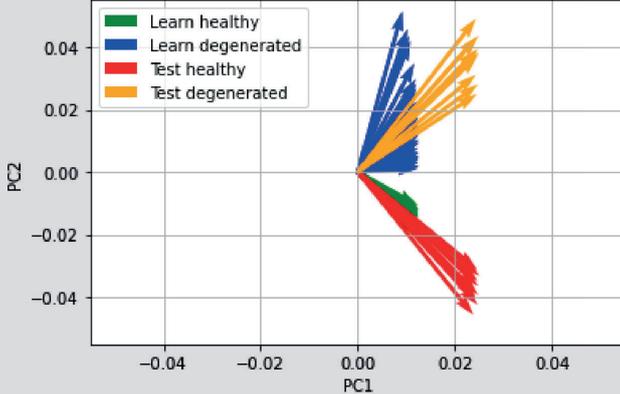


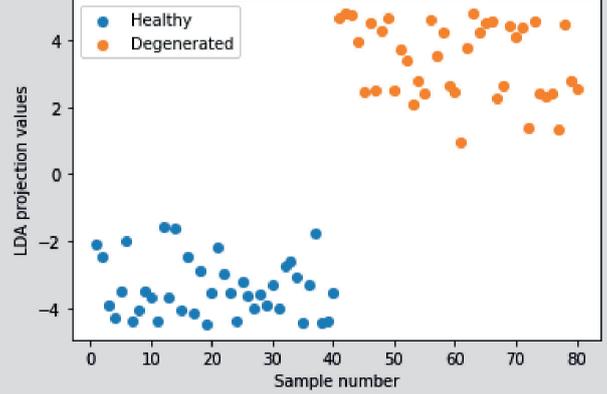
## Klassifizierung von Arthrosepatienten basierend auf klinischen Transkriptomdaten

PCA - Eigenvectors



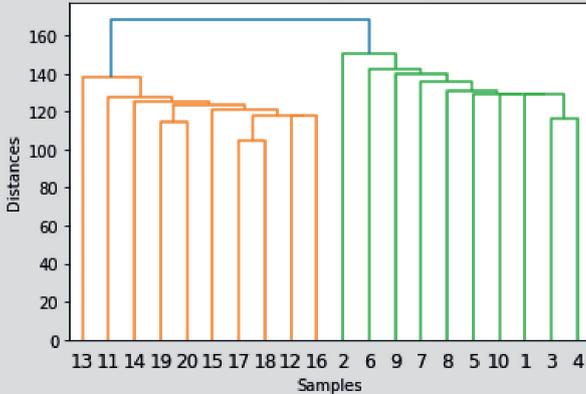
Die Principal Component Analysis (PCA) ermöglicht es, jedem Patient einen Eigenvektor zuzuordnen. Diese Vektoren lassen die Segmentierung zwischen gesunden und arthrotischen Patienten zu.

LDA scatter-plot - Learning

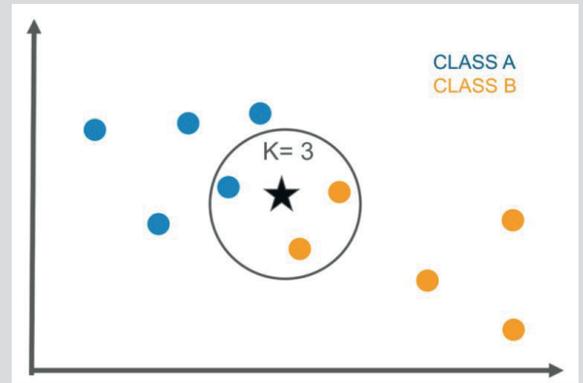


Die Linear Discriminant Analysis (LDA) findet eine Gerade, welche die Klassen möglichst voneinander trennt. Die Y-Achse stellt die Projektionswerte der Samples auf die trennende Gerade dar. Die X-Achse repräsentiert die Nummer der Samples.

Hierarchical Clustering - Dendrogram - Test



Mit Hilfe der Hierarchical Clustering Methode lassen sich Dendrogramme ableiten, welche die Ähnlichkeiten zwischen den Samples darstellen. Orange = Krank, Grün = Gesund



K-Nearest-Neighbor (KNN) bestimmt die Klasse von unklassierten Datenpunkten (Stern) anhand seiner K-nächsten Nachbarn.

### Problemstellung

Diese Bachelor-Thesis ist der Einstieg für die Entwicklung einer medizinischen Software. Diese Software soll in Zukunft helfen, Menschen, die an Osteoarthritis leiden, zuverlässiger identifizieren zu können, um ihnen frühzeitig eine optimale Behandlung zukommen zu lassen. Wie es in der Realität oft der Fall ist, sind die Grenzen zwischen physiologisch und pathologisch nicht immer klar voneinander trennbar. Auch können die Ursachen und Ausprägungen für eine Erkrankung an Osteoarthritis sehr vielfältig sein. Im ersten Teil der Arbeit ist der biologische Hintergrund zu Knorpelgewebe und Osteoarthritis aufgezeigt. Im zweiten Teil geht es um die Analyse von Transkriptomdaten aus menschlichen Chondrozyten.

### Lösungskonzept

Methoden aus dem Bereich Machine Learning helfen die Datensätze von gesunden und arthrotischen Patienten zu unterscheiden. Principal Component Analysis, Independent Component Analysis, Linear Discriminant Analysis und Hierarchical Clustering dienen für die Vorbehandlung der Daten. Anschliessend dient ein K-Nearest Neighbor Algorithmus dazu, die vorbehandelten Daten als gesund oder arthrotisch zu klassieren. Um künftige regulatorische Anforderungen zu erkennen in der Produktentwicklung ist eine Produktklassierung nach den entsprechenden Regulatorien notwendig.

### Ergebnisse

Die Machine Learning Methoden sind mit dem Python Modul Scikit-learn implementiert. Die verwendeten Methoden lassen deutliche Unterschiede in den Samples erkennen und diese auch zuverlässig klassieren. Die hier verwendeten Samples sind künstlich erzeugt und stellen ein idealisiertes Modell von Patientendaten dar. Für die künftige Weiterentwicklung gilt es, die verwendeten Methoden mit klinischen Patientensamples zu validieren. Die Software wird voraussichtlich in eine Klasse C fallen, gemäss In-Vitro-Diagnostic Regulation kurz IVDR.

### Autor

**Fabricio Kirchhofer**

Betreuer:  
Dr. Fabian Ille  
PD Dr. Philipp Stämpfli