



Arthroscopic Determination of Cartilage Proteoglycan Content and Collagen Network Structure with Near-Infrared Spectroscopy

JAAKKO K. SARIN ^{1,2}, OLLI NYKÄNEN,¹ VIRPI TITU,³ IRINA A. D. MANCINI,^{4,5} HAROLD BROMMER,⁴ JETZE VISSER,⁶ JOS MALDA ^{4,5,6}, P. RENÉ VAN WEEREN,^{4,5} ISAAC O. AFARA,¹ and JUHA TÖYRÄS^{1,2,7}

¹Department of Applied Physics, University of Eastern Finland, Kuopio, Finland; ²Diagnostic Imaging Center, Kuopio University Hospital, Kuopio, Finland; ³Institute of Biomedicine, Anatomy, University of Eastern Finland, Kuopio, Finland; ⁴Department of Equine Sciences, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands; ⁵Regenerative Medicine Utrecht, Utrecht University, Utrecht, The Netherlands; ⁶Department of Orthopaedics, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands; and ⁷School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, Australia

(Received 14 December 2018; accepted 24 April 2019; published online 6 May 2019)

Associate Editor Dan Elson oversaw the review of this article.

Abstract—Conventional arthroscopic evaluation of articular cartilage is subjective and insufficient for assessing early compositional and structural changes during the progression of post-traumatic osteoarthritis. Therefore, in this study, arthroscopic near-infrared (NIR) spectroscopy is introduced, for the first time, for *in vivo* evaluation of articular cartilage thickness, proteoglycan (PG) content, and collagen orientation angle. NIR spectra were acquired *in vivo* and *in vitro* from equine cartilage adjacent to experimental cartilage repair sites. As reference, digital densitometry and polarized light microscopy were used to evaluate superficial and full-thickness PG content and collagen orientation angle. To relate NIR spectra and cartilage properties, ensemble neural networks, each with two different architectures, were trained and evaluated by using Spearman's correlation analysis (ρ). The ensemble networks enabled accurate predictions for full-thickness reference properties (PG content: $\rho_{in\ vitro, Val} = 0.691$, $\rho_{in\ vivo} = 0.676$; collagen orientation angle: $\rho_{in\ vitro, Val} = 0.626$, $\rho_{in\ vivo} = 0.574$) from NIR spectral data. In addition, the networks enabled reliable prediction of PG content in superficial (25%) cartilage ($\rho_{in\ vitro, Val} = 0.650$, $\rho_{in\ vivo} = 0.613$) and cartilage thickness ($\rho_{in\ vitro, Val} = 0.797$, $\rho_{in\ vivo} = 0.596$). To conclude, NIR spectroscopy could enhance the detection of initial cartilage degeneration and thus enable demarcation of the boundary between healthy and compromised cartilage tissue during arthroscopic surgery.

Keywords—Equine, Deep learning, Neural networks, Arthroscopy, Mosaicplasty, Osteoarthritis, Post-traumatic osteoarthritis.

INTRODUCTION

Osteoarthritis (OA) is a prevalent disease characterised by joint pain, restricted mobility, and instability of joint.⁹ Idiopathic OA is especially common among the elderly population and may be characterized by erosion of articular cartilage—the connective tissue that enables smooth and frictionless motion of the joints. As articular cartilage is aneural and avascular tissue,²⁹ initial symptoms (e.g., joint pain) may occur in the later stages of the disease, making cartilage repair challenging. Development of OA can also be triggered by traumatic joint injuries.⁸ In contrast to idiopathic OA, post-traumatic OA (PTOA) can affect people of all ages.³⁵ Local traumatic cartilage lesions will cause the surrounding tissue to experience higher stresses and strains,³⁸ thus making the tissue more vulnerable to degeneration. Detection and intervention at the early disease stages (superficial proteoglycan (PG) loss and collagen damage) could be valuable for halting disease progression. Current methods to treat small chondral defects include arthroscopic debridement and lavage, osteochondral grafting, and autologous chondrocyte implantation.³⁷ Although tissue engineering has advanced rapidly in the last decade, no engineered cartilage currently exists that matches the properties of native cartilage.³⁷ In the event of insufficient repair, cartilage surrounding injury sites slowly degrades and will often exhibit signs of early to mid-stage degeneration (similar to those associated with early OA).¹⁵

Conventional imaging methods, such as radiography and magnetic resonance imaging, are well-suited

Address correspondence to Jaakko K. Sarin, Department of Applied Physics, University of Eastern Finland, Kuopio, Finland. Electronic mail: jaakko.sarin@uef.fi

for diagnosis of late stages of OA.³⁹ These methods are, however, insufficient for detecting the early signs of cartilage damage. Ligament tears and other traumatic injuries often require arthroscopic repair surgery. Currently, arthroscopic evaluation relies on visual assessment and scoring of cartilage lesions using established systems (e.g., International Cartilage Repair Society (ICRS)⁷) and manual probing with an arthroscopic hook. These methods are highly subjective⁶; thus, more quantitative techniques could be of great clinical significance.

Several quantitative techniques, including near-infrared spectroscopy (NIRS, 0.7–2.5 μm) and mid-infrared spectroscopy (MIRS, 2.5–15 μm), have been suggested for evaluating cartilage properties, such as biomechanical properties,^{25,27} cartilage thickness,²⁵ and biochemical composition,² in several *in vitro* studies. Clinical application of NIRS has only been presented in few studies^{31,32}; however, these studies are limited by the use of simplistic univariate analysis techniques, in which the ratio of two spectral peaks at 1175 and 1425 nm is evaluated.^{31,32} These spectroscopic techniques enable non-destructive *in situ* evaluation without the need for sample extraction and are sensitive to specific bond vibrations common in biological materials.¹⁰ In conventional diffuse reflectance spectroscopy, samples are irradiated with light and the reflected and scattered light is collected and analysed. Changes in tissue composition contribute to the magnitude of light absorption, whereas tissue structure contributes to the magnitude of reflection and scattering. Although the MIR spectral region is more specific with relatively stronger absorption bands,¹⁰ the technique suffers from poor penetration depth into biological tissues.²⁰ On the other hand, the NIR spectral region suffers from overlapping spectral features,¹⁰ making conventional univariate analysis ineffective. Currently, the availability of computational power and adaptation of state-of-the-art algorithms, such as neural networks (NNs), have enabled modelling of the relationship between NIR spectra and reference properties.²⁷

Neural networks (NNs) are analytical techniques that mimic the function of the brain and consist of several layers of neurons. The input layer receives the data and passes it on to hidden layers, whereas the output layer defines network output. In the neurons of the hidden layers, the inputs are weighted and processed through different activations (e.g., rectified linear units (ReLU¹⁸)). With regression problems, linear activation is often used in the output layer. The aim of the network is to optimize the hyper-parameters (weights and biases) to minimize network error. Even shallow NNs have shown superiority in modelling the relationships between spectra and reference properties

over principal component analysis (PCA) and partial least squares regression (PLSR)—the commonly applied analytical approaches in chemometrics.^{12,19} Deep learning applications, such as image recognition and classification, conventionally rely on convolutional NN (CNN) to process image data. For successful adaptation of deep learning, large training data are required for the modelling as limited data can often lead to overfitting. Several techniques have been introduced to minimize overfitting, including cross-validation,³⁶ dropout layers,³⁴ and early stopping.⁴¹ In cross-validation, an independent subgroup of data is withheld from the training data to monitor its prediction performance and thus to ensure a well-generalizable model. A dropout layer randomly deactivates neurons during training to make each neuron to learn broader features, whereas early stopping halts the training if no improvement is achieved after a certain number of iterations. Additional techniques for enhancing NN performance are data augmentation,²² in which the amount of training data is artificially increased (e.g., by the creation of slightly altered copies of the original data), and reduction of learning rate, in which learning rate is reduced to ensure finding the optimal solution.

We hypothesize that arthroscopic NIRS, combined with deep learning, can reliably predict cartilage PG content and collagen orientation angle in full-thickness and superficial regions of cartilage during *in vivo* arthroscopy. To test the hypothesis, *in vivo* and *in vitro* NIR spectra were acquired from tissue surrounding experimental cartilage repair sites in equine joints and compared to those acquired from matching anatomical locations in healthy equine joints. The surrounding tissue was investigated as the repair attempts were deemed unsuccessful and also to improve the currently challenging diagnosis of early cartilage degeneration. As reference, cartilage PG content and collagen orientation angle were determined with digital densitometry (DD) and polarized light microscopy (PLM), respectively. In addition, we applied one-dimensional CNNs (1D-CNNs) to investigate the relationship between cartilage NIR spectra and reference properties. Relative to earlier cartilage NIRS studies,^{14,26,32} the present study utilises a wider spectral region, adopts a more sophisticated analytical technique, and most importantly, demonstrates the feasibility of the NIRS for *in vivo* applications.

MATERIALS AND METHODS

On the medial femoral trochlear ridges of both femoropatellar joints of adult Shetland ponies ($N = 7$, 6 females and 1 male, Age = 8.8 ± 3.5 years), two

cylindrical ($d = 10$ mm) chondral lesions were surgically created, resulting in a total of 28 lesions.²⁷ Lesions were filled with four different experimental repair techniques (three varieties of gelatin methacryloyl (GelMA) hydrogels or fibrin glue) to investigate their potential—please refer to our earlier paper for further details on the repair techniques (open access).²⁷ The positions of these four experimental repairs were randomized equally between the four predetermined anatomical locations (left proximal, left distal, right proximal, and right distal femoral trochlea) to account for any variability influenced by the anatomical location. The ponies were sacrificed after 12 months follow-up and the tissue surrounding the repair sites was evaluated using a conventional arthroscope and a novel arthroscopic NIRS probe. Subsequently, samples were extracted after removing the skin and overlying tissues of the joint and frozen (at -20°C) until required for further analysis. No substantial changes in cartilage structure or composition are expected by subjecting the samples for a freeze–thaw cycle.²³ In addition, samples from matching anatomical location were extracted from the joints of healthy adult Shetland ponies ($N_{\text{control}} = 3$, Age = 10.3 ± 4.7 years) to enable comparison between the two groups: experimental and control ponies. The Ethics Committee of Utrecht University for Animal Experiments in compliance with the Institutional Guidelines on the Use of Laboratory Animals approved the study (Permission DEC 2014.III.11.098). The measurements were carried out in a surgical theatre at the Department of Equine Sciences, Utrecht University, The Netherlands. The control ponies were acquired from a slaughterhouse in Nijkerk, The Netherlands and thus no gender information was available. Unlike in humans,¹³ there are thus far for the equine species no indications for gender-related differences in the susceptibility to orthopaedic diseases, such as OA, nor in differences related to the healing capacity of articular injuries.

Arthroscopic NIR spectra ($n = 108$) were acquired by an experienced board-certified equine surgeon (with experience of > 500 arthroscopies, Diplomate European College of Veterinary Surgeons) under the guidance of a traditional arthroscope (4 mm, 30° inclination, Synergy HD3, Arthrex, Naples, FL, USA). The NIRS system consisted of spectrometers (AvaSpec-ULS2048L, $\lambda = 0.35\text{--}1.1$ μm , resolution = 0.6 nm and AvaSpec-NIR256-2.5-HSC, $\lambda = 1.0\text{--}2.5$ μm , resolution = 6.4 nm, Avantes BV, Apeldoorn, The Netherlands), a light source (AvaLight-HAL-(S)-Mini, $\lambda = 0.36\text{--}2.5$ μm , Avantes BV), and the custom-designed arthroscopic probe manufactured by Avantes BV. The reusable stainless-steel probe ($d = 3.25$ mm) is sterilisable in an autoclave at 121°C with the probe tip window ($d = 2$ mm) con-

taining 114 optical fibres ($d = 100$ μm) with 14 fibres used to collect the reflected and scattered light. Eight locations surrounding the cartilage repair sites were measured (Fig. 1) with the NIRS probe held in perpendicular contact with the cartilage surface. At each measurement point, 15 spectra were recorded, each being the average of ten successive spectra; the total duration of data acquisition was 2.4 s per measurement location. In addition, the traditional arthroscope and palpation hook were utilized to evaluate the measurement locations according to the ICRS scoring system.⁷ Arthroscopic images were recorded during the measurements to enable reliable location tracking. For joint distension, 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.

The NIR spectra ($n = 236$; experimental and control groups) were re-acquired *in vitro* under similar conditions as during arthroscopy procedure, i.e., including NIRS hardware, immersion solution, and temperature, apart from using the conventional arthroscope for navigation. Three successive spectra, each being an average of 10 spectra, were acquired at each measurement location. The different acquisition protocols for *in vivo* and *in vitro* measurements were utilized to account for operator-based movement and hardware-related variation, respectively.

In addition to spectral measurement, cartilage thickness was determined by imaging with optical coherence tomography (OCT, $\lambda = 1305 \pm 55$ nm, axial resolution < 20 μm , lateral resolution 25–60 μm ; Ilumien PCI Optimization System, St. Jude Medical, St. Paul, MN, USA).²⁷

Histological Processing and Imaging

Histological sections were prepared for each NIRS measurement location (Fig. 1a) by the process of fixing samples in formalin, subsequent decalcification in EDTA, and embedding in paraffin. For each NIRS measurement location, three sections were prepared with a microtome for DD (thickness = 3 μm) and PLM (thickness = 5 μm) imaging.²⁶ In the analysis, full-thickness and superficial (first 25%) cartilage PG content and collagen orientation angle were determined. The first 25% of cartilage was chosen to represent a layer in which collagen orientation changes from parallel orientation towards perpendicular orientation relative to cartilage surface.²⁹ This layer represents the conventional superficial and middle zones that are traditionally determined based on profiles of collagen orientation and optical retardance; however, due to the disruption of the collagen network in some of the present samples, this was not always feasible.

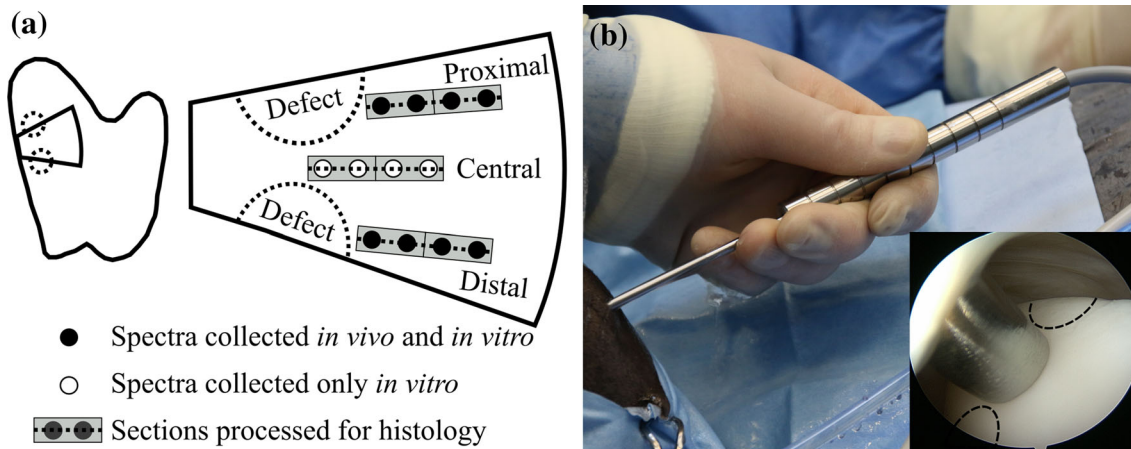


FIGURE 1. (a) The measurement locations (black and white circles) on an osteochondral sample along with the locations (grey rectangles) from which tissues were extracted for histological analysis (dashed line on a grey rectangle). On each measurement line (proximal, central, and distal), the measurement location closest to a defect was denoted as location 1 and the farthest as location 4. (b) The novel arthroscopic NIRS probe during *in vivo* spectral acquisition with the probe tip in contact with cartilage surface (inset, at Distal location 3). The distance between the locations of spectral acquisitions was equal to the outer diameter ($d = 3.25$ mm) of the NIRS probe.

To determine cartilage PG content *via* optical density (OD), the sections were stained with safranin-O and imaged using a DD system (Fig. 2). The system consists of a light microscope (Nikon Microphot-FXA, Nikon Co., Tokyo, Japan), equipped with a monochromatic light source and a 12-bit CCD camera (ORCA-ER, Hamamatsu Photonics K.K., Hamamatsu, Japan). The system was calibrated with filters having OD ranging from 0 to 3. In the analysis, the average OD profile of each location was determined by extracting depth-wise profiles perpendicular to the cartilage surface, interpolating them to 500 points, and averaging the interpolated profiles.

An Abrio PLM system (CRi, Inc., Woburn, MA, USA), mounted on a conventional light microscope (Nikon Diaphot TMD, Nikon Inc., Shinagawa, Tokyo, Japan), was used to determine collagen orientation in the samples (Fig. 2). The system consists of a green bandpass filter, a circular polarizer, and a computer-controlled analyser, composed of two liquid crystal polarizers and a CCD camera. In the orientation images, 0° corresponds with an orientation parallel to the cartilage surface and 90° with a position perpendicular to the cartilage surface. Similarly to the DD analysis, the depth-wise profiles perpendicular to cartilage surface were determined, followed by interpolation and averaging.

Spectral Pre-processing

Spectral measurements often include environmental or hardware-related noise; therefore, signal pre-processing is essential for ensuring reliable data for mod-

elling. The acquired spectra were pre-processed with several techniques, including smoothing, 1st derivative, and 2nd derivative pre-processing. Due to using two detectors with varying resolution, the wavelength regions were pre-processed separately with a 3rd degree Savitzky-Golay²⁸ algorithm using 139 points ($0.08 \mu\text{m}$, spectrometer: AvaSpec-ULS2048L) and 41 points ($0.26 \mu\text{m}$, spectrometer: AvaSpec-NIR256). These window lengths ensured sufficient noise reduction without any substantial loss of essential information.

Outlier Detection

In our previous study,²⁷ arthroscopic spectra were found to have contributions from the arthroscopic light source ($0.40\text{--}0.75 \mu\text{m}$); therefore, a spectral region of $0.40\text{--}0.80 \mu\text{m}$ was discarded from the analysis. To address the possibility of non-optimal contact between NIRS probe and cartilage surface during arthroscopic measurements, a minimum volume ellipsoid (MVE) approach was adopted. PCA was utilized to calculate the first three principal scores of pre-processed *in vitro* spectra, followed by calculation of scores for *in vivo* spectra by projecting them to the first space. As *in vitro* spectra (green, Figs. 3a and 3b) were always recorded with a reliable contact between the probe and tissue in a controlled environment, these scores were utilized to calculate the MVE (Fig. 3a). *In vivo* spectra falling outside the ellipsoid (red, Figs. 3a and 3c) were classified as outliers. In the analysis, the radii of MVE were uniformly altered to investigate their effect on NNs' prediction performance.

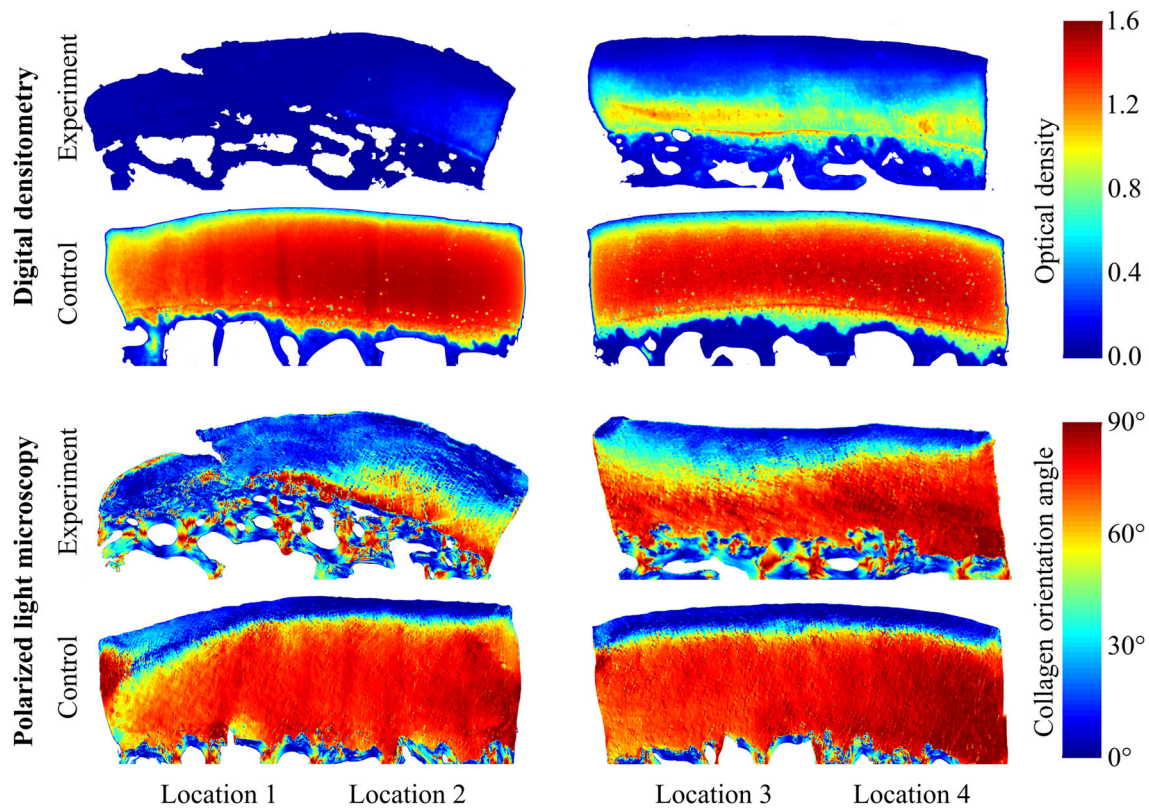


FIGURE 2. Example digital densitometry (optical density, i.e., proteoglycan content) and polarized light microscopy (collagen orientation angle) images of the experiment (worst case) and control (best case) groups from matching anatomical sites (distal locations 1-4). The arthroscopic photo shown as an inset in Fig. 1b was recorded during NIRS acquisition of the worst case at Distal location 3.

Data Preparation

To ensure well-generalized models, fivefold cross-validation was employed with 2 out of 10 ponies used as the validation set (each pony was used once in the validation). With each reference property, the ponies for the validation set were selected based on the pony-specific average value, i.e., in the first fold ponies with lowest and sixth lowest average values were assigned to the validation set (1-6, 2-7, etc.). This division was adopted to ensure representative calibration and validation sets (range of data), along with ensuring that the calibration set had no dependent information in the validation set. In addition, the reference properties and spectra were scaled with the MinMaxScaler and StandardScaler functions in the sklearn package,²¹ respectively.

Network Design and Training

As NNs require substantial amount of data to ensure well-generalizable models, data augmentation was adapted in this study by including the three repeated *in vitro* spectra from each location in the modelling. In addition, several studies have highlighted

the advantage of ensemble networks over a single network; thus, two separate network architectures were developed. To further optimize the modelling, the networks were initialized with three random seeds (7, 14, and 21). Therefore, the performance of the ensemble network was evaluated by averaging the outputs of these six variants. The network architectures were optimized by monitoring the overall performance of the validation sets.

The first network architecture was adapted from Bjerrum *et al.* by employing 1D-CNNs.⁵ The optimal network had two 1D convolutional layers (layer1_{kernel} = 8, layer1_{filter} = 64, layer2_{kernel} = 64, layer2_{filter} = 16) with ReLu activations.¹⁸ The outputs were flattened, followed by a dropout layer (ratio = 0.75), a fully connected dense layer (neurons = 128) with a ReLu activation, and the output layer with a linear activation. The second network architecture included a dense layer (neurons = 120) with a sigmoid activation, a dense layer (neurons = 60) with a ReLu activation, a dropout layer (ratio = 0.20), a dense layer (neurons = 15) with a ReLu activation, and the output layer with a linear activation. Root mean square errors (RMSE) of vali-

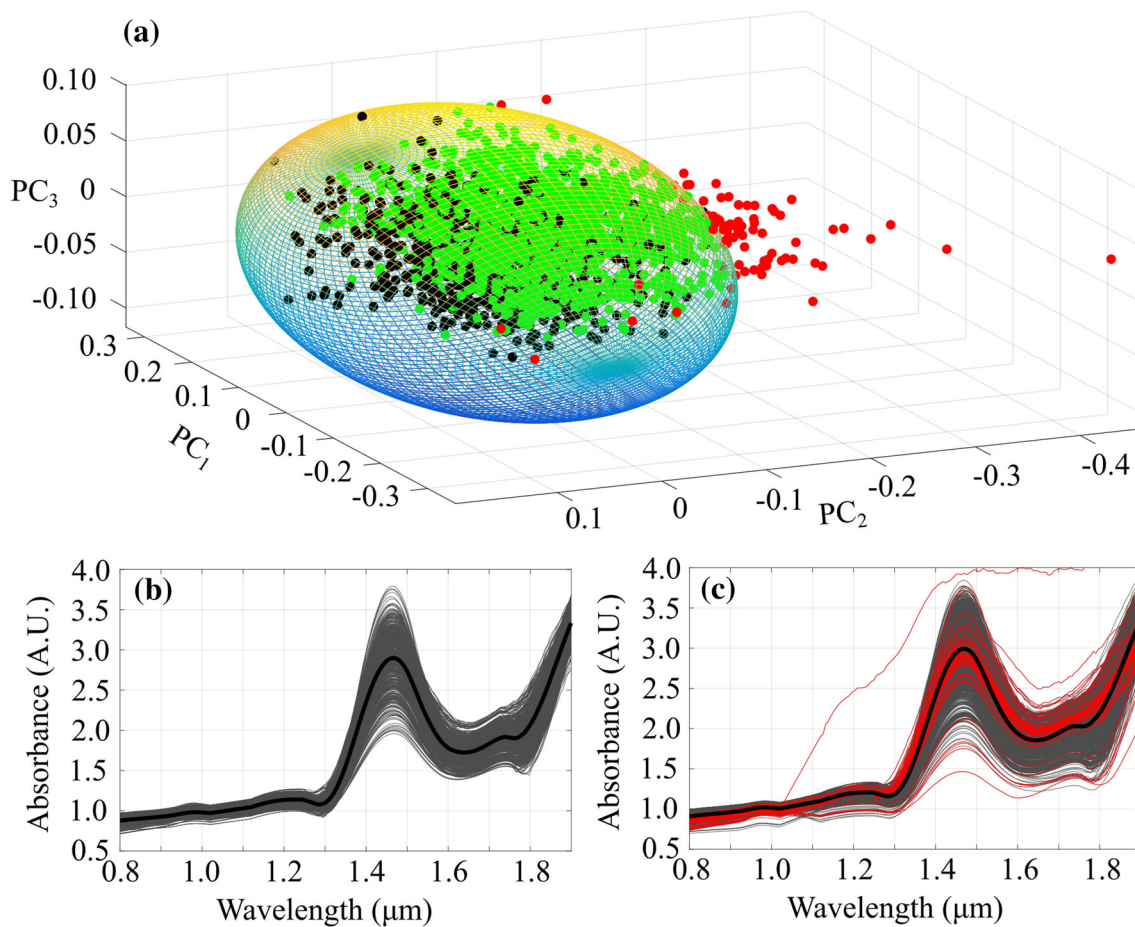


FIGURE 3. Minimum volume ellipsoid (MVE, a) calculated from the scores of *in vitro* (black) NIR spectral measurements. In addition, scores of *in vivo* spectral measurements within the MVE are presented (green) along with the outliers (red). In addition, smoothed *in vitro* (b) and *in vivo* (c) spectra with average spectra (black) are presented. In subfigure (c), outlier spectra are shown in red.

dation sets were used to evaluate the performance of both network architectures during the training. The training was performed using the Adam optimizer in Keras.¹¹ To limit the chance of overfitting, Keras EarlyStopping and ReduceLRonPlateau callbacks were utilized to halt training if no improvement was achieved after 25 epochs and to reduce the learning rate after 20 epochs, respectively.

Statistics

The performance of the ensemble network calibration (*in vitro*), validation (*in vitro*), and test (*in vivo*) sets was calculated by averaging the location-specific predictions, i.e., with *in vitro* data, the predictions of three repetitions were averaged and, with the *in vivo* data, the predictions from non-outlier spectra were averaged. Due to a non-normal distribution of reference parameters (Shapiro–Wilk test, $p < 0.0001$), non-parametric tests were used. SPSS (Version 25, SPSS Inc., IBM Company, Armonk, NY, USA) was used in the statis-

tical analysis. Statistical significance of differences in tissue properties between experimental and control groups was investigated using the Mann–Whitney U test with $p < 0.05$ as the limit for statistical significance. For testing the statistical differences between the measured and predicted reference values, a Wilcoxon signed rank test was used with $p < 0.05$ as the limit for statistical significance. Linear Mixed Model analysis was performed to determine underlying causes for the variability of reference properties by using the anatomical location (distal, proximal), repair group (1–4) and control, and measurement location (location 1–4) as fixed effects, whereas pony and leg (pony \times leg) (left or right) were utilized as random effects.

RESULTS

In the arthroscopic evaluation, the cartilage surrounding the repair sites was visually intact with no signs of cartilage fibrillation or lesions. However,

during cartilage palpation, all measurement locations around the repair were soft in comparison to native cartilage. Furthermore, cartilage stiffness was observed to increase when moving away from the repair sites. Thus, the measurement locations were systematically denoted with the score of ICRS 1a.

During initial preliminary modelling, 1st derivative pre-processing exhibited systematically better performance over the other pre-processing methods and was, thus, selected as the optimal pre-processing method in the final analyses. In the outlier detection, the best performance (smallest average RMSE of test set) was observed with slightly larger MVE (radii of 105%, Figs. 4a and 4c) with 51 outliers (out of 1620 collected during the arthroscopy). The average error of test sets substantially increased (correlation decreased) after reduction of radii by 25% (100% \rightarrow 75%), indicating extensive rejection of good spectra. A less substantial difference was observed when the radii were increased, thus indicating a low number of outlier spectra during the arthroscopic measurements.

The performance of the trained ensemble networks with both validation (*in vitro*) and test (*in vivo*) sets for

the cartilage reference properties, including cartilage thickness, full-thickness and superficial PG content and collagen orientation angle, is presented in Table 1. In addition, predictions based on validation and test sets enabled a comparison between the experimental and control groups (Fig. 5).

When comparing the measured reference values of experimental and control groups (Fig. 6), significant ($p < 0.05$) differences in full-thickness PG content were observed in locations closest to lesions on proximal and distal sites. However, no difference was apparent with full-thickness PG content or collagen orientation angle at locations further away from lesions, thus possibly indicating that locations closest to the lesion are more vulnerable to excessive stresses and strains and, hence, progressive degeneration as also demonstrated by Venäläinen *et al.*³⁸ However, contributions from other factors, including inflammatory signalling and cell-mediated factors, could not be accounted for in the analysis. Based on the linear Mixed Model, the measurement location had by far the most significant effect ($p < 1E-13$ with all reference properties), whereas anatomical location (thickness:

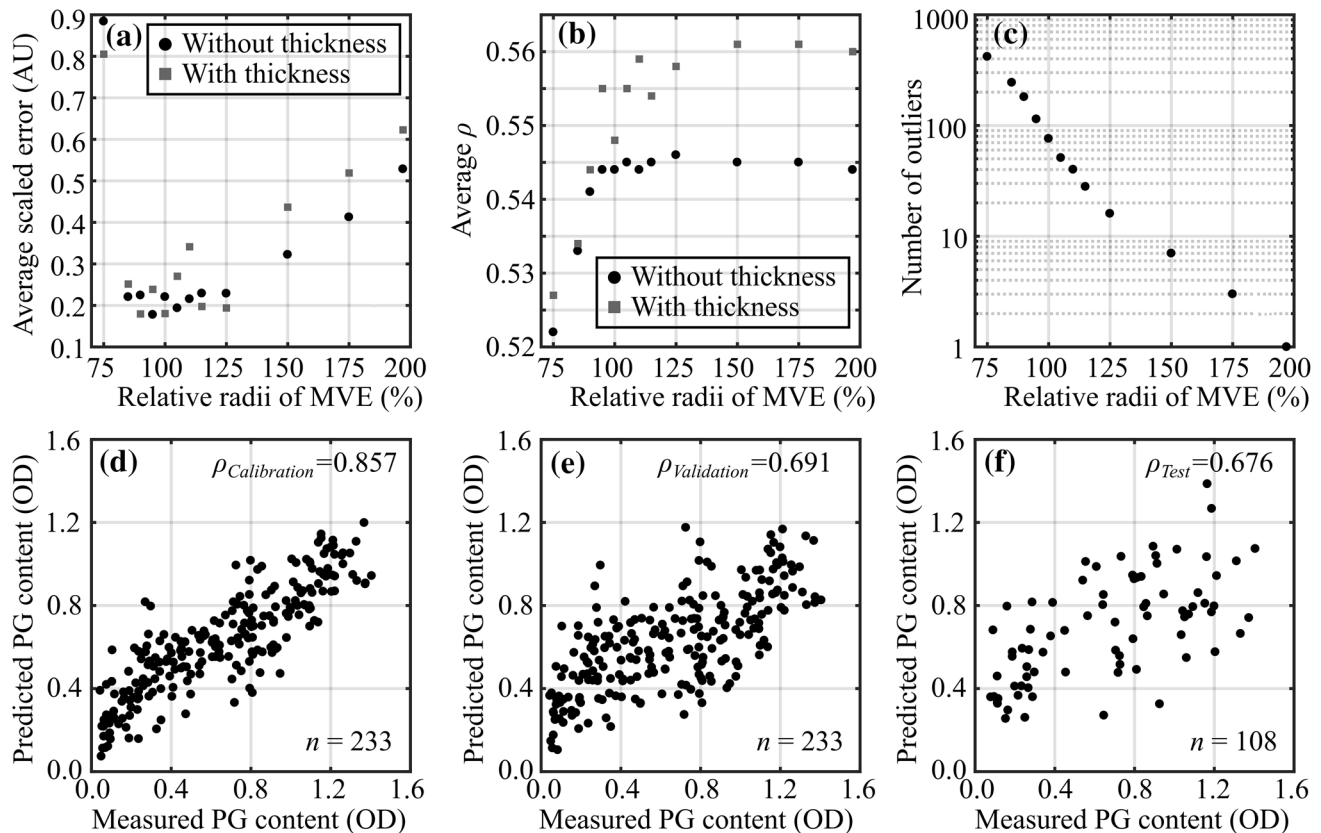


FIGURE 4. Detection of outlier *in vivo* spectra was performed by calculating the minimum volume ellipsoid (MVE) of the first three principal scores of *in vitro* spectra. For cartilage reference properties, the average scaled root mean square error (RMSE, a), average Spearman's rank correlation (ρ , b), and the number of outliers (c) are presented as a function of relative MVE radii with and without cartilage thickness. In addition, the relationships between measured and predicted full-thickness proteoglycan (PG) content (optical density, OD) for calibration (*in vitro*, d), validation (*in vitro*, e), and test (*in vivo*, f) sets are presented.

TABLE 1. Average cartilage reference properties and statistics on network performance, including Spearman's rank correlation (ρ , * $p < 0.001$), root mean square error (RMSE), and normalized RMSE (NRMSE)

Parameter	Cartilage thickness (mm)	Full-thickness PG content (OD)	Superficial PG content (OD)	Full-thickness collagen orientation angle (°)	Superficial collagen orientation angle (°)
Mean (range)	0.82 (0.14–1.36)	0.66 (0.04–1.41)	0.37 (0.04–1.17)	46.9 (10.1–77.2)	17.1 (3.6–56.0)
Calibration (<i>in vitro</i>)					
ρ	0.888*	0.857*	0.800*	0.781*	0.582*
RMSE	0.108	0.208	0.174	10.967	7.994
NRMSE	8.8%	15.3%	15.4%	16.4%	15.2%
Validation (<i>in vitro</i>)					
ρ	0.797*	0.691*	0.650*	0.626*	0.327*
RMSE	0.141	0.274	0.220	13.111	9.495
NRMSE	11.5%	20.1%	19.5%	19.6%	18.1%
Test (<i>in vivo</i>)					
ρ	0.596*	0.676*	0.613*	0.574*	0.316*
RMSE	0.191	0.296	0.248	13.439	11.642
NRMSE	17.2%	21.8%	22.0%	20.0%	22.2%

$p = 0.316$, PG: $p = 5.2E-6$, and collagen orientation angle: $p = 0.442$) and repair group (thickness: $p = 0.026$, PG: $p = 0.040$, and collagen orientation angle: $p = 0.646$) had significant effects in few occasions. In pairwise comparison between healthy (control group) and different repair groups (experimental group), a significant difference ($p = 0.034$) was observed only with cartilage PG content with one repair group (*GelMa reinforced*).

DISCUSSION

In this study, an arthroscopic NIRS was introduced, for the first time, for *in vivo* evaluation of cartilage thickness, PG content, and collagen orientation angle. Furthermore, no previous study has utilized 1D-CNNs to investigate the relationship between cartilage NIR spectra and reference properties; although, these relations have been previously demonstrated in several *in vitro* studies.^{3,25,26} The spectral region utilized in this study was similar or wider compared to the previous studies. The findings presented here demonstrate the clinical potential of NIRS for the evaluation of cartilage thickness, composition, and structure. Particularly, the prediction of superficial cartilage PG content could prove useful in detecting compromised cartilage prior to visible signs of degradation. The presented technique has the potential to improve the arthroscopic assessment of cartilage integrity. This could lead to better staging of OA and aid decision-making in cartilage repair surgery, for example, when selecting suitable autograft for mosaicplasty surgery or when assessing the boundaries of focal cartilage defects and healthy cartilage.

Significant differences in collagen orientation angle and PG content between the experimental and control

groups were only observed at the locations closest to lesions (Fig. 6). Although there were no other significant differences between the two groups, the observable difference substantially decreased when moving away from the lesions, with no apparent difference between both groups at the farthest location (location 4). The application of NIRS could, therefore, enable demarcation of the boundary between healthy and compromised cartilage tissue when performing repair arthroscopies. However, prior to clinical application, a substantial cartilage reference library would need to be collected to establish the normal variance of different anatomical locations. Linear mixed model analysis revealed that the type of repair strategies had no substantial effect on cartilage properties relative to the effect of measurement location (i.e., distance from the repair). Cartilage tissue closest to the repair site has been exposed to higher than normal stresses and strains, potentially accelerating the tissue degeneration. This observation is supported by the findings of Venäläinen *et al.*³⁸ Our findings are in line with those of Kaul *et al.*, who reported differences between repair tissue and adjacent articular cartilage after failed marrow-stimulation techniques (microfracture and Pridie drilling).¹⁵ In that study, the surrounding cartilage showed signs of early to severe OA after 8 months.

Spahn *et al.* and Hofmann *et al.* have previously applied NIRS in human arthroscopies^{14,32} and compared the reliability of the technique with scores from other modalities. Specifically, Spahn *et al.* reported spectral changes with varying severity of cartilage lesions³² and Hofmann *et al.* showed that NIRS values (based on a ratio of two spectral peaks^{14,31,32}) correlate with KOOS (Knee Injury and Osteoarthritis Outcome Score²⁴). The NIRS value varied between 0 (degenerated cartilage) and 100 (healthy cartilage) and was

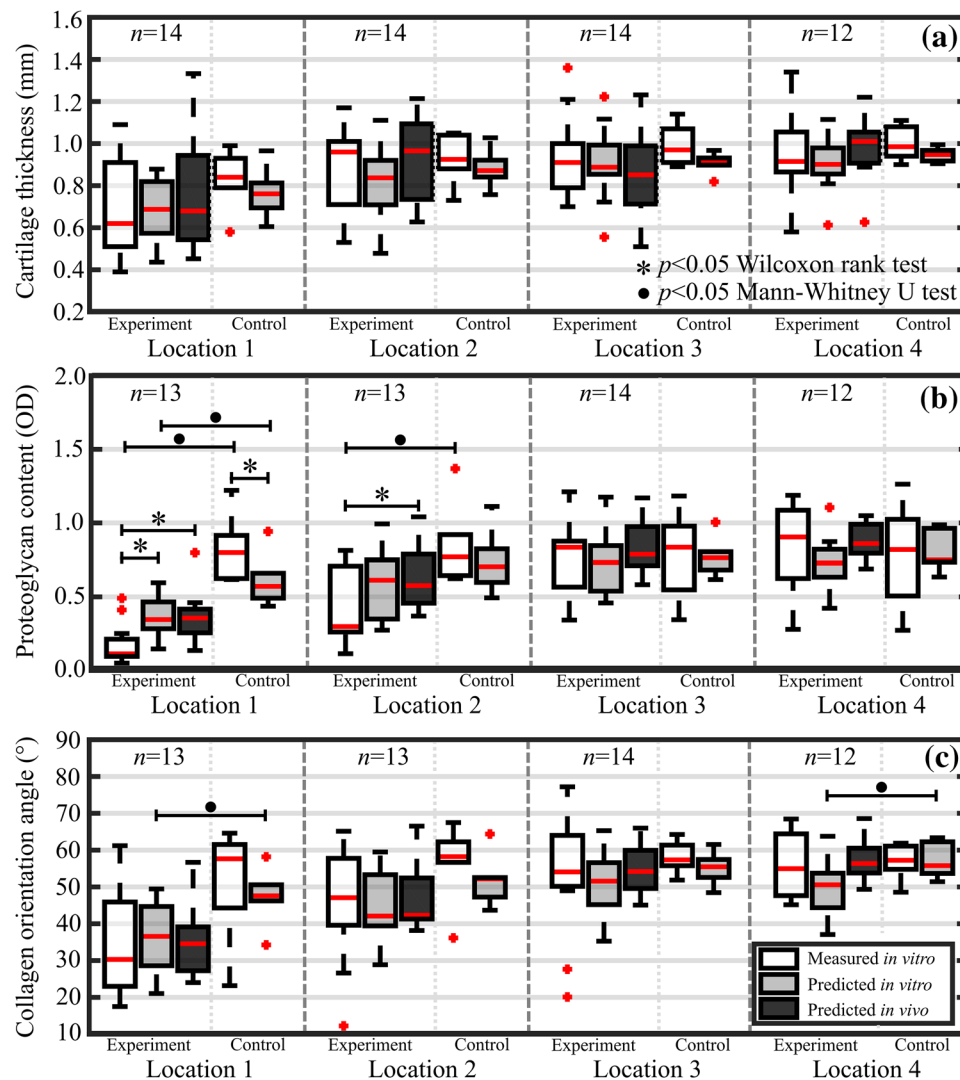


FIGURE 5. Boxplots for experimental and control groups with median (red line), quartiles (25% and 75%), and outliers (red crosses) of *in vitro* measured (white bars), *in vitro* predicted (grey bars), and *in vivo* predicted (black bars) for the 4 distal locations at increasing distances from the lesion. Values are presented for cartilage thickness (a), full-thickness proteoglycan content (b), and full-thickness collagen orientation angle (c). For each location, experiment and control groups had 12–14 and 6 measurements, respectively.

suggested to indicate diminished stiffness^{17,33} and altered water content.^{31,33} In addition to the suboptimal analysis technique, the studies utilized a narrower spectral region (1.1–1.7 μm) compared to the present study (0.8–1.9 μm). The inclusion of shorter and longer wavelengths within the wider spectral region enables evaluation of subchondral bone²⁷ and more specific evaluation of superficial layer, respectively, due to the wavelength-dependent penetration of light into the tissue.²⁰

Previous studies have reported the relationship between native cartilage PG content and its NIR spectra.^{3,26} Afara *et al.* demonstrated the potential of PG mapping with a similar spectral region (0.8–2.5 μm) using bovine samples.³ Although the reported

correlations were relatively higher, the study utilized PLSR with leave-one-out cross-validation, which is not as robust as *k*-fold cross-validation, since the calibration set always included measurements from the same joint as the validation (left out sample). In our previous *in vitro* study,²⁶ combination of a narrower spectral region (0.7–1.1 μm) and shallow NNs was demonstrated with samples from equine cartilage, resulting in similar correlations as presented here. With superficial cartilage PG content, the ensemble NN achieved a lower prediction error (RMSE) compared to the shallow NN in our previous study. The previous study also utilized OCT-based image classification to improve the prediction accuracy of the shallow NNs.²⁶ However, in the present study, all cartilage samples

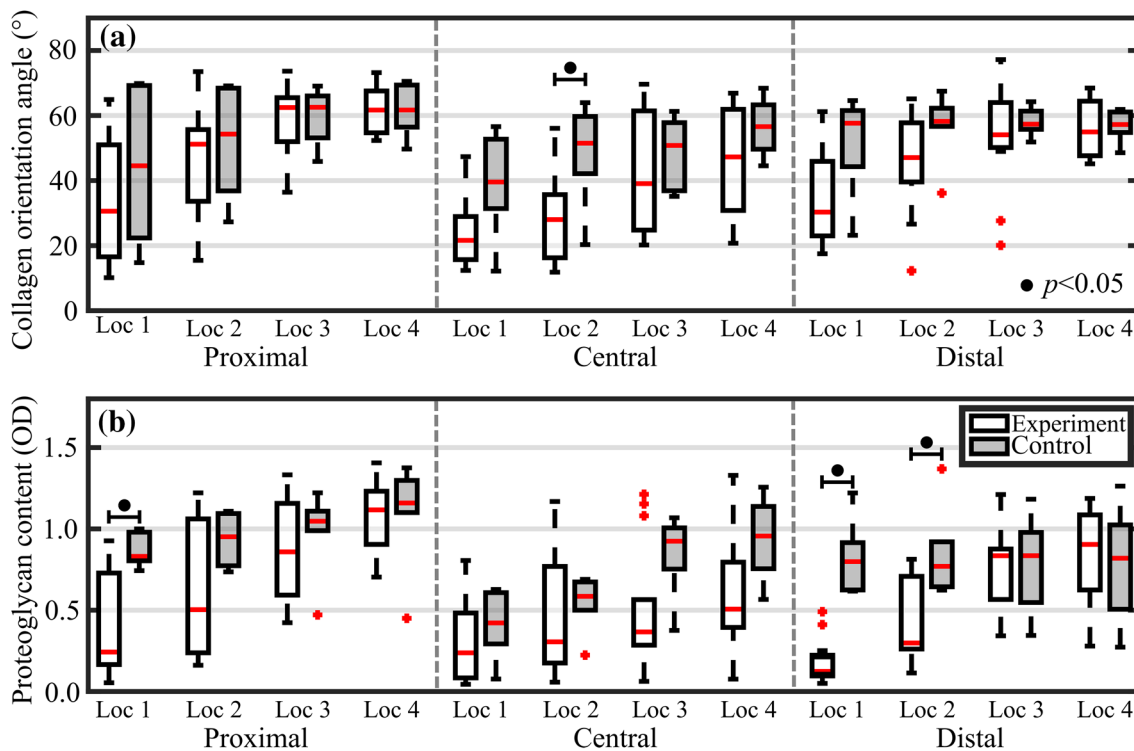


FIGURE 6. Boxplots for experimental and control groups with median (red line), quartiles (25 and 75%), and outliers (red crosses) of measured full-thickness collagen orientation angle (a) and proteoglycan content (b) for the proximal, central, and distal sites.

were assessed as visually intact based on conventional arthroscopic evaluation and the OCT images, making image-based classification impractical.

The relevant spectral features harnessed by the NNs for predicting the target (reference) tissue properties are due to the depth-dependent interactions (absorption, scattering, *etc.*) of NIR light with the tissue. These interactions are related to chemical bonds of cartilage constituents, *i.e.*, water, PG, and collagen. The NIR region is mostly affected by the first, second, and third overtone vibrations arising from stretching, bending, and combination bands. The most common bonds of cartilage constituents are OH, CH, NH, and CO. As cartilage matrix mainly consists of water (up to 80%),²⁹ the absorption of OH stretching and bending vibrations is most pronounced in the cartilage spectrum. Several optical windows (*e.g.*, 1.10–1.23 and 1.40–1.55 μm), not affected by the masking effect of water, have shown good prediction performance. The relationship between collagen orientation angle and NIR spectra is expected to arise from the birefringence property of cartilage, with healthy and intact collagen network reflecting more light back to the NIRS probe. With cartilage thickness, the NIR spectra are affected by the path length of light and the strong interaction at the interface of cartilage and bone (due to difference in refractive indices).^{4,40} The superior penetration depth of NIRS, compared to MIRS, promotes its application

for evaluation of whole cartilage tissue in clinical arthroscopies. In contrast to MIRS, NIRS is also capable of providing information on the underlying bone through articular cartilage in both equine and humans.^{1,27} These studies showed that the visible and short NIR spectral regions are optimal for estimating bone properties, such as subchondral plate thickness and volume fraction, due to their better penetration depth compared to longer wavelengths.²⁰

Although analyses of cartilage NIR spectra have mostly relied on conventional multivariate approaches (*e.g.*, PCA and PLS), recent studies have explored the potential of shallow NNs^{26,27} due to their ability to model complex non-linear features.^{12,19} In this study, deep NNs with two different architectures were developed and tested with positive outcomes. With deep learning, a substantial amount of data is required to ensure reliable optimization and to limit the chance of overfitting. To account for this constraint, cross-validation, dropout layers, and early stopping were utilized. In cross-validation, the division was based on pony level (*i.e.*, calibration and validation set did not include data from the same pony), therefore ensuring that the validation set had independent data.

Outlier detection is essential in diagnostics to ensure reliable estimation of tissue properties.^{27,30} In our earlier study utilizing the same NIR spectral data,²⁷ 8 out of 15 spectra were eliminated based on contri-

butions from the light source of the conventional arthroscope during measurements. However, this does not account for the relative orientation of the NIRS probe to the arthroscope as inclination of NIRS probe tip (non-optimal contact) towards arthroscopic light source will result to stronger contributions compared to an inclination to opposite direction; therefore, the MVE approach was adopted. In the study of Spahn *et al.*,³⁰ no explicit reasoning for outlier classification was given. The MVE approach, based on the controlled *in vitro* measurements, enables fast and effective detection of outliers as observed with the prediction results.

During the arthroscopic NIRS measurements, the narrow cavity of the equine joint restricted probe mobility and, therefore, limited the alignment between the probe and cartilage surface. However, the number of outlier measurements was relatively low (3.1%), demonstrating the clinical feasibility of the technique. The main limitation of this study was the contribution of the arthroscopic light source, as this limits the use of data from the visible spectral region, which may be useful for predicting subchondral and trabecular bone properties.²⁷ As an animal model, equines were well-suited for this study due to similar cartilage thickness with humans.¹⁶ In addition, for example, racehorses suffer from joint injuries that are often evaluated and treated arthroscopically, thus demonstrating the usefulness of the technique in a relevant veterinary field.

In conclusion, NIRS in combination with deep NNs can be utilized for predicting cartilage thickness, PG content, and collagen orientation angle during *in vivo* arthroscopy, enabling differentiation between the experimental and control groups. Therefore, adaptation of NIRS may prove useful clinically, e.g. when arthroscopically selecting suitable allograft tissue for mosaicplasty surgery or when assessing the boundaries of focal cartilage defects and healthy cartilage.

ACKNOWLEDGMENTS

Open access funding provided by University of Eastern Finland (UEF) and Kuopio University Hospital. 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 (Project 267551), the Orion Research Foundation sr, and the Finnish Foundation of Technology Promotion financially supported this study. The European Community's Seventh Framework Programme (FP7/2007–2013) under Grant Agreement 309962 (Hy-

droZONES), and the Dutch Arthritis Society (LLP-12 and LLP-22) supported this research. Nikae CR te Moller is acknowledged for assistance during arthroscopic measurements.

CONFLICT OF INTEREST

The authors have no conflicts of interest regarding the study or preparation of the manuscript.

OPEN ACCESS

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

REFERENCES

- ¹Afara, I. O., C. Florea, I. A. Olumegbon, C. T. Eneh, M. K. H. Malo, R. K. Korhonen, and J. Töyräs. Characterizing human subchondral bone properties using near-infrared (NIR) spectroscopy. *Sci. Rep.* 8:9733, 2018.
- ²Afara, I. O., M. Hauta-Kasari, J. S. Jurvelin, A. Oloyede, and J. Töyräs. Optical absorption spectra of human articular cartilage correlate with biomechanical properties, histological score and biochemical composition. *Physiol. Meas.* 36:1913–1928, 2015.
- ³Afara, I. O., H. Moody, S. Singh, I. Prasad, and A. Oloyede. Spatial mapping of proteoglycan content in articular cartilage using near-infrared (NIR) spectroscopy. *Biomed. Opt. Express* 6:144–154, 2015.
- ⁴Ascenzi, A., C. Fabry, S. The, B. Cytology, and N. Aug. Technique for dissection and measurement of refractive index of osteones. *J. Biophys. Biochem. Cytol.* 6:139–142, 1959.
- ⁵Bjerrum, E. J., M. Glahder, and T. Skov. Data augmentation of spectral data for convolutional neural network (CNN) based deep chemometrics, 2017. <http://arxiv.org/abs/1710.01927>.
- ⁶Brismar, B. H., T. Wredmark, T. Movin, J. Leandersson, and O. Svensson. Observer reliability in the arthroscopic classification of osteoarthritis of the knee. *J. Bone Joint Surg. Br.* 84:42–47, 2002.
- ⁷Brittberg, M., and C. S. Winalski. Evaluation of cartilage injuries and repair. *J. Bone Joint Surg. Am.* 85:58–69, 2003.
- ⁸Brown, T. D., R. C. Johnston, C. L. Saltzman, J. L. Marsh, and J. A. Buckwalter. Posttraumatic osteoarthritis: a first estimate of incidence, prevalence, and burden of disease. *J. Orthop. Trauma* 20:739–744, 2006.
- ⁹Buckwalter, J. A., and J. Martin. Degenerative joint disease. *Clin. Symp.* 47:1–32, 1995.
- ¹⁰Burns, D. A., and E. W. Ciurczak. Handbook of Near-Infrared Analysis (3rd ed.). Boca Raton: CRC Press, 2009.
- ¹¹Chollet, F. and others. Keras., 2015. keras.io.

- ¹²Goldshleger, N., A. Chudnovsky, and E. Ben-Dor. Using reflectance spectroscopy and artificial neural network to assess water infiltration rate into the soil profile. *Appl. Environ. Soil Sci.* 2012. <https://doi.org/10.1155/2012/439567>.
- ¹³Hame, S. L., and R. A. Alexander. Knee osteoarthritis in women. *Curr. Rev. Musculoskelet. Med.* 6:182–187, 2013.
- ¹⁴Hofmann, G. O., J. Marticke, R. Grossstück, M. Hoffmann, M. Lange, H. K. W. Plettenberg, R. Braunschweig, O. Schilling, I. Kaden, and G. Spahn. Detection and evaluation of initial cartilage pathology in man: a comparison between MRT, arthroscopy and near-infrared spectroscopy (NIR) in their relation to initial knee pain. *Pathophysiology* 17:1–8, 2010.
- ¹⁵Kaul, G., M. Cucchiari, K. Remberger, D. Kohn, and H. Madry. Failed cartilage repair for early osteoarthritis defects: A biochemical, histological and immunohistochemical analysis of the repair tissue after treatment with marrow-stimulation techniques. *Knee Surgery Sport. Traumatol. Arthrosc.* 20:2315–2324, 2012.
- ¹⁶Malda, J., J. C. de Grauw, K. E. M. Benders, M. J. L. Kik, C. H. A. van de Lest, L. B. Creemers, W. J. A. Dhert, and P. R. van Weeren. Of mice, men and elephants: the relation between articular cartilage thickness and body mass. *PLoS ONE* 8:e57683, 2013.
- ¹⁷Marticke, J. K., A. Hösselbarth, K. L. Hoffmeier, I. Marintschev, S. Otto, M. Lange, H. K. W. Plettenberg, G. Spahn, and G. O. Hofmann. How do visual, spectroscopic and biomechanical changes of cartilage correlate in osteoarthritic knee joints? *Clin. Biomech.* 25:332–340, 2010.
- ¹⁸Nair, V., and G. E. Hinton. Rectified Linear Units Improve Restricted Boltzmann Machines. 2010.
- ¹⁹Ni, Y., G. Zhang, and S. Kokot. Simultaneous spectrophotometric determination of maltol, ethyl maltol, vanillin and ethyl vanillin in foods by multivariate calibration and artificial neural networks. *Food Chem.* 89:465–473, 2005.
- ²⁰Padalkar, M. V., and N. Pleshko. Wavelength-dependent penetration depth of near infrared radiation into cartilage. *Analyst* 140:2093–3100, 2015.
- ²¹Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay. Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12:2825–2830, 2011.
- ²²Perez, L., and J. Wang. The effectiveness of data augmentation in image classification using deep learning. 2017. <https://doi.org/10.1109/igarrs.2016.7730324>.
- ²³Qu, C., M. Hirviniemi, V. Tiitu, J. S. Jurvelin, J. Töyräs, and M. J. Lammi. Effects of freeze-thaw cycle with and without proteolysis inhibitors and cryopreservant on the biochemical and biomechanical properties of articular cartilage. *Cartilage* 5:97–106, 2014.
- ²⁴Roos, E. M., and S. Toksvig-Larsen. Knee injury and osteoarthritis outcome score (KOOS)—validation and comparison to the WOMAC in total knee replacement. *Health Qual. Life Outcomes* 1:17, 2003.
- ²⁵Sarin, J. K., M. Amissah, H. Brommer, D. Argüelles, J. Töyräs, and I. O. Afara. Near infrared spectroscopic mapping of functional properties of equine articular cartilage. *Ann. Biomed. Eng.* 44:3335–3345, 2016.
- ²⁶Sarin, J. K., L. Rieppo, H. Brommer, I. O. Afara, S. Saarakkala, and J. Töyräs. Combination of optical coherence tomography and near infrared spectroscopy enhances determination of articular cartilage composition and structure. *Sci. Rep.* 7:10586, 2017.
- ²⁷Sarin, J. K., N. C. R. te Moller, I. A. D. Mancini, H. Brommer, J. Visser, J. Malda, P. R. van Weeren, I. O. Afara, and J. Töyräs. Arthroscopic near infrared spectroscopy enables simultaneous quantitative evaluation of articular cartilage and subchondral bone in vivo. *Sci. Rep.* 8:13409, 2018.
- ²⁸Savitzky, A., and M. J. E. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36:1627–1639, 1964.
- ²⁹Sophia Fox, A. J., A. Bedi, and S. A. Rodeo. The basic science of articular cartilage: structure, composition, and function. *Sports Health* 1(461–468):2009, 2009.
- ³⁰Spahn, G., H. M. Klinger, M. Baums, M. Hoffmann, H. Plettenberg, A. Kroker, and G. O. Hofmann. Near-infrared spectroscopy for arthroscopic evaluation of cartilage lesions: results of a blinded, prospective, interobserver study. *Am. J. Sports Med.* 38:2516–2521, 2010.
- ³¹Spahn, G., H. Plettenberg, M. Hoffmann, H.-T. Klemm, C. Brochhausen-Delius, and G. O. Hofmann. The frequency of cartilage lesions in non-injured knees with symptomatic meniscus tears: results from an arthroscopic and NIR-(near-infrared) spectroscopic investigation. *Arch. Orthop. Trauma Surg.* 137:837–844, 2017.
- ³²Spahn, G., H. Plettenberg, E. Kahl, H. M. Klinger, T. Mückley, and G. O. Hofmann. Near-infrared (NIR) spectroscopy. A new method for arthroscopic evaluation of low grade degenerated cartilage lesions. Results of a pilot study. *BMC Musculoskelet. Disord.* 8:47, 2007.
- ³³Spahn, G., H. Plettenberg, H. Nagel, E. Kahl, H. M. Klinger, T. Mückley, M. Günther, G. O. Hofmann, and J. A. Mollenhauer. Evaluation of cartilage defects with near-infrared spectroscopy (NIR): an ex vivo study. *Med. Eng. Phys.* 30:285–292, 2008.
- ³⁴Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15:1929–1958, 2014.
- ³⁵Stiebel, M., L. E. Miller, and J. E. Block. Post-traumatic knee osteoarthritis in the young patient: therapeutic dilemmas and emerging technologies. *Open Access J. Sport. Med.* 5:73–79, 2014.
- ³⁶Stone, M. Cross-validated choice and assessment of statistical predictions. *J. R. Stat. Soc. Ser. B* 36:111–147, 1974.
- ³⁷Tuan, R. S., A. F. Chen, and B. A. Klatt. Cartilage regeneration. *J. Am. Acad. Orthop. Surg.* 21:303–311, 2013.
- ³⁸Venäläinen, M. S., M. E. Mononen, J. Salo, L. P. Räsänen, J. S. Jurvelin, J. Töyräs, T. Virén, and R. K. Korhonen. Quantitative evaluation of the mechanical risks caused by focal cartilage defects in the knee. *Sci. Rep.* 6:37538, 2016.
- ³⁹von Engelhardt, L. V., M. Lahner, A. Klusmann, B. Bouillon, A. Dávid, P. Haage, and T. K. Lichtinger. Arthroscopy vs MRI for a detailed assessment of cartilage disease in osteoarthritis: diagnostic value of MRI in clinical practice. *BMC Musculoskelet. Disord.* 11:75, 2010.
- ⁴⁰Wang, K., J. Wu, R. E. Day, and T. B. Kirk. Utilizing confocal microscopy to measure refractive index of articular cartilage. *J. Microsc.* 248:281–291, 2012.
- ⁴¹Zhang, T., and B. Yu. Boosting with early stopping: convergence and consistency. *Ann. Stat.* 33:1538–1579, 2005.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.