Mankin score with the spectral region $1.11\text{--}2.27 \mu\text{m}$, whereas Afara et al.¹⁸ presented slightly poorer performance (PLS: calibration $R^2 = 0.83$, $RMSECV = 1.6$) on Mankin score using the spectral region of $0.40\text{--}1.10 \mu\text{m}$. In addition, Stumpfe et al.¹⁹ demonstrated moderate correlation (PLS: $\rho = 0.54$) between the spectral region 0.95 to $1.70 \mu\text{m}$ and Mankin score but no error statistics or any means of cross-validation were reported. Marticke et al.³⁰ reported a negative relationship ($\rho = -0.47$) between arthroscopic ICRS score and a spectral characteristic value (estimated as a linear combination of 2 spectral peaks), whereas Spahn et al.³¹ reported a positive association with spectral ratio (ratio of the same 2 peaks). Direct comparison of the aforementioned studies is limited due to their simplistic analysis approach. Overall, the findings of this study agree with *in vitro* performance of previous studies and extend technique validity and application for *in vivo* application.

Previous studies associating PG content with NIR spectra have focused on bovine^{20,21} and equine

cartilage.^{8,12} Few studies have also investigated the PG content of engineered cartilage constructs.³²⁻³⁴ The bovine studies of Brown et al.²⁰ and Afara et al.²¹ both used the spectral region 0.8 to $2.5 \mu\text{m}$, PLS regression, and assessed PG content after various stages of artificial depletion. Brown et al.²⁰ presented a significant distinction between normal and PG depleted samples, whereas Afara et al.²¹ demonstrated a superior cross-validated performance (calibration $R^2 = 93.76$, $RMSECV = 0.573$) based on data within the spectral regions 0.8 to $1.0 \mu\text{m}$ and 1.55 to $1.84 \mu\text{m}$. The calibration correlation of Afara et al.²¹ is substantially greater than that found in this study ($\rho = 0.739$, $RMSECV = 0.185$), whereas the cross-validated errors are similar, considering the ranges of OD values ($0.2\text{--}1.5$) and PG scores ($0\text{--}4$) in this and their study,²¹ respectively. The higher correlation could be explained by their use of leave-one-out (vs k-fold) cross-validation and smaller sample size. Sarin et al.⁸ applied a similar cross-validation scheme and also validated the performance for arthroscopy on equine. The validation performance of their CNN ($\rho = 0.691$, $RMSECV = 0.274$) was fairly similar. Furthermore, the validation performance on arthroscopic spectra was inferior to the aforementioned performance, similar to the present study. Most importantly, it's evident that prediction of PG content is possible and reproducible both for equine and human cartilage; although, their thickness varies greatly ($0.14\text{--}1.36 \text{ mm}$ and $1.23\text{--}5.90 \text{ mm}$, respectively^{8,13}). Due to the superior performance of sophisticated machine-learning approaches over conventional chemometrics techniques,³⁵ the combination of CNN and protocols accounting for data dependency could benefit future studies.

The NIRS literature of the aforementioned publications provides the groundwork for clinical adaptation of NIRS in cartilage assessment. Nowadays, arthroscopic cartilage evaluation usually relies on visual estimation and instrument palpation, which has limited reliability in cartilage defect grading.^{4,36} Here, NIRS enables the estimation of PG content and defect severity in a situation resembling that of arthroscopic surgery without the need for destructive sample extraction. This nondestructive evaluation could enable the objective classification of cartilage defects and further enable the monitoring of potential treatment options in cartilage repair. In addition, the technique enables the mapping of the extent of traumatic cartilage injuries that would assist in the treatment decision-making and open new possibilities to evaluate the prognosis of a cartilage defect in joint trauma. For example, post-traumatic cartilage degeneration is a well-known consequence of an anterior cruciate ligament rupture, even after a successful anterior cruciate ligament reconstruction.^{37,38} NIRS could enable the evaluation of this post-traumatic process in the knee and other joints.

Outlier detection is essential in both the initial model training and validation, followed by ensuring the validity of new data (i.e., no unreliable predictions). For this study, median-based statistics were chosen due to their inherent property of being less sensitive to outliers (compared with average). Estimation of spectral (multivariate) outliers is challenging and, thus, dimension reduction techniques, such as PCA, have been popular.^{8,13,15} Furthermore, due to the high water content of cartilage, the spectral separation between the measurements with poor and good contact is especially challenging and requires further validation (i.e., a designated study). Therefore, the spectral outliers highly resemble the nonoutlier spectra. In 2 studies by Sarin et al.,⁸ a 3-dimensional volume was created based on PCA scores of in vitro data and arthroscopic spectra falling outside this volume were deemed outliers. The exclusion percentages were 4.5% to 23.5% and 3.1%.^{8,15} Prakash et al.¹³ used a similar outlier exclusion as in this study with similar accuracy of 50% to 70% in the test set but a relatively lower percentage of outliers (33%). Interestingly, after outlier exclusion, their correlation coefficients improved but error variance substantially increased. In this study, both the error and its variance decreased after outlier exclusion. Several studies^{14,31,39} focusing on in vivo spectral measurements do not provide any means for outlier exclusion, although highlighting the difficulty of in vivo measurements.³⁹ In future studies, outlier detection may also benefit from the latest innovations, such as extended isolation forests.⁴⁰

Limitations

Some limitations were evident in this study. The number of cadavers was relatively low and could lead to limited range (nonrepresentative) of reference properties; however, several locations were assessed in extensive laboratory measurements, thereby sufficiently increasing the number of observations. A relatively high percentage of outlier spectra depicts the challenge in probe alignment within the joint cavity to ensure optimal contact with the cartilage surface, as also highlighted by Spahn et al.³⁹ However, the percentage of outlier locations (i.e., all 15 spectra excluded) was substantially lower, indicating that a successful measurement was acquired in most of the locations.

Conclusions

NIRS is capable of nondestructive evaluation of cartilage integrity (i.e., histologic scores and PG content) under similar conditions as in clinical arthroscopy.

Acknowledgments

Tarja Huhta and Linda Rantamaa are acknowledged for preparing the histologic sections. Nina Hänninen,

M.Sc., and Miitu Honkanen, M.Sc., are acknowledged for assistance with histologic scoring.

References

1. 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* 2015;50:712-719.
2. Porter M, Shadbolt B. Accuracy of standard magnetic resonance imaging sequences for meniscal and chondral lesions versus knee arthroscopy. A prospective case-controlled study of 719 cases. *ANZ J Surg* 2021;91:1284-1289.
3. von Engelhardt LV, Lahner M, Klusmann A, Bouillon B, Dávid 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 Musculoskelet Disord* 2010;11:75.
4. 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;129:1117-1121.
5. Pritzker KPH, Gay S, Jimenez SA, Ostergaard K, Pelletier JP, Revell K, et al. Osteoarthritis cartilage histopathology: Grading and staging. *Osteoarthr Cartil* 2006;14:13-29.
6. van der Sluijs JA, Geesink RGT, van der Linden AJ, Bulstra SK, Kuyer R, Drukker J. The reliability of the mankin score for osteoarthritis. *J Orthop Res* 1992;10:58-61.
7. Brittberg M, Winalski CS. Evaluation of cartilage injuries and repair. *J Bone Joint Surg Am* 2003;85-A:58-69 (suppl).
8. 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;47:1815-1826.
9. Pailhé R, Mounier A, Boisson B, Rouchy RC, Voros S, Chipon E, et al. Qualitative and quantitative assessment of cartilage degeneration using full-field optical coherence tomography ex vivo. *Osteoarthr Cartil* 2018;26:285-292.
10. Torniainen J, Ristaniemi A, Sarin JK, Mikkonen S, Afara IO, Stenroth L, et al. Near infrared spectroscopic evaluation of ligament and tendon biomechanical properties. *Ann Biomed Eng* 2019;47:213-222.
11. Ala-Myllymäki J, JTJ Honkanen, Töyräs J, Afara IO. Optical spectroscopic determination of human meniscus composition. *J Orthop Res* 2016;34:270-278.
12. Sarin JK, Rieppo L, Brommer H, Afara IO, Saarakkala S, Töyräs J. Combination of optical coherence tomography and near infrared spectroscopy enhances determination of articular cartilage composition and structure. *Sci Rep* 2017;7:10586.
13. 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. *Osteoarthr Cartil* 2019;27:1235-1243.
14. Spahn G, Felmet G, Hofmann GO. Traumatic and degenerative cartilage lesions: Arthroscopic differentiation

- using near-infrared spectroscopy (NIRS). *Arch Orthop Trauma Surg* 2013;133:997-1002.
15. Sarin JK, te Moller NCR, Mohammadi A, Prakash M, Torniainen J, Brommer H, et al. Machine learning augmented near-infrared spectroscopy: In vivo follow-up of cartilage defects. *Osteoarthr Cartil* 2021;29:423-432.
 16. 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;68:1168-1175.
 17. Spahn G, Plettenberg H, Nagel H, Kahl E, Klinger HM, Mückley T, et al. Evaluation of cartilage defects with near-infrared spectroscopy (NIR): An ex vivo study. *Med Eng Phys* 2008;30:285-292.
 18. 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;36:1913-1928.
 19. Stumpfe ST, Pester JK, Steinert S, Marintschev I, Plettenberg H, Aurich M, et al. Is there a correlation between biophotonical, biochemical, histological, and visual changes in the cartilage of osteoarthritic knee-joints? *Muscles Ligaments Tendons J* 2013;3:157-165.
 20. Brown CP, Bowden JC, Rintoul L, Meder R, Oloyede A, Crawford RW. Diffuse reflectance near infrared spectroscopy can distinguish normal from enzymatically digested cartilage. *Phys Med Biol* 2009;54:5579-5594.
 21. Afara IO, Moody H, Singh S, Prasadam I, Oloyede A. Spatial mapping of proteoglycan content in articular cartilage using near-infrared (NIR) spectroscopy. *Biomed Opt Express* 2015;6:144-154.
 22. Houhou R, Bocklitz T. Trends in artificial intelligence, machine learning, and chemometrics applied to chemical data. *Anal Sci Adv* 2021;2:128-141.
 23. Cui C, Fearn T. Modern practical convolutional neural networks for multivariate regression: Applications to NIR calibration. *Chemom Intell Lab Syst* 2018;182:9-20.
 24. Prakash M, Sarin JK, Rieppo L, Afara IO, Töyräs J. Accounting for spatial dependency in multivariate spectroscopic data. *Chemom Intell Lab Syst* 2018;182:166-171.
 25. Abdi H, Williams LJ. *Principal component analysis* Wiley Interdiscip Rev Comput Stat, 2010;433-459;2, 2010.
 26. Engel J, Gerretzen J, Szymańska E, Jansen JJ, Downey G, Blanchet L, et al. Breaking with trends in pre-processing? *TrAC Trends Anal Chem* 2013;50:96-106.
 27. Torniainen J, Afara IO, Prakash M, Sarin JK, Stenroth L, Töyräs J. Open-source python module for automated preprocessing of near infrared spectroscopic data. *Anal Chim Acta* 2020;1108:1-9.
 28. 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;8:13409.
 29. UP Palukuru, McGoverin CM, Pleshko N. Assessment of hyaline cartilage matrix composition using near infrared spectroscopy. *Matrix Biol* 2014;38:3-11.
 30. Marticke JK, Hösselbarth A, Hoffmeier KL, Marintschev I, Otto S, Lange M, et al. How do visual, spectroscopic and biomechanical changes of cartilage correlate in osteoarthritic knee joints? *Clin Biomech (Bristol, Avon)* 2010;25:332-340.
 31. Spahn G, Plettenberg H, Kahl E, Klinger HM, Mückley T, Hofmann GO. Near-infrared (NIR) spectroscopy. A new method for arthroscopic evaluation of low grade degenerated cartilage lesions. Results of a pilot study. *BMC Musculoskelet Disord* 2007;8:47.
 32. Baykal D, Irrechukwu O, Lin P-C, Fritton K, Spencer RG, Pleshko N. Nondestructive assessment of engineered cartilage constructs using near-infrared spectroscopy. *Appl Spectrosc* 2010;64:1160-1166.
 33. McGoverin CM, Hanifi A, UP Palukuru, Yousefi F, Glenn PBM, Shockley M, et al. Nondestructive assessment of engineered cartilage composition by near infrared spectroscopy. *Ann Biomed Eng* 2016;44:680-692.
 34. Hanifi A, Palukuru U, McGoverin C, Shockley M, Frank E, Grodzinsky A, et al. Near infrared spectroscopic assessment of developing engineered tissues: Correlations with compositional and mechanical properties. *Analyst* 2017;142:1320-1332.
 35. 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.
 36. 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;131:377-381.
 37. Carbone A, Rodeo S. Review of current understanding of post-traumatic osteoarthritis resulting from sports injuries. *J Orthop Res* 2017;35:397-405.
 38. Diemer F, Zebisch J, Saueressig T. Consequences of anterior cruciate ligament rupture: A systematic umbrella review. *Sportverletzung-Sportschaden* 2022;36:18-37.
 39. Spahn G, Klinger HM, Baums M, Hoffmann M, Plettenberg H, Kroker A, et al. Near-infrared spectroscopy for arthroscopic evaluation of cartilage lesions: Results of a blinded, prospective, interobserver study. *Am J Sports Med* 2010;38:2516-2521.
 40. Hariri S, Kind MC, Brunner RJ. Extended isolation forest. *IEEE Trans Knowl Data Eng* 2018;33:1479-1489.