Infrared spectroscopy is suitable for objective assessment of articular cartilage health

V. Virtanen a,*, V. Tafintseva b, R. Shaikh c, E. Nippolainen c, J. Haas d, e, I.O. Afara c, f, J. Töyräs c, g, H. Kröger b, J. Solheim b, B. Zimmermann b, A. Kohler b, B. Mizaikoff d, e, M. Finnilä a, L. Rieppo a, S. Saarakkalaa a, i

a Research Unit of Medical Imaging, Physics and Technology, Faculty of Medicine, University of Oulu, Oulu, Finland
b Faculty of Science and Technology, Norwegian University of Life Sciences, Ås, Norway
c Department of Applied Physics, University of Eastern Finland, Kuopio Campus, Kuopio, Finland
d Institute of Analytical and Bioanalytical Chemistry, Ulm University, Ulm, Germany
e Hahn-Schickard, Ulm, Germany
f School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, Australia
g Kuopio University Hospital, Kuopio, Finland
h Department of Orthopedics, Traumatology and Hand Surgery, Kuopio University Hospital, Kuopio, Finland
i Department of Diagnostic Radiology, Oulu University Hospital, Oulu, Finland

ARTICLE INFO

Keywords:
Fourier transform infrared spectroscopy
Cartilage
Osteoarthritis
Arthroscopy

ABSTRACT

Objective: To evaluate the feasibility of Fourier transform infrared attenuated total reflectance (FTIR-ATR) spectroscopy to detect cartilage degradation due to osteoarthritis and to validate the methodology with osteochondral human cartilage samples for future development towards clinical use.

Design: Cylindrical (d = 4 mm) osteochondral samples (n = 349) were prepared from nine human cadavers and measured with FTIR-ATR spectroscopy. Afterwards, the samples were assessed with Osteoarthritis Research Society International (OARSI) osteoarthritis cartilage histopathology assessment system and divided into two groups: 1) healthy (OARSI 0–2) and 2) osteoarthritic (OARSI 2.5–6). The classification was done with partial least squares discriminant analysis model utilizing cross-model validation. Receiver operating characteristics curve analysis was performed and the area under curve (AUC) was calculated.

Results: For all samples combined, classification accuracy was 73% with AUC of 0.79. Femoral samples had accuracy of 74% and AUC of 0.77, while tibial samples had accuracy of 66%, and AUC of 0.74. Patellar samples had accuracy of 84% and AUC of 0.91.

Conclusions: The results indicate that FTIR-ATR spectroscopy can differentiate between healthy and osteoarthritic femoral, tibial and patellar human tissue. If combined with a fiber optic probe, FTIR-ATR spectroscopy could provide additional objective intraoperative information during arthroscopic surgeries, which could improve clinical outcomes.

1. Introduction

Articular cartilage, the highly specialized connective tissue covering the ends of bones in diarthrodial joints, enables the near frictionless movement of joints [1]. Articular cartilage is avascular and aneural and, hence, has limited self-repair capacity [1]. Cartilage cells, i.e., chondrocytes, occupy 1–3% of total cartilage volume [2]. Chondrocytes are surrounded by water and extracellular matrix (ECM), which is a complex network of collagen fibers and proteoglycans [2].

Injuries to articular cartilage present an exceptionally challenging medical problem due to lack of the tissue self-regenerative capability [3]. Moreover, untreated cartilage injuries may develop into osteoarthritis (OA) with symptoms like swelling, joint pain and impaired mobility. OA is the most common chronic joint disease. Globally, the socio-economic burden of OA is rising due to the ageing population and obesity [4]. The annual cost of OA range between 1330 and 10,452 euros per capita in Europe [5]. Currently, there is no cure for OA, although numerous repair procedures exist for treating knee joint injuries. The goal of
cartilage repair surgery is to restore the cartilage tissue, and it is typically carried out during arthroscopy [3]. The effectiveness of the repair procedures relies on an accurate diagnosis of the severity and extent of the articular cartilage damage [6]. At present, orthopaedic surgeons rely on visual assessment and manual probing to estimate cartilage health during the arthroscopic joint repair surgeries [7]. Thus, the current arthroscopic assessment of cartilage is highly subjective and has poor reproducibility [8,9].

Fourier transform infrared (FTIR) spectroscopy, a popular type of vibrational spectroscopy, is based on molecular vibrations of investigated material, and it can provide specific information on tissue biochemical composition without any sample processing [10,11] (Fig. 1). There are studies showing that an FTIR spectrometer can be coupled with an attenuated total reflectance (ATR) based fiber optic probe to enable in vivo measurements [12,13]. Fiber optic FTIR spectra of degenerated tibial human articular cartilage have been shown to correlate with Mankin grade (histological assessment of cartilage) [14] and Collins grade (macroscopic assessment of cartilage) [15]. Moreover, multivariate data analysis of FTIR spectra can differentiate between the type of collagen, i.e. type I or II, in connective tissues [16]. Consequently, these results suggest that FTIR spectroscopy could provide objective information about cartilage tissue health during arthroscopy to help making better-informed decisions and, thus, potentially improve surgery outcomes.

The aim of the present study was to evaluate the feasibility of FTIR-ATR spectroscopy for assessing articular cartilage health. Earlier FTIR-ATR studies on human tissue have been limited to the study of tibial cartilage [14]. In contrast, more versatile sample set used in this study includes tibial, femoral, and patellar cartilage. A Multivariate discriminant model was built to classify the cartilage FTIR spectra into relatively intact and degraded classes according to Osteoarthritis Research Society International (OARSI) histopathological grading [18].

2. Methods

2.1. Study design and experimental samples

The sample set consisted of a total of 349 samples which were extracted from both knees of 9 human cadavers. The complete study design is described in the flowchart presented in Fig. 2. Ethical permission was obtained from Ethics Committee of Northern Savo prior to performing the study, permission number 123/2015 (58/2013). The samples were prepared by drilling cylindrical $(d = 4 \text{ mm})$ osteochondral plugs with a dental drill from central locations of femoral, tibial, and patellar cartilage. After extraction, the sample plugs were immersed in phosphate buffered saline (PBS) and deep frozen at $-80^\circ \text{ C}$ for storage.

Fig. 1. A schematic representation of Fourier transform infrared (IR) attenuated total reflection (ATR) measurement setup.

Prior to FTIR spectroscopic measurements, deep frozen samples were left to thaw at room temperature for 30 min. Defrosted samples were subsequently attached with instant glue (Cyanoacrylate, Loctite 401, Henkel Corporation) from bone to plastic petri dishes to ensure easier clamping of the tissue to enable good contact with the ATR sampling cell (Fig. 1).

Spectra were recorded with a Bruker Alpha HR spectrometer (Bruker Optics GmbH, Ettlingen, Germany) equipped with a globar mid-IR source and a deuterated triglycine sulfate (DTGS) detector. The Bruker OPUS 8.1 (Bruker Optics GmbH, Ettlingen, Germany) spectroscopy software was used for data acquisition and instrument control. The spectrometer was equipped with a Bruker Platinum ATR sampling cell (Bruker Optics GmbH, Ettlingen, Germany). The cell is composed of a single reflection diamond ATR crystal embedded in tungsten carbide which provides excellent inertness towards biomedical samples. Further, the Platinum ATR cell contained an adjustable pressure clamp that facilitated adjustment of an optimum contact pressure for reproducible acquisition of FTIR-ATR spectra. During measurement, the osteochondral plugs were immersed in PBS to avoid dehydration and to maintain a physiological environment. For each spectrum, 128 scans were averaged. The ATR crystal was cleaned with isopropanol and background was measured in air between each sample measurement. Spectra were recorded within a spectral window of 4000 cm$^{-1}$ to 600 cm$^{-1}$ at a spectral resolution of 2 cm$^{-1}$. Each sample was measured three times except for one sample for which one replicate is missing.

2.2. Fourier transform infrared spectroscopy

The samples were fixed in formalin and decalcified with 5% EDTA solution after FTIR spectroscopic measurements. Following dehydration, the samples were embedded in paraffin for histology. Thereafter, 3 μm
thick sections were cut and stained with Safranin O. These histological sections were evaluated according to the OARSI grading system [18], which grades the lesion depth to assess the severity of OA with grades ranging from 0 (healthy) to 6 (severe OA with bone exposed). In this study, randomly ordered and blind-coded sections were scored by three graders first independently, and afterwards a consensus grade was decided by uniform agreement. The grades were then pooled to form a sample set with largely intact tissue (OARSI grade: 0–2) and another sample set with clear surface damage, such as fissures and loss of cartilage matrix (OARSI grade: 2.5–6).

2.4. Data analysis

Spectra from samples with insufficient cartilage contact to the ATR crystal were excluded by removing all spectra with absorbance less than 0.05 at 1240 cm$^{-1}$. This removed 22 spectra in total. The remaining 1,024 spectra were included in further analysis. Pre-processing by weighted multiplicative signal correction (MSC) [19,20] was applied. Regions of 1800–1780 cm$^{-1}$ and 900–800 cm$^{-1}$ were up-weighted by 15 to mitigate baseline effects. Afterwards, the spectral range was truncated to 1720-900 cm$^{-1}$.

A partial least squares discriminant analysis (PLS-DA) classifier was used for discriminating the spectra into healthy (OARSI 0–2) and osteoarthritic (OARSI 2.5–6) groups. In PLS-DA, latent variables (i.e. PLS components) are calculated from independent (spectra) and dependent (sample groups, dummy variable: 0 (healthy) and 1 (osteoarthritic)) variables by maximising the covariance between the two variable sets. This reduces the dimensionality of the data and, thus, the effect of the multi-collinearity problem [21]. Models with up to 20 PLS components were calculated, and the model with the smallest number of PLS components, where the error is within 5% of the true minimum across all 20 PLS components, was selected as the final model. The cut-off for classification between healthy and osteoarthritic groups was 0.5.

Cross-model validation (CMV) [22] was used to reduce the possibility of overfitting and to check the stability of the results. For each CMV loop, samples from one cadaver at a time were taken out completely from the training set and used as the test set. The remaining training set was subjected to a leave-one-cadaver-out cross validation to optimize the model parameters (number of PLS components). The resulting PLS-DA model was then used for classifying the samples of the test set. This was repeated 9 times to classify the samples of all cadavers. The number of PLS components for each CMV step is shown in Table 1.

| CMV step number | Number of PLS components |
|-----------------|--------------------------|
| 1               | 7                        |
| 2               | 11                       |
| 3               | 9                        |
| 4               | 8                        |
| 5               | 10                       |
| 6               | 8                        |
| 7               | 10                       |
| 8               | 6                        |
| 9               | 9                        |

2.5. Statistical analysis

To evaluate PLS-DA model's performance, receiver operating characteristics curve analysis was obtained and area under curve (AUC) was calculated utilizing MATLAB built-in function perfcurve.

Intraclass correlation coefficient (ICC) was calculated to determine the reliability of OARSI grading between the three assessors. ICC was calculated from two-way random effects analysis of variance, with absolute agreement [23]. Inter-observer variability between the three observers showed ICC of 0.82.

3. Results

Mean FTIR-ATR spectra of the fingerprint region (1720-900 cm$^{-1}$) of the two sample groups are presented in Fig. 3 and the spectral difference between groups is shown in Fig. 4. Healthier group shows higher absorbance in most of the prominent bands, such as the amide I (1700-1600 cm$^{-1}$), amide II (1580-1490 cm$^{-1}$) and amide III (1300-1200 cm$^{-1}$). Collagen related peaks at 1458 cm$^{-1}$, 1378 cm$^{-1}$ and 1338 cm$^{-1}$ also show slightly higher absorbance in the healthier group. In the carbohydrate region, C-O stretch at 1165 cm$^{-1}$ shows higher intensity in healthier groups, with smaller variation across the region from 1100 cm$^{-1}$ to 1000 cm$^{-1}$.

PLS-DA was used for classifying the FTIR-ATR cartilage spectra into healthy and diseased groups. CMV accuracy was 72.8% and AUC was 0.79 for all samples combined. We also investigated the results by selecting the samples from femur, tibia, and patella separately. Classification of femoral samples had 73.6% accuracy and 0.77 AUC, tibial samples 66.0% accuracy and 0.74 AUC, while classification of patellar samples had 83.7% accuracy and 0.91 AUC. Confusion matrices for all samples combined and for each of the anatomical locations separately are shown in Table 2, while CMV accuracies and AUCs are shown in Table 3. For all samples combined, accuracy, sensitivity, and specificity for each step of the cross-model validation is shown in Fig. 5. The accuracy and sensitivity was notably worse for step #1 compared to rest of the CMV steps.

4. Discussion

The aim of this study was to evaluate the feasibility of FTIR-ATR spectroscopy for assessment of human cartilage health. In contrast to earlier studies, we included cartilage from joint surfaces of multiple bones, i.e., tibia, femur, and patella. A good overall classification accuracy of 72.8% was obtained with a PLS-DA model. This is a promising result considering the limited amount of data for modelling in the current study and conservative method applied to validate the developed PLS-DA model. The present results indicate that objective spectroscopic approach could provide an improvement over the currently used subjective methods, which have a limited accuracy [9].

Previous FTIR-ATR spectroscopic studies have used Mankin scoring and Collins grade as the reference classification methods. Mankin score is a histological score that considers progressive features of OA, i.e., lesion depth, but also features related to disease activity, such as cellular features and proteoglycan depletion. In contrast, Collins grade is a macroscopic grade that does not capture early tissue-level changes. In our study, we decided to use the OARSI histological grading as it is designed to reflect only the progression of OA [18]. In classification analysis, we...
and amide III (1300–1200 cm\(^{-1}\)) are both associated with the changes in the proteoglycan contents [26]. While these changes can be visually seen in the average and difference spectra, they might not be statistically significant by themselves, thus requiring the utilization of the multivariate methods.

PLS-DA classifiers are routinely used in spectroscopic analysis of biological tissues [16,27–30]. The validation for classification methods used in previous publications are not always clear as it is not necessarily mentioned whether the technical replicates of the same specimen or specimen extracted from the same individual are used in both the training and the test sets. We chose to validate the PLS-DA model using CMV to reduce the chance of overfitting. In our analysis, we separated the cadavers so that all samples extracted from one cadaver are used as the test samples in a single CMV step. In other words, samples of the same cadaver are not simultaneously present in training and test sets. This provides more conservative results, but also gives us confidence that the developed model generalizes better for unseen data. The present results indicate that FTIR-ATR spectroscopy has potential to be utilized as an objective tool during arthroscopy and give more confidence for surgeons during intraoperative decision making. Fiber optic spectroscopic methods have shown promise towards clinical use as summarised in recent review paper [11], and our study is one more step towards applications for cartilage health assessment.

The present study is limited by the relatively low number of samples. Furthermore, the uneven distribution of OARSI grades made more fine-grained classification analysis with multiple different OARSI grades unfeasible. In particular, the number of cadavers in the dataset is critical for generalization of the model. Fig. 5 shows that when testing the performance of the models on samples from certain cadavers, the performance of the model can vary. For example, when testing the model with cadaver #1 samples as the test set, the accuracy and sensitivity are notably worse compared to other CMV folds where other cadavers were utilized as independent test sets. Lower sensitivity in this case indicates that the model predicted relatively more of the degenerated samples as healthy. It is possible there is some biological reason behind this, and thus, increasing the number of cadavers would capture the biological variation between the individuals better and, therefore, result in more robust models. However, the overall performance of the PLS-DA model is good, as the accuracy and AUC were 72.8% and 0.79, respectively. Compared to earlier studies, we also investigated the location-based differences in this study. While the patellar samples showed high accuracy and AUC, the results are biased since 102 out of 123 samples had OARSI grade higher than 2. Femoral samples show slightly poorer performance at only 66.6% accuracy and 0.74 AUC. The healthier samples from tibia seemed to be more likely to be classified as degenerated: 43% or 52 out of 120 OARSI 0–2 grade.

Table 2

| Anatomical location | Diseased (predicted) | Healthy (predicted) |
|---------------------|----------------------|---------------------|
| Femur               | 74% (176)            | 26% (61)            |
| Tibia               | 73% (124)            | 27% (47)            |
| Patella             | 88% (90)             | 12% (12)            |
| All samples combined| 38% (8)              | 62% (13)            |

Table 3

Cross-model validation (CMV) accuracy and area under curve (AUC) for receiver operating characteristics curve of the partial least squares discriminant analysis (PLS-DA) model for each validation step of cross-model validation.

| Sample location | CMV Accuracy | AUC  |
|-----------------|--------------|------|
| Femur           | 73.6%        | 0.77 |
| Tibia           | 66.0%        | 0.74 |
| Patella         | 83.7%        | 0.91 |
| All combined    | 72.8%        | 0.79 |
samples were misclassified as OARSI 2.5–6 group samples in the tibial group. The result is likely explained by tibial cartilage having slightly different composition and structure compared to the femoral cartilage [31]. Thus, the tibia samples are more easily misclassified as the majority of the samples used for training the model were from femoral cartilage.

Another limitation related to the technique itself is the low penetration depth (few micrometres) of infrared light in FTIR-ATR sampling. Thus, the method is limited to the analysis of cartilage surface. However, this does not prevent the detection of cartilage degeneration. FTIR-ATR spectroscopy is highly sensitive to changes in the molecular composition, and even small alterations in the constituents of the cartilage surface can be detected. The composition of cartilage changes at different depth of the tissue, and FTIR imaging studies of histological sections have shown that there are major differences between the spectra acquired from different depths [24,25]. If the degeneration has progressed so far that the original surface is no longer present, the classification models could also incorporate the depth-related spectral differences in addition to the degeneration related changes in classifying the cartilage FTIR-ATR spectra. It is also possible that PBS content remaining within the tissue during measurements varies between healthier and more degenerated samples. Thus, some of the absorbance differences in the Amide I region could be attributed to that effect. ATR technique relies on stable contact between the surface of the tissue and the ATR crystal. If the surface is degenerated enough that cartilage is missing, measurements will not be reliable. Since this amount of degeneration is easily seen by the surgeon, it does not affect potential clinical use.

As a summary, current methods, such as the International Cartilage Regeneration & Joint Preservation Society (ICRS) classification, for clinical arthroscopic evaluation of cartilage health are subjective and show relatively poor reproducibility [9,32]. Fiber optic FTIR-ATR spectroscopy during arthroscopic joint surgeries could provide valuable information during the surgery and allow objective measures to guide the surgeon. Highly damaged cartilage can be visually seen during arthroscopy, but objective methods for assessing the extent of damaged regions might help guide the extent of interventions taken by the surgeon. Identifying regions with early-stage tissue degeneration, which cannot be seen with current methods, could be important for prevention of OA. Our findings indicate that FTIR-ATR spectroscopy is a promising method for objective assessment of cartilage health.

Contributions

Vesa Virtanen: Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. Valeria Tafintseva: Methodology, Software, Validation, Data curation, Writing – review & editing. Rubina Shaikh: Investigation, Writing – review & editing. Ervin Nippolainen: Investigation, Data curation, Writing – review & editing. Julian Haas: Investigation, Writing – review & editing. Heikki Kröger: Investigation, Writing – review & editing. Johanne Solheim: Software, Data curation, Writing – review & editing. Boris Zimmermann: Software, Writing – review & editing. Achim Kohler: Resources, Supervision. Boris Mizaikoff: Resources, Writing – review & editing. Mikko Finnila: Writing – review & editing. Lassi Rieppo: Conceptualization, Methodology, Software, Validation, Investigation, Writing – review & editing. Supervision, Funding acquisition. Simo Saarakkala: Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Role of the funding source

This work was supported by the European Union’s Horizon 2020 research and innovation programme (H2020-ICT-2016-2017) project MIRACLE (grant agreement number 780598), and the Academy of Finland (project numbers 310466 and 315820).

Declaration of competing interest

All authors declare no conflicts of interest.

Icmje disclosure form

In the interest of transparency, we ask you to disclose all relationships/activities/interests listed below that are related to the content of your manuscript. “Related” means any relation with for-profit or not-for-profit third parties whose interests may be affected by the content of the manuscript. Disclosure represents a commitment to transparency and does not necessarily indicate a bias. If you are in doubt about whether to list a relationship/activity/interest, it is preferable that you do so. The following questions apply to the author’s relationships/activities/interests as they relate to the current manuscript only.

The author’s relationships/activities/interests should be defined broadly. For example, if your manuscript pertains to the epidemiology of hypertension, you should declare all relationships with manufacturers of antihypertensive medication, even if that medication is not mentioned in the manuscript.

In item #1 below, report all support for the work reported in this manuscript without time limit. For all other items, the time frame for disclosure is the past 36 months.

| Name all entities with whom you have this relationship or indicate none (add rows as needed) | Specifications/Comments (e.g., if payments were made to you or to your institution) |
|---|---|
| All support for the present manuscript (e.g., funding, provision of study materials, medical writing, article processing charges, etc.) | HORIZON2020: Grant agreement number 780598 |
| No time limit for this item. | MIRACLE Academy of Finland |
| | Project numbers 310466 and 315820 |

(continued on next page)
Please place an “X” next to the following statement to indicate your agreement:

___X I certify that I have answered every question and have not altered the wording of any of the questions on this form.

Authors’ disclosure

Authorship. All authors should have made substantial contributions to all of the following: (1) the conception and design of the study, or acquisition of data, or analysis and interpretation of data, (2) drafting the article or revising it critically for important intellectual content, (3) final approval of the version to be submitted. By signing below each author also verifies that he (she) confirms that neither this manuscript, nor one with substantially similar content, has been submitted, accepted or published elsewhere (except as an abstract). Each manuscript must be accompanied by a declaration of contributions relating to sections (1), (2) and (3) above. This declaration should also name one or more authors who take responsibility for the integrity of the work as a whole, from inception to finished article. These declarations will be included in the published manuscript.

Acknowledgement of other contributors. All contributors who do not meet the criteria for authorship as defined above should be listed in an acknowledgements section.

Examples of those who might be acknowledged include a person who provided purely technical help, writing assistance, or a department chair who provided only general support. Such contributors must give their consent to being named. Authors should disclose whether they had any writing assistance and identify the entity that paid for this assistance.

Conflict of interest. At the end of the text, under a subheading “Conflict of interest statement” all authors must disclose any financial and personal relationships with other people or organisations that could inappropriately influence (bias) their work. Examples of potential conflicts of interest include employment, consultancies, stock ownership, honoraria, paid expert testimony, patent applications/registrations, and research grants or other funding.

Declaration of Funding. All sources of funding should be declared as an acknowledgement at the end of the text.

Role of the funding source. Authors should declare the role of study sponsors, if any, in the study design, in the collection, analysis and interpretation of data; in the writing of the manuscript; and in the decision to submit the manuscript for publication. If the study sponsors had no such involvement, the authors should state this.

Studies involving humans or animals. Clinical trials or other experimentation on humans must be in
accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000. Randomized controlled trials should follow the Consolidated Standards of Reporting Trials (CONSORT) guidelines and be registered in a public trials registry.

Studies involving experiments with animals were in accordance with institution guidelines.

Please sign below to certify your manuscript complies with the above requirements and then upload this form at https://www.editorialmanager.com/oac/