Politechnika Wrocławska
Wydział Chemiczny

PRACA DYPLOMOWA

Raman spectroscopy in understanding the pathomechanism of osteoarthritis

Spektroskopia Ramana w zrozumieniu patomechanizmu choroby zwyrodnieniowej stawów

Autor
inż. Maciej Pawelski

Praca została wykonana pod opieką
dr inż. Marleny Gąsior-Głogowskiej
Katedra Inżynierii Biomedycznej
Wydział Podstawowych Problemów Techniki

słowa kluczowe: Raman, FT-Raman, hydroxyapatite, mineralization, bone quality

Osteoarthritis is a common inflammatory cartilage and bone disease, leading to disabilities and in some cases death. Most characteristic symptoms include pain and edema of hands and feet joints. Since this disease attacks by surprise, it is crucial to perform a quick diagnosis. Established methods can be supported by vibrational spectroscopy analysis. Calculation of several factors based on the Raman spectrum of the sample may help in the diagnosis. The aim of this work is a development of software package allowing an automatization of such analysis.

Table of Contents

I. Introduction 3

II. Osteoarthritis 4

III. Raman spectroscopy 7

IV. Analysis automation 11

V. Results and Conclusions 18

VI. References 19

I. Introduction

Osteoarthritis (OA) is a common chronic connective tissue disease of immunological background and unknown etiology, characterized by unspecific inflammation of symmetrical joints, extra-articular changes and systemic symptoms. It leads to disabilities and in some cases death. Main target for the disease are people in their forties and fifties, with women susceptible three times more than men. It's most characteristic symptoms include pain and edema of hands and feet joints and morning stiffness. Besides, OA may also cause more general symptoms like mild fever, muscle ache, tiredness, lack of appetite, loss of body mass. In some cases, Secondary OA may also be induced by diseases such as alkaptonuria, Ehler-Danlos syndrome, Wilson's disease or hemochromatosis. Due to these changes, patients may suffer from disabilities preventing from movement and/or performing general daily tasks, which may lead to death in extreme cases. [1]

However, swift diagnostics may help in early detection of the disease. Usually, it consists of several laboratory examinations, such as ESR (Erythrocyte Sedimentation Rate), fibrinogen and CRP concentration. However, data obtained this way is not directly linked to presence of OA. Examining actual bone tissue yields more direct results. Thanks to Raman spectroscopy, one may examine a bone sample for several key factors that may point to the premature presence of the disease. [2] These factors are:

- **degree of mineralization** – quotient of phosphate and amide I band, indicates a more mineralized collagen matrix and points to the architecture of the bone. Diseased tissue has greater degree of mineralization than healthy one.
- **substitution of carbonate ions in locations of phosphate ions** – quotient of carbonate and phosphate bands, relates to the functional properties of bone and its structural order. Increasing value indicates thickened subchondral and increased structural disorder, may be used as a marker for OA.
- **degree of crystallization (size)** – reciprocal of full width at half maximum (FWMH) of phosphate band, indicates greater crystallinity (a well-ordered crystal lattice) with decreasing bandwidth.

Bandwidths acquired from Raman spectroscopy provide useful information about bone type, its age, disease occurrence, diet or physical activity. Bone structure determines its biomechanical properties. With these factors in mind, analyzing each sample separately requires a lot of time. Raman scattering induced by input laser irradiation makes this particular method valuable for analysis [3]:

- provides means to establish vibrational symmetry of the spectrum
- molecular orientation affects the irradiation, making it easier to tell oriented molecules from random ones.

To assess applicability of FT-Raman spectroscopy for OA detection, an automation process would speed up the process. That's where automating and programming techniques come in handy. Thanks to Python programming language and its external libraries, we have been able to create a desktop application running on Windows operating system which performs necessary calculations automatically. This way, analysis of multiple samples can be easily done in matter of seconds. The drawback is, however, that the program is tuned specifically for one type of FT-Raman setup.

Since the main idea of the project was to speed up the analysis process, the program was designed to work on minimalistic, simple graphical user interface, friendly for people not used to run shell scripts and/or computer science at all. Python allows for

quick processing of data without usage of low-level nuances which are so common for languages such as C. Its robust external libraries (namely NumPy and matplotlib) make advanced mathematical calculations easier and faster, as well as allow for quick plot generation for processed data.

II. Osteoarthritis

Osteoarthritis is a form of arthritis caused by degeneration of joints which results in their pain and disfunction. It is the most prevalent joint disease in the world. As chances of osteoarthritis increase with age, it is one of the most frequent diseases among older people. It's often accompanied by disabilities of knees and hips, resulting in raised difficulty of walking, running or any other leg related commotion. [1]

OA is a chronic disease of connective tissue, characterized by unspecific inflammation of joints and presence of extra-articular manifestations, leading to disabilities or even premature death. Frequency of occurrence of OA is between 0.3% and 1.5% of global population, which are usually people in their forties and fifties. Women tend to suffer from the disease 3 times more often than men. In Poland, it's estimated that 1% of population suffers from OA, which causes around 400 000 people to suffer from related disabilities and one in four of patients need operational treatment because of massive degradation of a massive joint (e.g. knee joint). After 5 years of sickness, half of the patients lose ability to work with all of patients being disabled after 10 years. Approximately 70% of patients are affected by periods of exacerbation and relative remission. 15% of patients are affected mildly by the disease; 10% are affected by long-lasting remission. OA may also rarely have an episodic (palindromic) course. Remissions are more often observed in men than in women. [4]

Disease's usual targets consist of small joints in hands and feet, which may cause ache, edema or exudation in a joint. These symptoms may also be accompanied by morning stiffness, caused by accumulation of edema fluid in diseased tissue, muscle ache, loss of body mass or lack of appetite. Disease's symptoms include symmetric pain and edema in joints of hands and feet, less often knee and arm. Usually, OA develops insidiously in several weeks. 10-15% of patients are affected by symptoms very suddenly in several days – in these cases joint inflammation is not symmetric; mild fever, muscle ache and tiredness are also observed. Most often targets of OA are interphalangeal, metacarpal and metatarsal joints. From there, disease may move to joints of arms, knees, hips and elbows. Joints of upper parts of the body are much more often inflamed than the lower parts.[5]

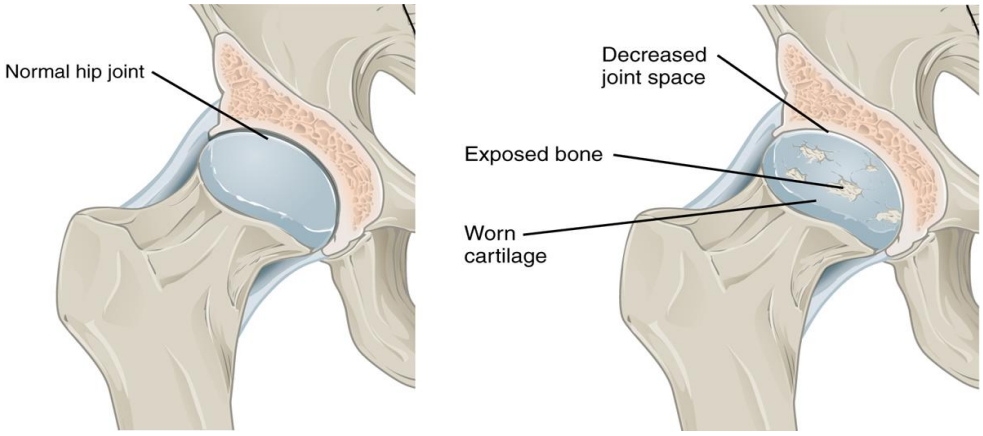

Figure 1: Healthy (left) vs diseased (right) joint. [6]

OA causes different joints suffer from different symptoms:

- **Hand** – in the initial period interphalangeal joints suffer from fusiform edema, decline of interosseus and lumbrical muscles, rash in the vicinity of the withers of the thumb. Later there may be observed deformations of fingers (most often elbow deviation) and hand subluxation of phalanges, which cause considerable encumbrance of finger movability. Due to bone changes, narrowing of joint gap and destruction of tendon-ligamentous apparatus of wrist, there may observed a stiffening of the latter. Growing synovial membrane may press on median nerve, which may lead to carpal tunnel syndrome.
- **Arm** – synovial membrane inflammation of both joints, corobrachial ligament, rotator cuffs, bursae, neighboring muscles and tendons.
- **Elbow** – ache and extension limitation, may lead to lasting flexion contracture.
- **Phalangeal and metatarsal** – inflamed very often in the beginning of the disease, deformation of fingers similarly to hand joints.
- **Ankle** – may be inflamed in heavy, progressing form of the disease, may cause instability and bad positioning of feet.
- **Hip** – ache in the vicinity of groin, increased difficulty of walking.
- **Knee** – rarely inflamed in the beginning, exudation in the joint causes kneecap baling. May cause Baker's cyst to form, which breaking causes diffusion of fluid to shank tissues, causing considerable shank edema and contracture in knee joint.

Figure 2: Rheumatoid nodules in fingers. [7]

OA may also cause changes outside of joints:

- **Rheumatoid nodules** – subcutaneous, painless, occur on straighten surface, mostly on forearm's surface and places susceptible to pressing (e.g. buttocks), tendons, above joints; they also occur in internal organs.
- **Circulatory system** – pericardium inflammation, changes in heart muscle and valves (cardiomyopathy), pulmonary hypertension, atherosclerosis, blood clots and blockages.

- **Respiratory system** – pleura inflammation, rheumatoid nodules within lungs, bronchioles inflammation and lung fibrosis.
- **Eyes** – cornea, sclera and conjunctiva inflammation,
- **Kidneys** – usually related to drug side effects, kidney inflammation, secondary amyloidosis.
- **Other** – inflammation of small and medium blood vessels that may cause necrosis of distal parts of fingers, skin and internal organs; nervous system changes (carpal tunnel syndrome), polyneuropathy, compression of spinal nerve roots caused by destruction of neck spine joints, spleen enlargement, lymph nodes enlargement in axillary, elbow and submandibular parts.

Though pathogenesis of OA is not perfectly known, genetic and environmental factors (e.g. smoking) play a major role. Some of the potential factors related to the increased morbidity of OA include occurrence of common epitope in class II HLA molecules, polymorphism of some matrix metalloproteinases (MMPs), including MMP-1 and MMP-3 as well as polymorphism of pro-inflammatory cytokines, especially of cancer α necrosis factors (TNF- α), interleukins 1β (IL-1β) and interleukins 4. [5]

Frequency of occurrence of several class II tissue compatibility antigens point in patients point to importance of genetic factors in the genesis of the disease. These antigens have been expressed in 70% of people suffering from OA and only 30% of healthy people. Some of these antigens have been found in inhabitants of Europe, North America and Japan.

A number of laboratory tests are usually carried out for diagnosis of the disease, consisting of:

- Erythrocyte Sedimentation Rate – >30 mm after 1 hour
- Increased concentration of fibrinogen and C-reactive protein (CRP)
- Normocytic and hypochromic anemia
- Thrombocythemia (in case of very active form of disease)
- Thrombocytopenia (as drug-related complications)
- Increased concentration of α1 and α2 globulins in plasma
- Rhematoid factor in blood within IgM class
- ACPA (anti-citrullinated protein antibody) – sensitivity >50%, specificity 98%

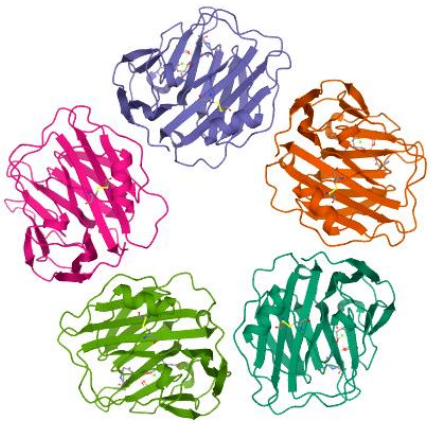

Figure 3: The annular, pentameric shape of C-reactive protein. [8]

In addition to laboratory tests, several other examinations are performed:

- Synovial fluid – presence of inflammatory fluid, increased RF (rheumatoid factor), possible ragocytes
- RTG – allows to see bone and joints changes
- USG – allows detection of synovial membrane inflammation and presence of synovial fluid; may reveal erosion of joint surface; may reveal loss of fiber architecture or tendon rupture
- MR – similar function to USG, can also reveal bone marrow edema and allows for much quicker detection of destructive changes

Goal of the treatment is clinical remission or at least little activity of the disease if remission is not possible. Such goal should be achieved within around 6 month. Pharmacological treatment include usage of drugs modifying course of the disease, which have fundamental significance in OA treatment and should be used immediately after the diagnosis due to their ability to prevent destructive changes within joints. Drug choice depends od activity and duration of the disease as well as previous treatment and prognostic factor (RF). Usually, first drug of choice is methotrexate, used to slow down multiplication of unwanted cells. For ad hoc calming of the symptoms nonsteroidal anti-inflammatory drugs are used. If the disease attacks a single joint, injection of glicocorticosteroids may be considered [5].

Equally important as drug usage in OA treatment is rehabilitation. Patients are advised to take part in kinesinotherapy and physical therapy. The former increases muscle strength improves physical fitness as well as prevents from contractures and deformations, while latter has analgesic, anti-inflammatory and relaxing effect for muscles. To keep the mind equally maintained, psychological support is highly recommended. In extreme cases, surgery may be considered. Such cases may be high pain despite maximum preservative treatment or destruction of the joint, resulting in limitation of movement to the point of disability. Several surgery types may be used [5]:

- synovectomy,
- reconstructive surgery,
- arthrodesis,
- alloplasty,
- corrective surgery.

III. Raman Spectroscopy

Biological sample are characterized by its intricate structure and varying chemical composition. Often, the examined component occurs in very small concentration within the sample. However, spectroscopic methods are well designed for examinations of such samples. Spectroscopy is nondestructive, has low invasiveness and enables for detection

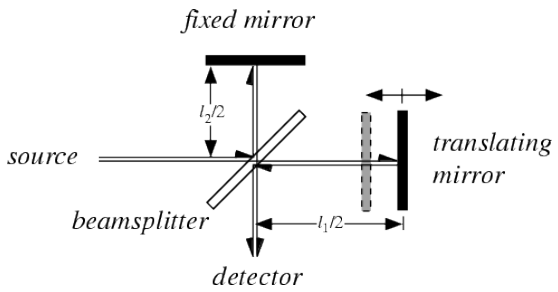

Figure 4: Fourier spectrometer general mode of action. [9]

of changes within tissues even in the early stages of their formation. They are also quick and can be performed in vivo, allowing for skipping of painful biopsy operations [3].

Featured in this work is Raman spectroscopy, which uses non-elastic scattering of electromagnetic radiation phenomenon. This phenomenon is related to change in energy of scattered photons. This change yields information about oscillatory energy levels of the examined sample. Lasers of wide range are the source of radiation in Raman spectroscopy. As a result, one gets a spectrum consisting of oscillatory bandwidth related to vibrations of molecules, which are used in chemical analysis. In comparison to oscillatory infrared spectrum, intensity of the bandwidths is decided by change of polarizability during the vibration instead of change of electric dipole moment. Thus, in Raman spectroscopy, the prevailing vibrations are those changing polarizability of the bond, such as C=C or S-S. [3]

The Raman effect, on which Raman spectroscopy is based was first discovered by Indian physicist Chandrasekhara Venkata Raman in 1920s. But it was not until 1980s that Raman spectroscopy was commonly used in biological sample examination, thanks to application of lasers. Due to light scattering by the sample, photons of changed energy occur alongside unchanged photons. Thus, in spectrum there are additional Stokes and anti-Stokes bandwidths, which have accordingly decreased and increased frequencies. These bandwidths are placed on both sides of Rayleigh bandwidth and are usually much weaker than it.

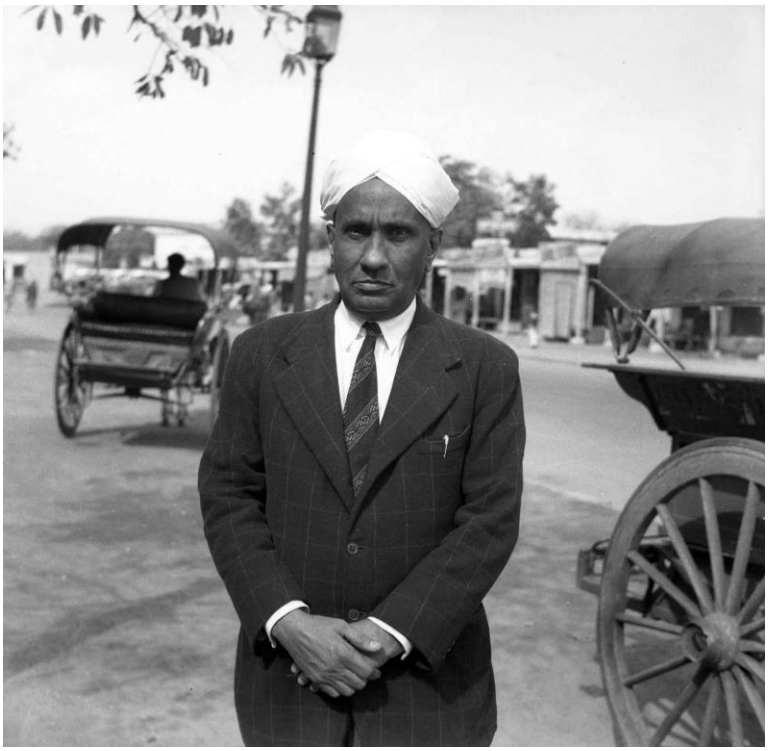

Figure 5: Chandrasekhara Venkata Raman (1888 - 1970) [10]

The Raman Effect can be explained either by wavelike or corpuscular representation of electromagnetic radiation. The first one assumes that the electromagnetic wave affects molecules as a collection of electric charges. Electric field of an electromagnetic wave polarizes this collection, causing it to become a dipole. Its size depends on how easy the molecule is to be polarized. This dipole oscillates with the changing electric field of the electromagnetic field and its own molecular oscillations.

This way, an electromagnetic wave is scattered. Such wave has several components. The main one is the Rayleigh component, which has the same frequency as the wave. Components which have their frequencies modified by their own vibrational frequencies are called the Raman components. With either decreased or increased frequency, the component is called Stoke or anti-Stoke respectively. Stoke bandwidths are usually stronger, thus more often analyzed in Raman spectroscopy. However, due to Rayleigh bandwidth being three orders of magnitude stronger than Stoke and anti-Stoke bandwidths, the apparatus required for Raman spectroscopy needs to be exceptionally sensitive.

In corpuscular representation, electromagnetic radiation can be treated as a photon stream. Let us consider photons that reach the analyzed sample and photons emitted by that sample. Radiation caused by the laser excites the molecules of the analyzed sample to a particular energy state. By returning to ground state, these molecules emit photons. Similarly to wavelike representation, if frequency of the photons remains unchanged, then Rayleigh scattering is observed. If the frequency changes, Raman scattering is observed. However, it is quite weak, since only one in 10^6 of radiation quanta is affected by it. [3]

The most useful bandwidth in Raman spectroscopy is Stokes bandwidth, which are much stronger than anti-Stokes one. This is due to fact, that most of the samples are analyzed in room temperature, meaning they are found in the zero oscillation level. In Raman spectroscopy, Raman spectrum of Stokes scattering is obtained. [11]

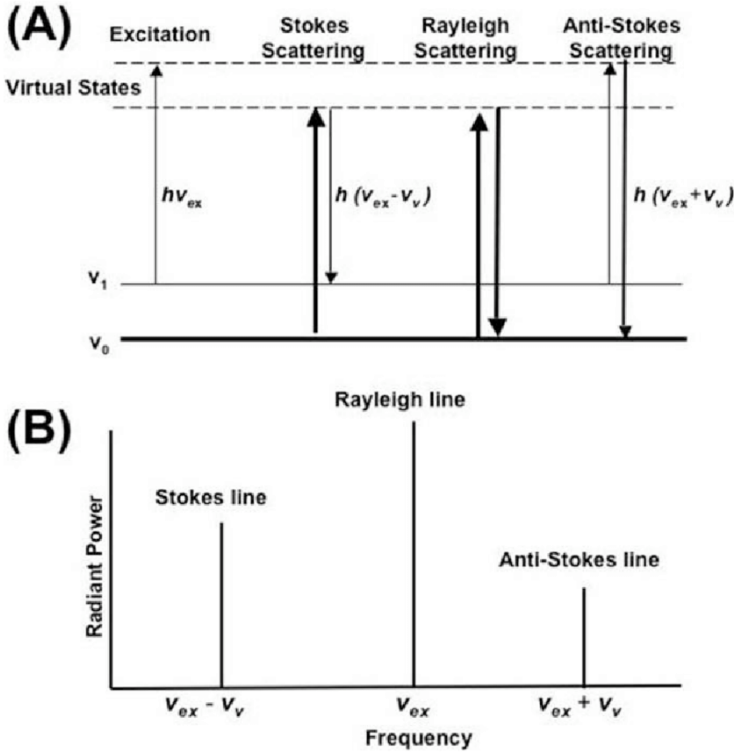

Figure 6: Diagrams of energy levels (A) and radiant power (B) of Raman effect. [11]

In this particular work, Raman spectroscopy was used to analyze three main factors – degree of mineralization, degree of crystallization and substitution of carbonate ions in locations of phosphate. Thanks to this technique, we're able to determine various characteristics of bone structure. One of the main pros of Raman spectroscopy is the fact, that the sample doesn't need any chemical preparation, which is the case with, for

example, IR spectroscopy, which needs sample to be dehydrated. For these reasons, this technique is widely used wherever one needs to examine bone composition, including dentistry or implantology. [12]

First one of the examined factors is degree of mineralization. It described by a ratio of mineral content to the collagen content within bone. Increased value points to a more mineralized collagen matrix, which causes the bone to be more brittle due to collagen molecules being less mobile. For this factor, bandwidths of importance are phosphate bandwidth, which can be found at around 960 cm^{-1} and amide I bandwidth, which can be found at around 1670 cm^{-1}.

Since bone is a biological sample, its bioapatite mineral crystallites may reveal some substitutions within the crystal lattice. The substitution which constitutes for the second factor is the substitution of carbonate ions for phosphate ions. These ions are represented within 1070 cm^{-1} and 960 cm^{-1} bands respectively. This factor points to the changes in mineral crystal form, changing symmetry and creating gaps within the crystal lattice. The change of shape may thus prove additional mechanical challenges due to alteration of strain. This value also increases greatly with age, which may point to the age of the bone.

The last relevant factor is the degree of crystallization. It is calculated by inversing the value of full width at half maximum (FWMH) of the phosphate band (960 cm^{-1}). Increasing value of crystallinity suggests a more ordered crystal lattice, which may be related to a decrease of plasticity of the bone and alteration of strain within the matrix to mineral crystals. This poses a danger of production of cracks within bone structure and cause structural disorder. [12]

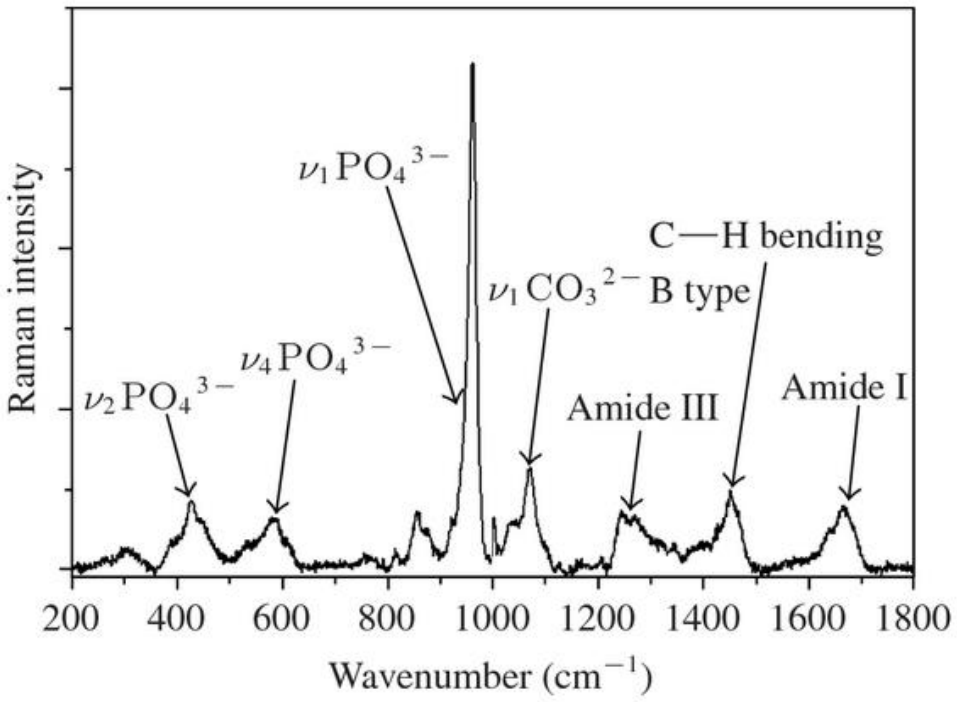

Figure 7: A typical Raman spectrum of bone sample with all important bandwidths marked. [13]

Even with help of typical software, analysis of great number of samples takes long time. Since OA is a quickly developing disease, time is crucial in diagnosis and treatment. Spectroscopic research usually yields thousands of results, which manual analysis is rather pointless. Other beneficial factors include:

- less variability,
- greater control,
- reduced operational costs,
- optimized work time.

IV. Analysis automation

Development of automatization is not a temporary trend, it's a direction that points modern business and industry to the future. Companies invest large amount of money into automating their processes to provide products and services of the highest quality. Through optimalization of processes, one can achieve the best results, thus usage of modern and efficient automation tools has become crucial for cutting-edge businesses. It has become a necessity in modern world, a condition to maintain and develop the product or service within big, competition-filled market. [14] Through automatization, one can:

- increase work performance
- improve product or service quality
- lower labor cost
- mitigate effects of labor deficiency
- eliminate routine, monotonous tasks
- improve work safety
- shorten process realization
- enable implementation of processes which may not be done by hand.

Thanks to various programming tools, automatization is becoming more and more prevalent in day-to-day lives. Examples include automated fridges, vacuum cleaners, cooking appliances (Thermomix) or washing machines. With the increasing need for automation, laboratory appliances are also being automated.

Since manual analysis of thousands of spectra is a tremendously monotonous and time-consuming task, automatization of such process is extremely beneficial, especially in case of diagnosis of such a quickly developing disease like OA. In this work, Python programming language together with some of its external libraries have been used to create application that automizes the analysis process.

Created in early 1990's by Dutch programmer Guido van Rossum, Python quickly gained reputation as an easy and accessible programming language. Since it leaves more low-level programming intricacies, it has become a preferred language for learning programming. Due to its simple syntax, great extensibility, portability and implementation of multiple programming paradigms it swiftly found its way in various programming area such as:

- Web application development (Django and Flask frameworks)
- Machine learning and artificial intelligence
- System deployment and provisioning (Ansible)
- Scientific and numerical applications (NumPy)
- Game development
- Image processing

Especially in scientific field, the simplicity of Python has been appreciated by programmers and scientists around the globe, resulting in takeover of this programming

area, which was previously led by languages such as Perl or Fortran. Simple Python syntax allows to quickly and efficiently write scripts working on, for example, protein database files or spectra. Extensive numerical libraries allow for easy data manipulation and analysis, decreasing researcher's work time. Python has also spawned many external libraries and programs designed strictly for biological applications, including:

- Biopython – library of tools for biological computation
- PyMOL – molecular visualization system
- PyCogent – library for genomic biology
- Biskit – allows for manipulation and analysis of protein complexes, macromolecular structures and molecular dynamics trajectories.

Application which is the key element of this work has been divided into two separate parts. First one performs the analysis on the acquired data. Second one provides a simple, user-friendly interface. The analysis has been divided into several steps:

1. acquisition of data
2. removal of wave number values below 250 cm^{-1} which are overshadowed due to Rayleigh scattering
3. baseline cut
4. search for closest to the ν1 phosphate band at 960 cm^{-1} assigned for stretching vibrations of PO_4 groups, characteristic in mineral phase Raman spectra.
5. normalization of the plot to $\nu 1PO_4$ band at 960 cm^{-1}
6. reading of intensity, FWMH and area of bands at 960, 1070 and 1670 cm^{-1}, where the last two corresponds to carbonate ions and proteins (Amide I), respectively.
7. analytic comparison – carbonate substitution, sample mineralization & crystal

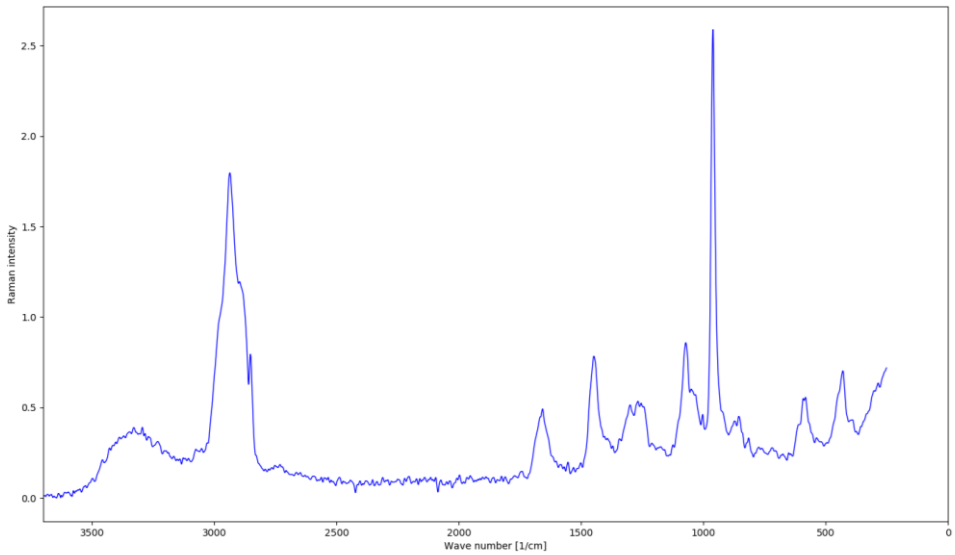

Figure 8: Raw FT-Raman spectrum for human bone sample.

To work, the application needs several libraries to be imported. These are NumPy, SciPy, matplotlib and peakutils, all of which are external.

```
import numpy as np
from scipy import sparse
from scipy.sparse.linalg import spsolve
from matplotlib import pyplot as plt
from scipy.signal._savitzky_golay import savgol_filter as sgf
from scipy.signal import peak_widths, argrelextrema
import peakutils as pku
```

Figure 9: Libraries imported into the analysis part of the program.

The initial form of a spectrum is a CSV file with two columns of values: wave number (X axis, $1/cm^{-1}$) and Raman intensity (Y axis, no unit). Following the loading of data, program performs the initial transformations, which is ensuring that dot is used as decimal separator and removal of wave number values below 250 which are overshadowed due to Raman scattering. Therefore, values on X axis are greater than or equal to 250.

For further processing, spectrum needs to be smoothed which is done using a digital filter. The goal is to increase the quality of the signal by increasing the signal-to-noise ratio while maintaining its initial shape. In the application, Savitzky-Golay filter function, which is a part of SciPy library, is used. Smoothing is achieved in a process called convolution, by fitting subsequent subsets of data to a low degree polynomial by means of linear least squares method. This filtering technique is based on established mathematical procedures and was designed by. A. Savitzky and M. J. E. Golay in 1964. [16] Two arguments are passed into this function:

- window length – the number of coefficients
- order of polynomial – used to fit the samples

These values have been set to 15 and 2 respectively, which are most usual for both smoothing and retaining the original shape.

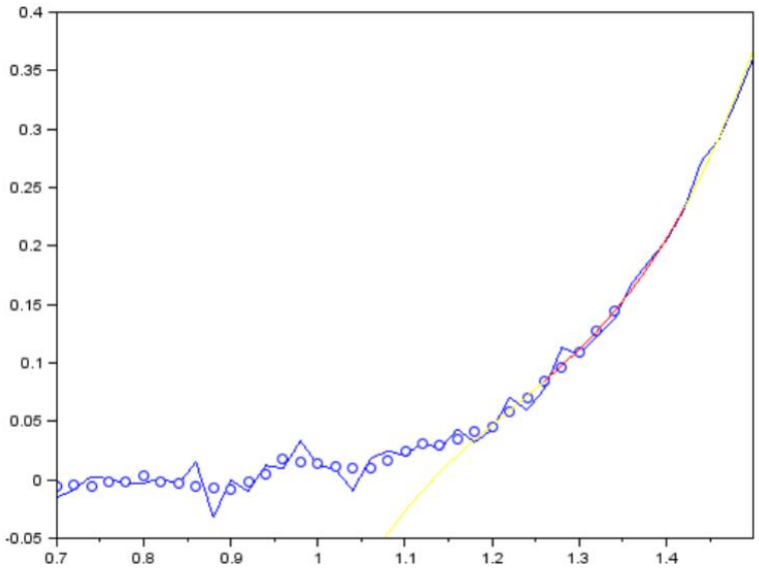

Figure 10: Showcase of Savitzky-Golay's filter convolution process. [17]

Smoothed data is then processed further by subsequent baseline cutting function. It is based on the Asymmetric Least Squares Smoothing algorithm, created by P. Eilers and H. Boelens in 2005 [18]. Intensities of the spectrum are a vector $y = \{y_1, y_2, \ldots, y_i\}$, defined as i. Additional smoothing vector $z = \{z_1, z_2, \ldots, z_i\}$ is faithful to y. Penalized least squares function:

$$F = \sum_i (y_i - z_i)^2 + \lambda \sum_i (\Delta^2 z_i)^2$$

is then minimized with:

$$\Delta^2 z_i = (z_i - z_{i-1}) - (z_{i-1} - z_{i-2}) = z_i - 2z_{i-1} + z_{i-2}, i \in [1, 2, 3, \ldots, n]$$

where Δ is a second-order differential operator. The λ parameter adjusts the balance between fitness and smoothness of the data and should be within range of 10^2 –10^7 for heavy smoothing. Then, fitness weights are defined as w vector and the function is minimized:

$$F = \sum_i w_i (y_i - z_i)^2 + \lambda \sum_i (\Delta^2 z_i)^2$$

Minimization of the equation above may lead to a subsequent equation:

$$(W + \lambda D^T D) z = W y$$

With W = diag(w), W is a diagonal matrix for w vector, D is a second-order differential matrix ($Dz = \Delta^2 z$) and T is the transposition of the matrix. For additional precision, ensuring that proper noise is being removed, penalty parameter p is introduced. Its value ranges from 0 to 1 and is calculated thusly: $w_i = p$ if $y_i > z_i$ and $w_i = 1 - p$ otherwise. [18] Within the program, the function based on the algorithm is being called with following parameters:

- $p = 0.001$.
- number of iterations = 10
- $\lambda = 10^7$

```
def baseline_als(y, lam, p, niter=10):
  L = len(y)
  D = sparse.diags([1,-2,1],[0,-1,-2], shape=(L,L-2))
  w = np.ones(L)
  for i in range(niter):
    W = sparse.spdiags(w, 0, L, L)
    Z = W + lam * D.dot(D.transpose())
    z = spsolve(Z, w*y)
    w = p * (y > z) + (1-p) * (y < z)
  return z
```

Figure 10: Asymmetric Least Squares Smoothing algorithm implemented in Python

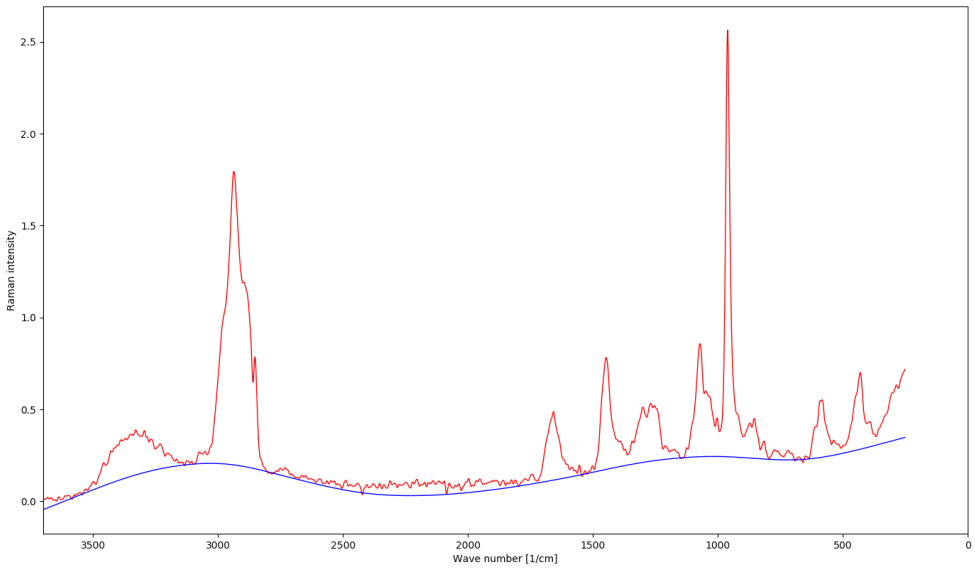

Figure 11: Baseline (blue) cut with the Asymmetric Least Squares Smoothing algorithm

Following baseline cutting, data needs to be normalized for proper analysis and comparison between spectra. Since the most characteristic band in the bone Raman spectra is the 960-phosphate band, it has been chosen for normalization. To normalize the data, every value has to be divided by the value of the phosphate band peak. Application uses peakutils (pku) library to find the desired peak indexes. To ensure that only relevant peaks are found, function pku.indexes accepts two additional arguments. First one is normalized threshold, allowing only peaks with amplitude higher than the threshold to be detected. Second one is minimal distance between peaks. To prevent irrelevant peaks in the neighborhood of 960 band from appearing, these arguments have been set to values 0.2 and 30 accordingly.

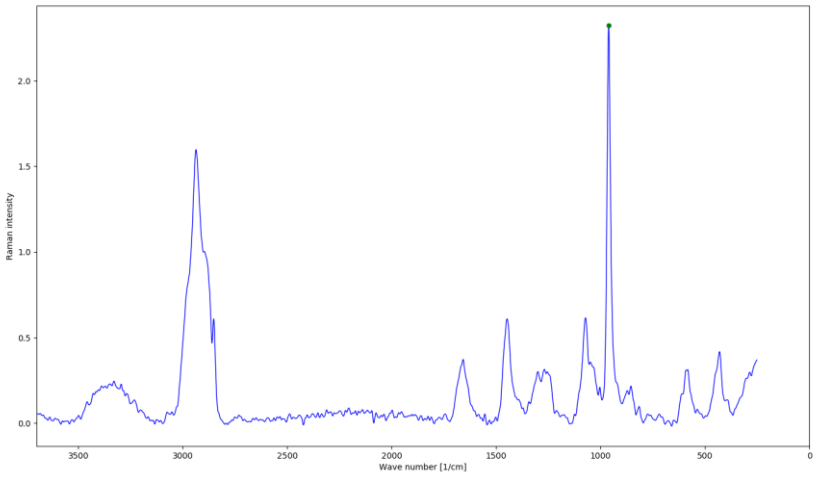

Figure 12: FT-Raman spectrum of human bone with the $\nu1(PO_4)$ band at 960 cm^{-1} marked (green).

Out of available peaks found by pku.indexes, program is looking for the one lying closest to 960 wave number. This is done by a simple function which takes an array of values, subtracts 960 from the absolutes of these values and returns the element with the minimum absolute difference.

Found 960 peak value is then used to normalize the data. This is done by dividing every value by the value of phosphate peak, so that the peak itself has the value of 1 after normalization. Normalization allows the samples to be compared and further analyzed.

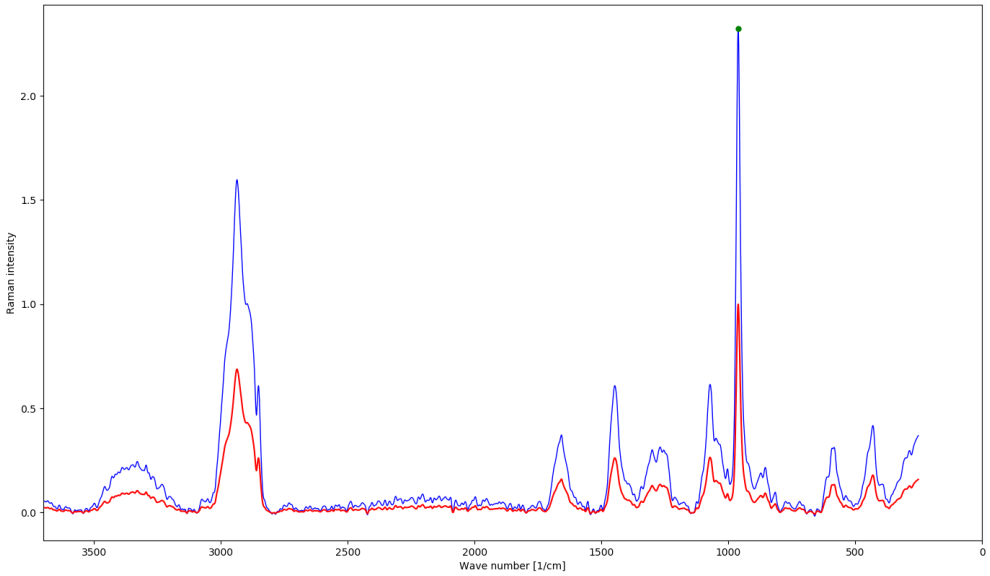

Figure 14: FT-Raman spectrum of human bone before (blue) and after (red) normalization.

The crucial step for the analysis is finding other desired peaks – carbonate (1060) and amide I (1670). This is done similarly to the previous step. However, to achieve more precise results, program compares areas of the desired bandwidths instead of their peak height. Instead of searching for peak values, program now searches for the minimal values. Then, chooses the two values surrounding the peak by their indexes with regards to the set index threshold. Area is then calculated by integrating along the axis limited by the two surrounding minima. This is done by NumPy's trapz function, which uses the composite trapezoidal rule of integration.

The last value required for analysis is full width at half maximum of the phosphate band. SciPy library provides the exact function for this – peak_widths. This tool takes the array of values as an argument and calculates the band's width at the set height. Value obtained is then reciprocated to achieve the degree of crystallization.

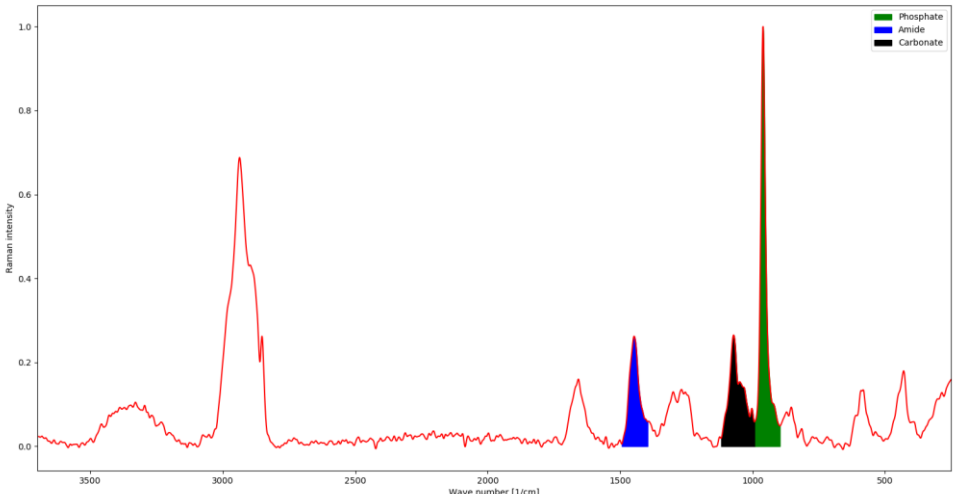

Figure 15: FT-Raman spectrum of human bone with the desired bandwidths colorized.

If the program runs a single analysis, results obtained in this part of the program may are plotted using matplotlib library, which allows for advanced plot manipulation and results are shown in a popup window. Multiple analysis creates CSV file with results for all the selected samples.

To ensure the accessibility of the application, a simple graphical interface has been created using PyQt, a Python binding for the Qt cross-platform C++ framework. In a minimal window, user can choose between two modes of analysis – single and multiple. The first one processes only one sample, shows plot and enables for its manipulation. Second one takes multiple files and performs the analysis one by one, saving the results to the external CSV file. The path of this file may be changed in the *Settings* menu of the program. By defualt, it is set to the installation path and is saved in the *ramanizator.conf* configuration file, also in the installation path.

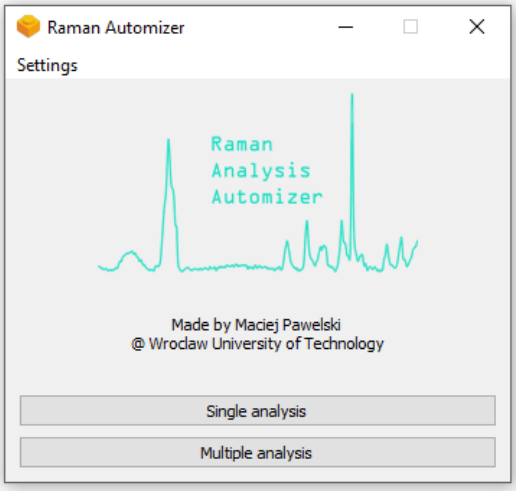

Figure 15: Raman analysis automizer UI.

Program has been designed to work on Windows machines. Thanks to FBS library, one may quickly turn PyQt based program to an easily installable and runnable application. FBS does not limit the program only to Windows. It may be just as well used to create packages for Linux and Mac systems.

Results for 13G-mapa-0009.CSV	
Mineralization	2.173
Substitution	0.611
Phosphate band FWHM	20.54
Crystallinity	0.049
Results for 30s-64skany-550mW-4cm-brzeg-1.csv	
Mineralization	11.57
Substitution	0.192
Phosphate band FWHM	38.47
Crystallinity	0.026
Results for 50S_128sk_4cm_450mmW_poj_2.csv	
Mineralization	3.921
Substitution	0.409
Phosphate band FWHM	36.82
Crystallinity	0.027

Table 1: Results imported into Microsoft Excel

V. Results and conclusions

Results of calculations have been compared between a model healthy pig bone and four subsequent stages of the disease in human bone sample, based on the angle of deformation of the knee joint of the patient. Human samples have been taken from tibia ephiphyses of patients qualified for arthroplasty operation, due to considerable osteoarthritis. [19] These results have been compiled in Table 2. Spectra have been taken by Nicolet NXR 9650 FT-Raman spectrometer. Laser beam has been provided by Nd:YAG excitation laser (1064 nm, 550 mW). Range of spectra has been set between 4000-250 cm^{-1}, with resolution of 4 cm^{-1} and averaging 64 scans. I would like to hereby thank dr inż Anna Nikodem and mgr inż. Justyna Wolicka for sharing these results.

Mineralization increases with each stage of the disease and is the main factor pointing to the presence of OA. Carbonate substitution is also increasing with the

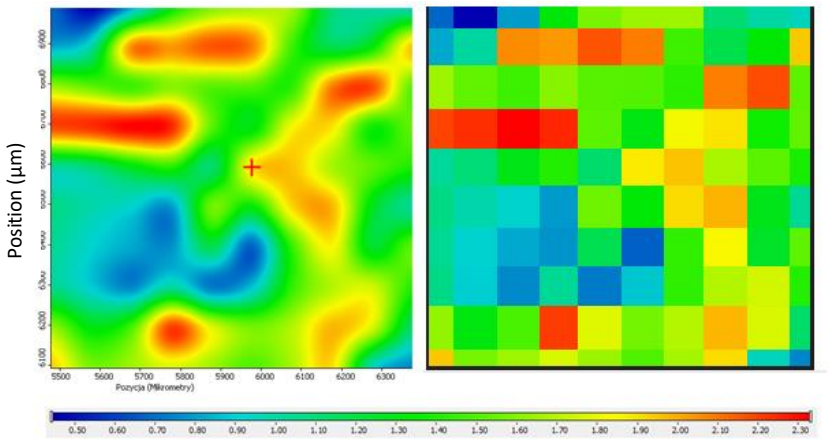

Figure 16: Mineralization map for the pig sample.

exception of Stage 1, which indicate the remodeling activity of bone matrix. Crystallization value for a human bone sample is generally lower than for a pig bone sample and oscillates around the same value of 0.026 with a decreasing trend.

	Mineralization	Substitution	Crystallization	
Healthy	2.41	0.53	0.046	Average
	0.19	0.079	0.00062	Std. deviation
	0.085	0.035	0.00027	Confidence
Stage 1	2.05	0.74	0.027	Average
	0.68	0.24	0.0010	Std. deviation
	0.30	0.10	0.00044	Confidence
Stage 2	2.47	0.55	0.026	Average
	0.93	0.26	0.0027	Std. deviation
	0.41	0.11	0.00099	Confidence
Stage 3	2.74	0.57	0.026	Average
	1.48	0.23	0.0012	Std. deviation
	0.65	0.10	0.00051	Confidence
Stage 4	2.90	0.62	0.025	Average
	2.29	0.28	0.0021	Std. deviation
	1.00	0.12	0.00091	Confidence

Program provides the general overview and trend regarding the change of values for the related factors. This change can be used to further investigate development of osteoarthritis. Better values could be achieved by further tuning the variables within the program for the particular sample and spectrum. Differences between manual and automatic calculations may also be present due to different approaches to e.g. area calculation and bandwidth selection. Particularly helpful would be implementation of machine learning algorithms, which could be used to teach the program to precisely select precise bandwidths. As expected, mineralization and substitution values increase in the subsequent stages of the disease, proving that diseased bone tissue has a more mineralized collagen matrix and disordered subchondral. It's worth noting that not all of the values of these factors may change during course of the disease, as shown with the crystallization values. This shows, that size of the crystals within these particular samples are consistent.

Automation of the analysis done in this work has taken steps in speeding up the process. However, further development is possible and leaves much space to extend capabilities of the application and the analysis itself by addition of tools for e.g. statistical evaluation of data and results. This way, quick diagnosis of osteoarthritis may save more patients from cruel fate of the disease.

I. References

1. S Glyn-Jones, A J R Palmer, R Agricola, A J Price, T L Vincent, H Weinans, A J Carr, Osteoarthritis, The Lancet, Volume 386, Issue 9991, 2015, Pages 376-387, ISSN 0140-6736
2. Kozielski M, Buchwald T, Szybowicz M, Błaszczak Z, Piotrowski A, Ciesielczyk B. Determination of composition and structure of spongy bone tissue in human head of femur by Raman spectral mapping. J Mater Sci Mater Med. 2011 Jul;22(7):1653-61. doi: 10.1007/s10856-011-4353-0. Epub 2011 May 31. PMID: 21626309; PMCID: PMC3127018.
3. Nykiel, P, SPEKTROSKOPIA RAMANA: NOWOCZESNA TECHNIKA W DIAGNOSTYCE MEDYCZNEJ I ANALIZIE BIOCHEMICZNEJ, Biul. Wydz. Farm. WUM, 2013, 4, 27-36
4. Sobierajski, T, Codzienność z reumatoidalnym zapaleniem stawów, Ogólnopolska Federacja Stowarzyszeń Reumatyków, ISBN: 978-83-953807-0-9
5. https://www.mp.pl/interna/chapter/B16.II.16.1. (visited 13.05.2021)
6. https://en.wikipedia.org/wiki/Osteoarthritis#/media/File:0910_Oateoarthritis_Hip_B.png (visited 13.05.2021)
7. https://lh3.googleusercontent.com/proxy/1FfnVPoZ-cARHYgMOOdhnBJ8mkmNIQSBFhwlTDlPr_8zDiLsBc_15TgXg7Bam2G7CLL3Q4gEGJW3Oefku81mlAP31B4l7b7XeydrrFQPSGHM1jI8zcd8IGOEvYfvqgSK4P6B2048-kJhVR3dNghKv6F3vkI (visited 13.05.2021)
8. https://cdn.rcsb.org/images/structures/gn/1gnh/1gnh_assembly-1.jpeg4 (visited 13.05.2021)
9. https://scienceworld.wolfram.com/physics/fimg192.gif (visited 17.05.2021)
10. https://3.bp.blogspot.com/-kVVFoo5l2Fw/WgKWnYAPacI/AAAAAAAABdA/-qGctm2s1-0ObwhA-x01GJhrxt1wCSoZwCLcBGAs/s1600/SIR-C-CV-RAMAN.jpg (visited 17.05.2021)
11. https://www.researchgate.net/profile/Dimitra-Stratis-Cullum/publication/235158125/figure/fig1/AS:648595397169152@1531648601265/a-Energy-level-diagram-demonstrating-Raman-scattering-b-resulting-spectrum-not-drawn.png
12. Paris C, Lecomte S, Coupry C. ATR-FTIR spectroscopy as a way to identify natural protein-based materials, tortoiseshell and horn, from their protein-based imitation, galalith. Spectrochim Acta A Mol Biomol Spectrosc. 2005 Nov;62(1-3):532-8. doi: 10.1016/j.saa.2005.01.023. PMID: 16257757.
13. M. Kazanci, P. Roschger, E.P. Paschalis, K. Klaushofer, P. Fratzl, Bone osteonal tissues by Raman spectral mapping: Orientation–composition, Journal of Structural Biology, Volume 156, Issue 3, 2006, Pages 489-496, ISSN 1047-8477
14. https://www.researchgate.net/profile/Miroslaw-Szybowicz/publication/255697046/figure/fig1/AS:304768124768269@1449673791524/Representative-Raman-spectrum-collected-from-spongy-bone-tissue-showing-the-major-bands.png (visited 17.05.2021)
15. Christopher D. Wickens & Stephen R. Dixon (2007) The benefits of imperfect diagnostic automation: a synthesis of the literature, Theoretical Issues in Ergonomics Science, 8:3, 201-212, DOI: 10.1080/14639220500370105
16. William H. Press and Saul A. Teukolsky, Savitzky-Golay Smoothing Filters, Computers in Physics 4, 669 (1990); doi: 10.1063/1.4822961

17. https://upload.wikimedia.org/wikipedia/commons/8/89/Lissage_sg3_anim.gif (visited 19.05.2021)
18. Eilers, Paul & Boelens, Hans. (2005). Baseline Correction with Asymmetric Least Squares Smoothing. Unpubl. Manuscr.
19. J. Wolicka, Analiza wpływu szpotawości na właściwości fizyczne i strukturalne tkanek podporowych nasady bliższej kości piszczelowej, Wydział Mechaniczny Politechniki Wrocławskiej, 2020