Bundesministerium
für Bildung
und Forschung

Otto-Friedrich-Universität Bamberg



# Erklärbare KI für medizinische Anwendungen

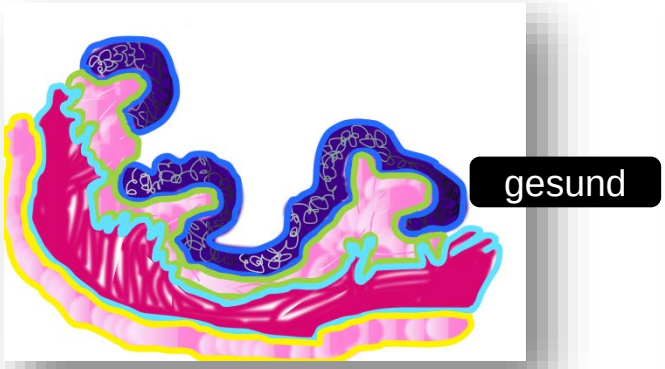
**B.Sc. Bettina Finzel**, Prof. Dr. Ute Schmid
Kognitive Systeme, Otto-Friedrich-Universität Bamberg

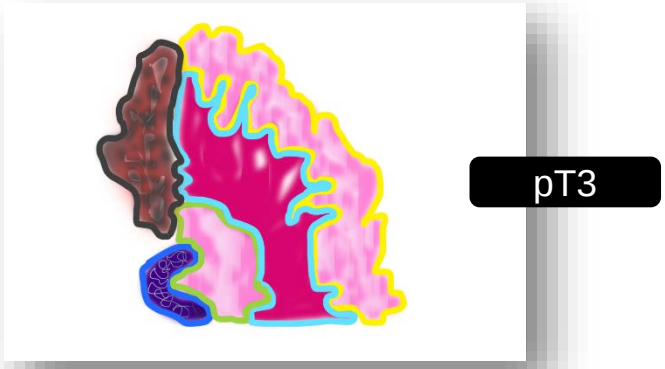
DGBMT TraMeExCo, 29.3.2019, DGE-BV 2019, Stuttgart

Gefördert durch das BMBF, Kennzeichen 01IS18056B (TraMeExCo)

# Anwendungsfall pT Klassifizierung

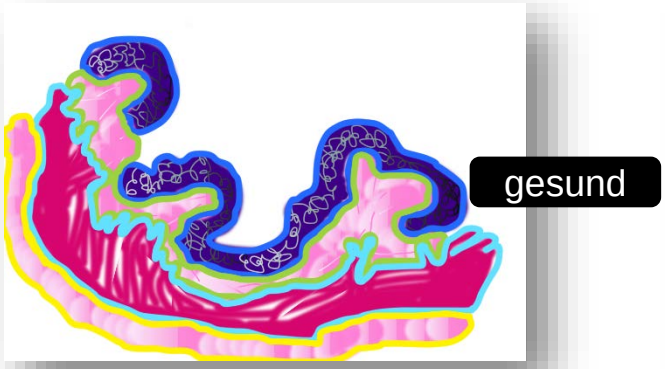
- Bestimmung des Krankheitsstadiums (Größe und Ausdehnung des Tumors)
  - Gesundes Gewebe
  - Tis: Carcinoma in situ, Krebszellen in oberen Schichten der Darmschleimhaut
  - T1: Tumor beschränkt sich auf Darmschleimhaut
  - T2: Zusätzlich zur Schleimhaut ist Muskulatur befallen
  - T3: Tumor ist in alle Schichten der Darmwand eingewachsen
  - T4: Tumor hat sich in benachbarte(s) Gewebe / Organe ausgebreitet

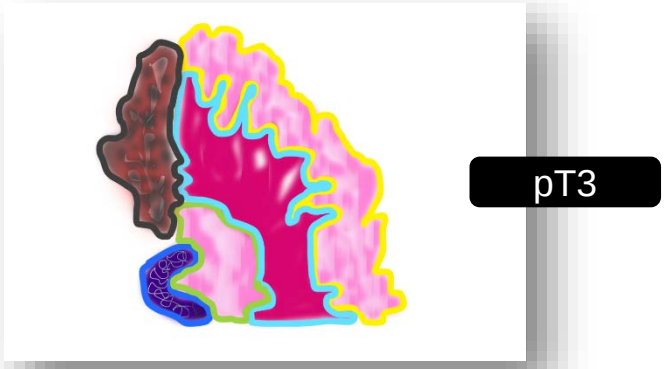


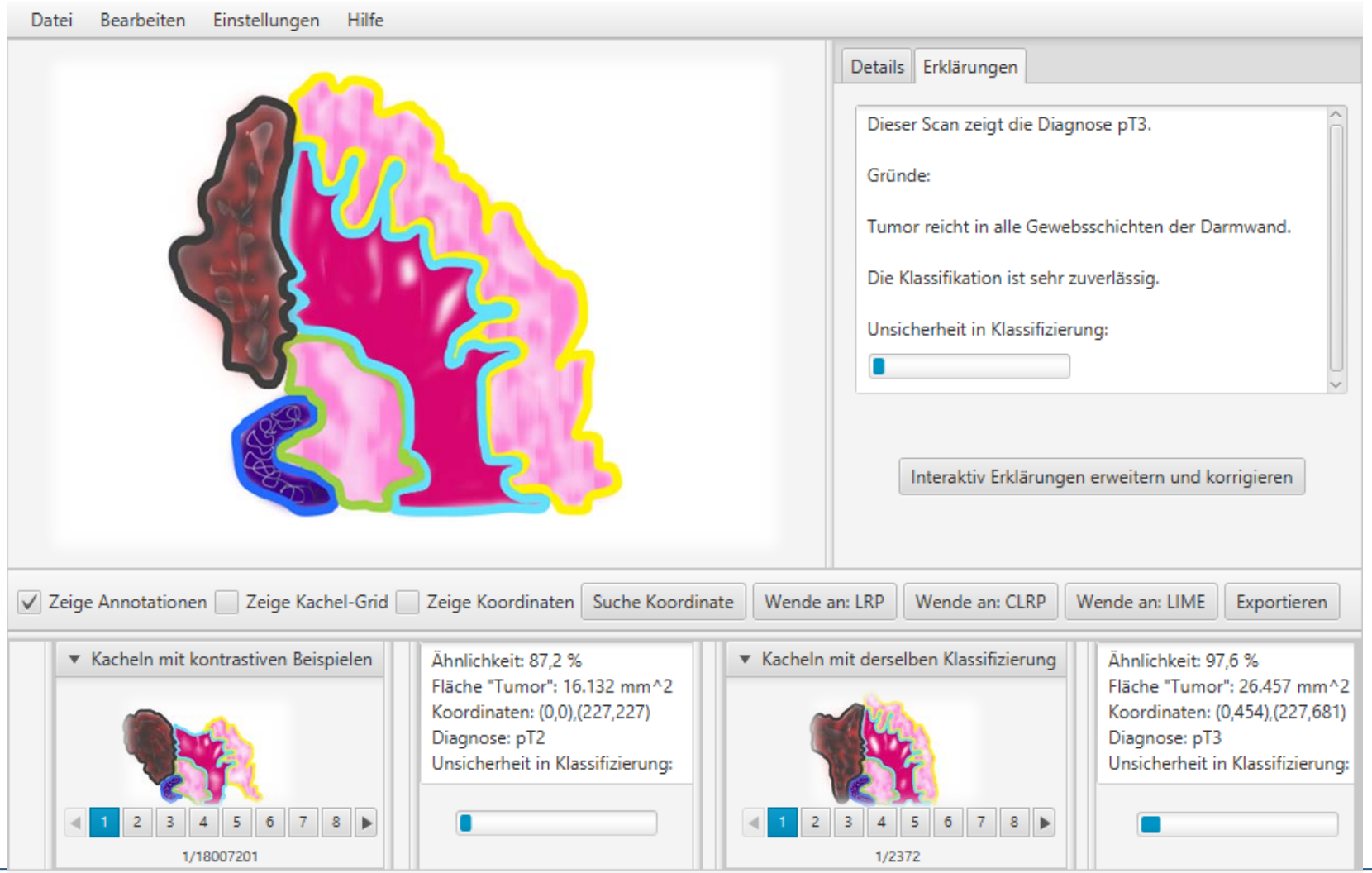
# Anwendungsfall pT Klassifizierung

- Bestimmung des Krankheitsstadiums (Größe und Ausdehnung des Tumors)
  - **Gesundes Gewebe**
  - Tis: Carcinoma in situ, Krebszellen in oberen Schichten der Darmschleimhaut
  - **T1: Tumor beschränkt sich auf Darmschleimhaut**
  - **T2: Zusätzlich zur Schleimhaut ist Muskulatur befallen**
  - **T3: Tumor ist in alle Schichten der Darmwand eingewachsen**
  - T4: Tumor hat sich in benachbarte(s) Gewebe / Organe ausgebreitet



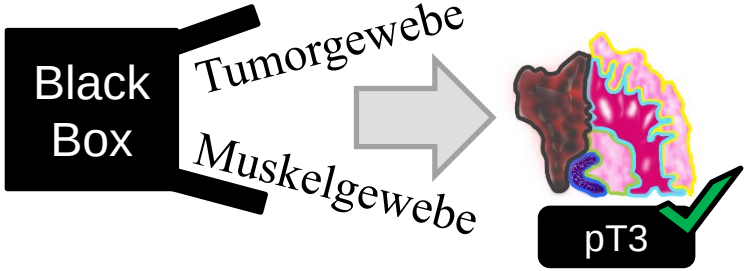
# Transparente Diagnoseunterstützung

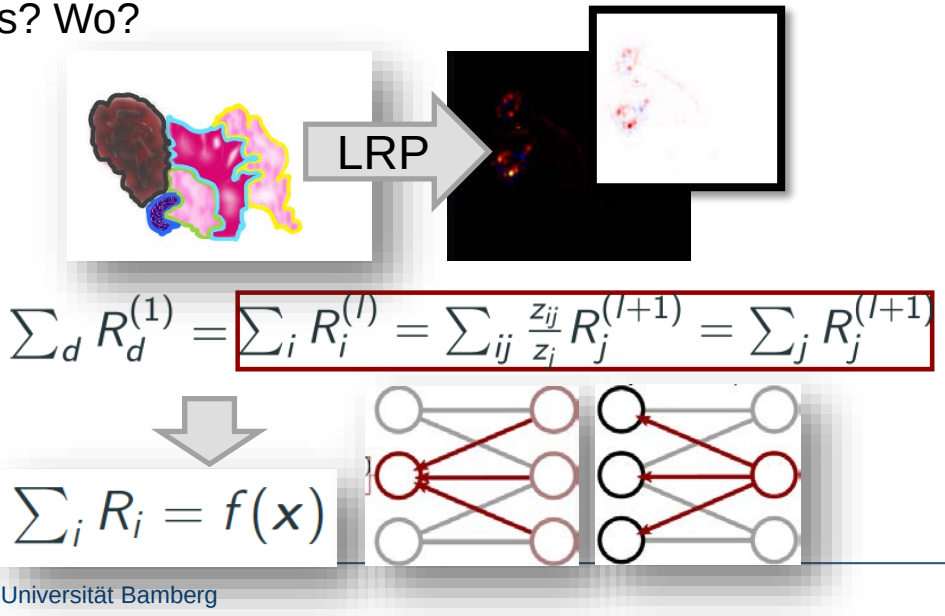


# Erklärbare KI: Was hat das Neuronale Netz gelernt?

- Erklärbare KI macht Black-Box Klassifizierer transparent, indem sie offenlegt, welche Merkmale für eine bestimmte Klassifikation als relevant betrachtet wurden
  - Beispiel: **Layer-wise Relevance Propagation** (Bach et al., 2015)

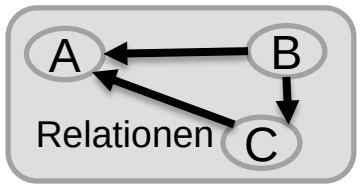
  → Vorhandensein eines Tumors? Wo?

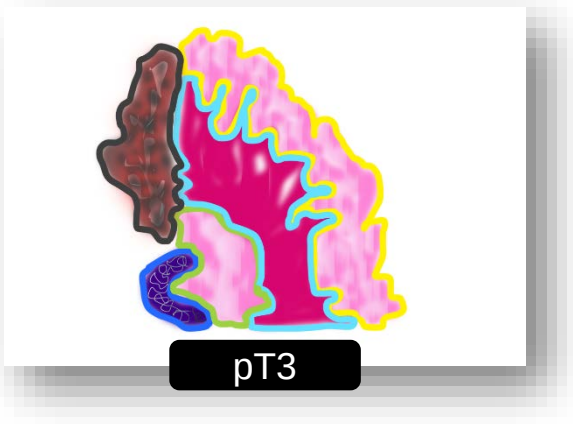


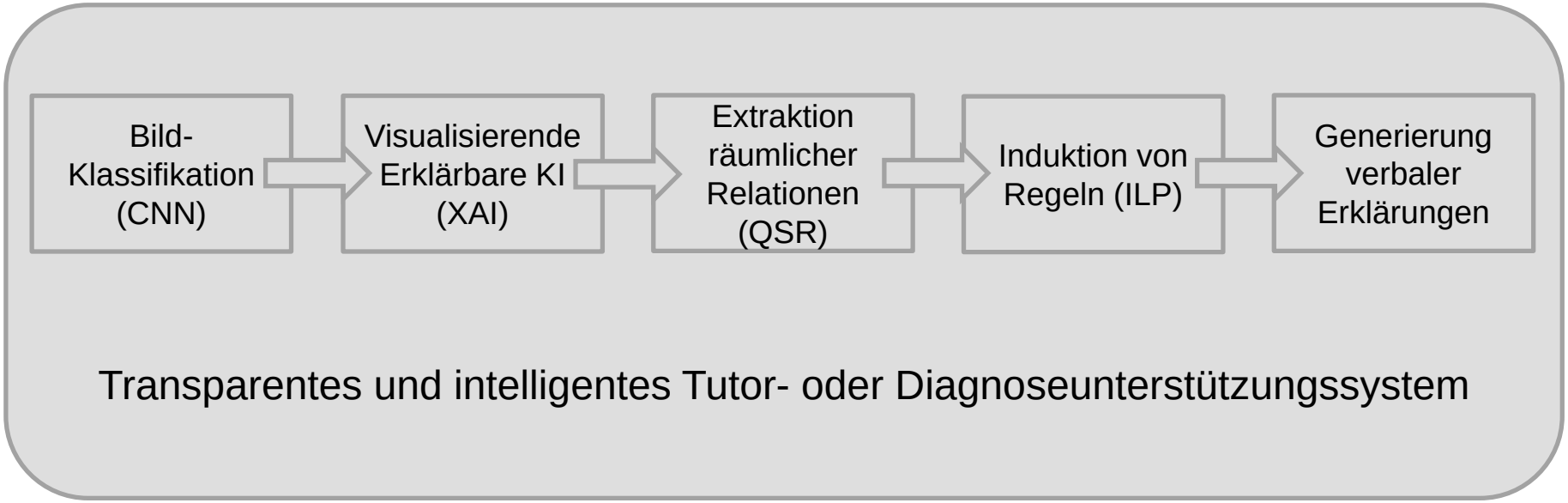
# Erklärbare KI: Was hat das Neuronale Netz gelernt?

- Visualisierende Erklärbare KI kann mit regel-basierten interpretierbaren maschinellen Lernverfahren kombiniert werden (Rabold et al., 2018), um festzustellen, ob komplexe Zusammenhänge, die zu einer bestimmten Klassifizierung führen, berücksichtigt wurden
  - **Inductive Logic Programming** (Muggleton & De Raedt 1994)
  - → Welche Zusammenhänge gelten im mit pT3 diagnostizierten Scan?
    - Tumor wächst in Schleimhaut hinein
    - Tumor wächst in Bindegewebe hinein
    - Tumor wächst in Muskelgewebe hinein
    - → **Räumliche Relationen** (z.B. Renz, 2002)

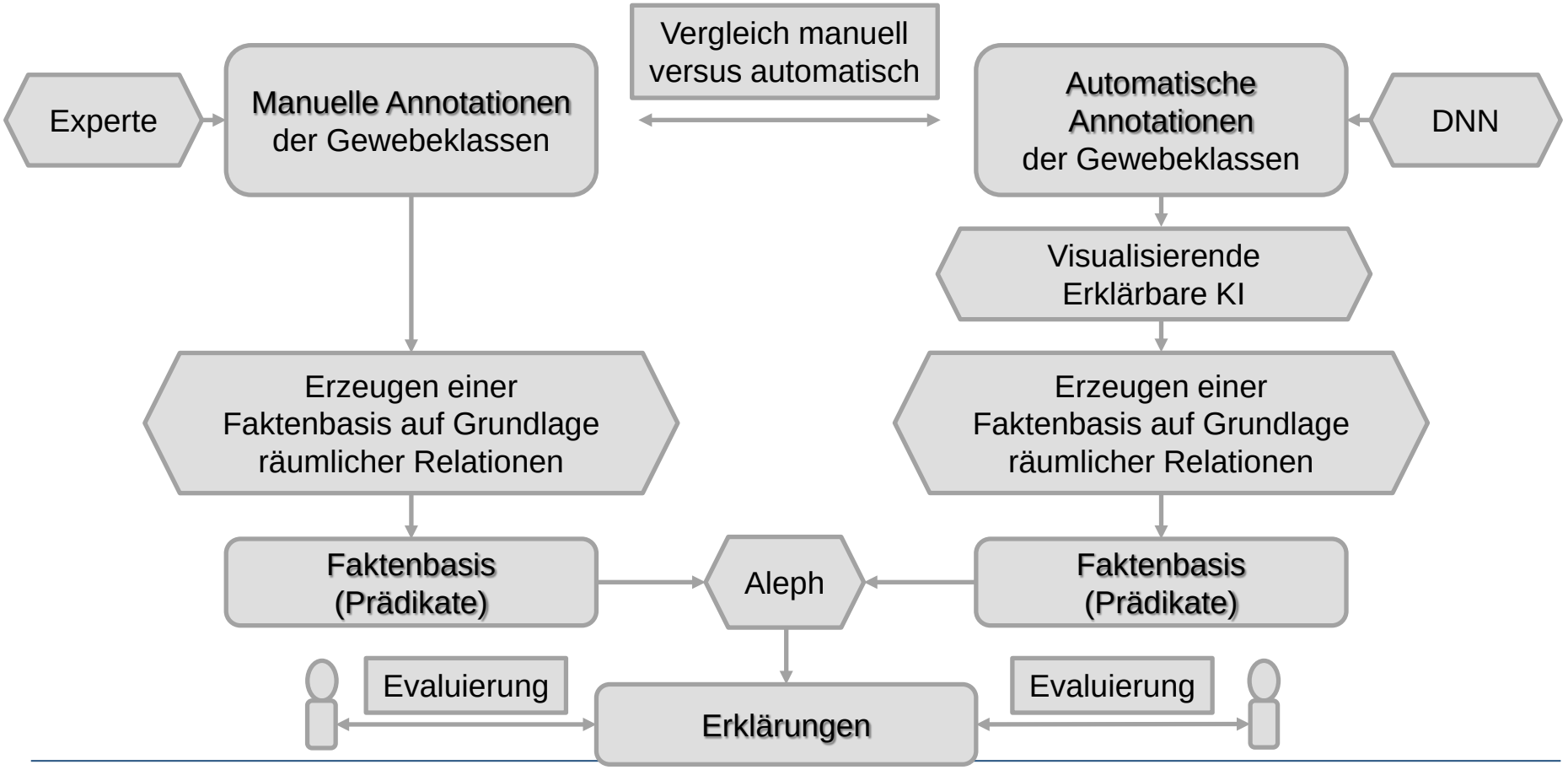


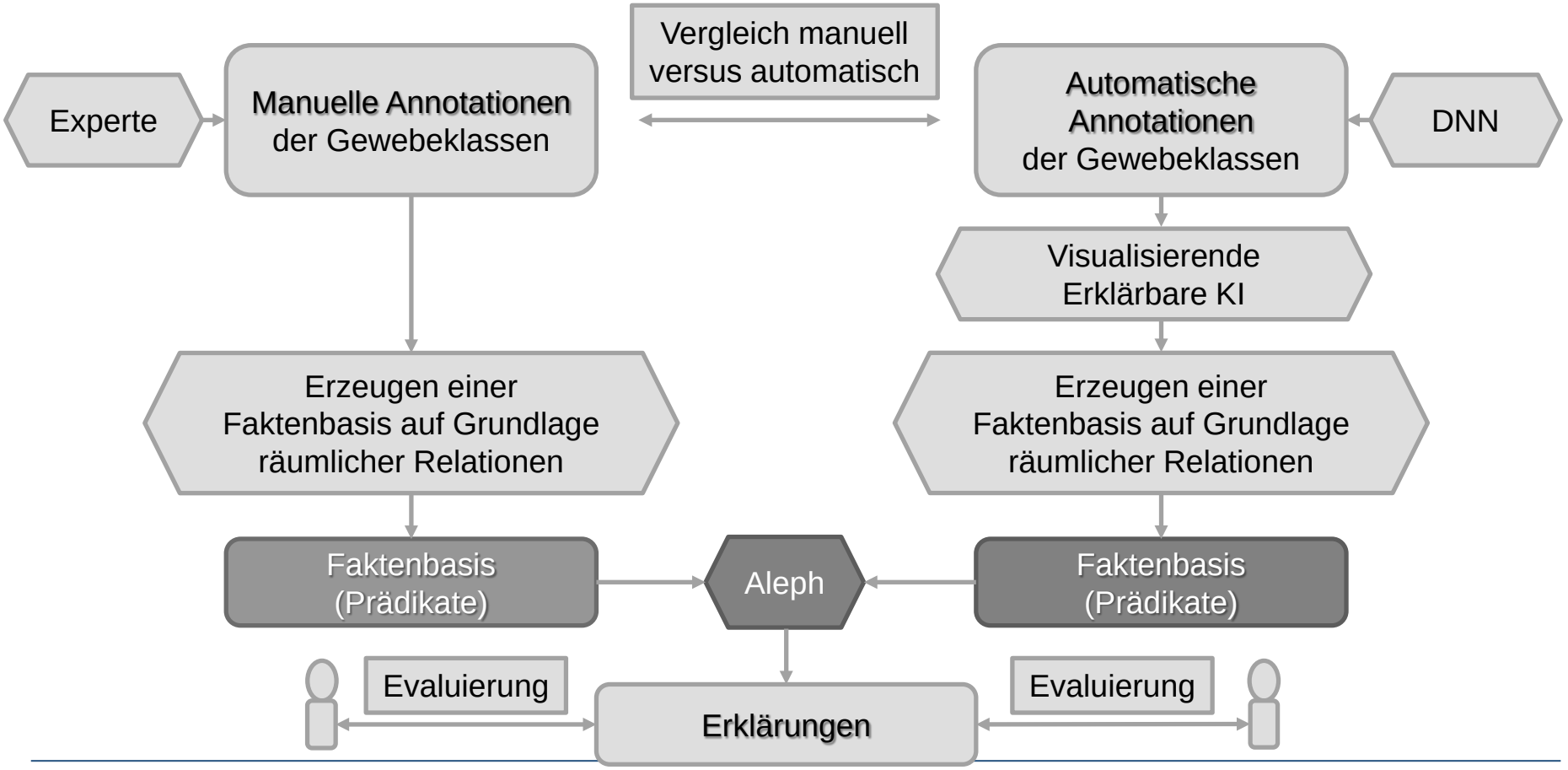
# Unser Ansatz



# Unser Ansatz

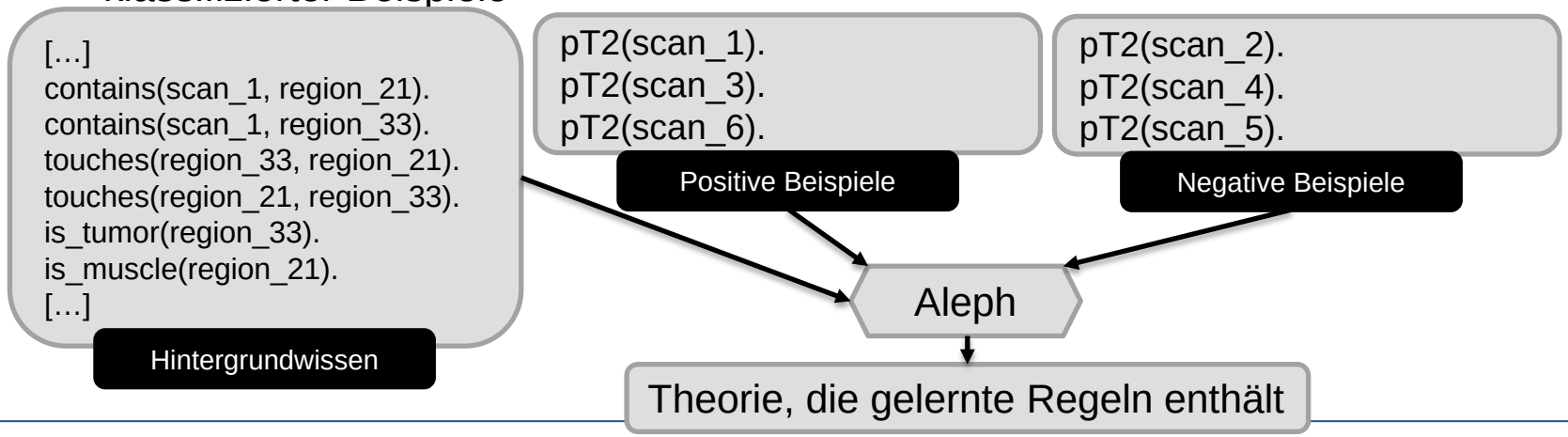


# Unser Ansatz



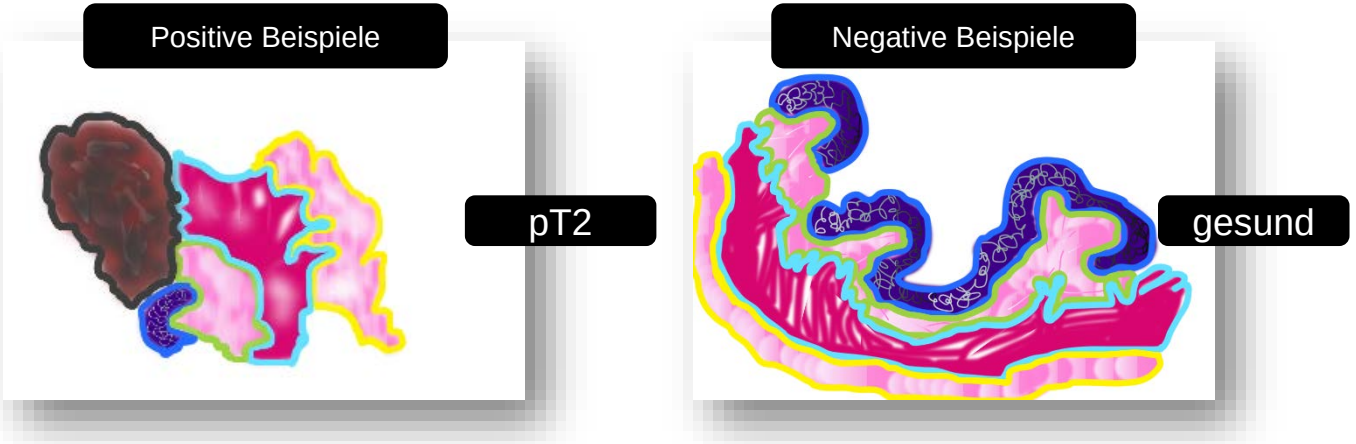
# Implementierung

- **Hintergrundwissen:** räumliche Relationen von jedem Scan (werden aus manuellen Annotationen und automatischen Visualisierungen extrahiert)
- **Symbolische Repräsentation:** Prolog Prädikate
- **Anwendung des Aleph Algorithmus (Induktive Logische Programmierung):** Generierung von Regeln aus Hintergrundwissen und Beispielen (pos. und neg.) zur Erklärung ganzer Klassen oder einzelner klassifizierter Beispiele



# Implementierung

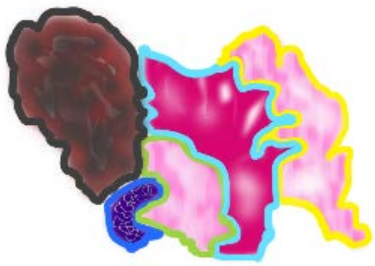
[theory]
[Rule 1] [Pos cover = 3 Neg cover = 0]
**pT2(A) :-**
**is_tumor(B), contains(A,B).**
Accuracy = 1.0



# Implementierung

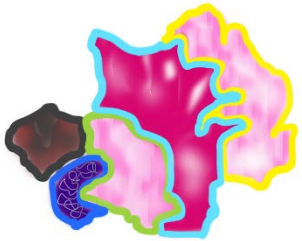
[theory]
[Rule 1] [Pos cover = 3 Neg cover = 0]
**pT2(A) :-**
**is_muscle(B), touches(B,C), is_tumor(C), contains(A,B).**
Accuracy = 1.0

Positive Beispiele



pT2

Negative Beispiele

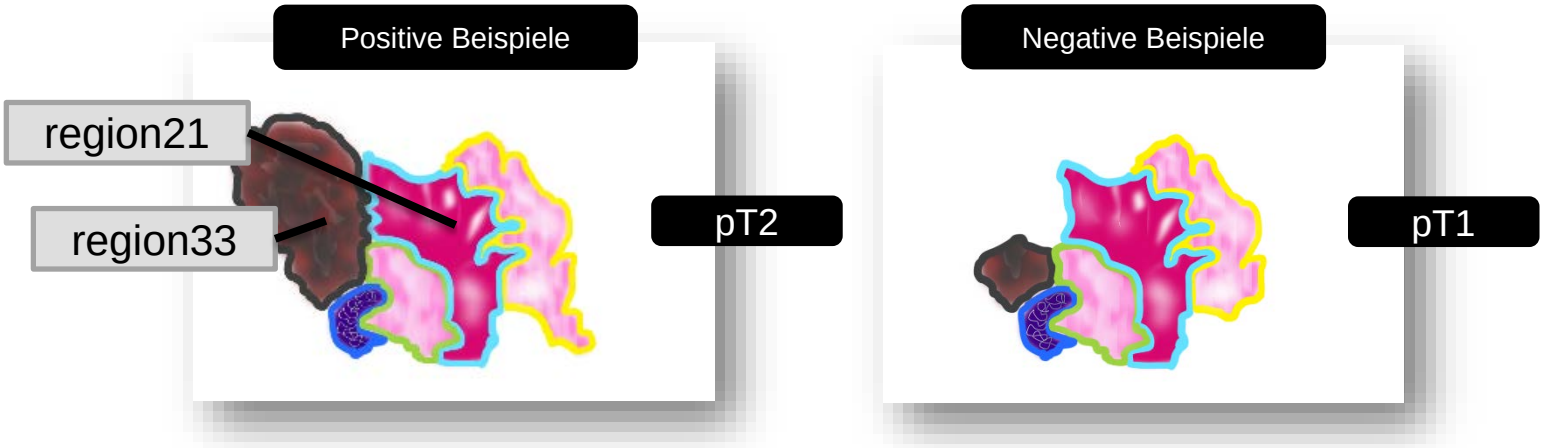


pT1

# Implementierung

**pT2(scan1)**

**is_muscle(region21), touches(region21,region33), is_tumor(region33), contains(scan1,region21).**



# Zusammenfassung

- XAI Methoden und Interpretable ML können genutzt werden, um automatische Klassifizierungen mikroskopischer Schnitte zu erklären
- XAI Methoden ermöglichen es, relevante Regionen zu identifizieren und zu visualisieren
- Interpretable Machine Learning eignet sich, um verbale Erklärungen zu erzeugen und komplexe Strukturen zu untersuchen
- Die generierten Erklärungen können zur Validierung des Klassifizierers verwendet werden
  - → Re-Training Bedarf identifizieren
  - → Experten-Feedback berücksichtigen

**Fazit: Durch die Kombination von XAI Methoden und Interpretable Machine Learning können performante Black-Box Verfahren transparent gemacht werden und in den Diagnoseprozess von Experten einbezogen werden**

# Literatur


- Bach, S., A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, and W. Samek (2015). On Pixel-wise Explanations for Non-Linear Classier Decisions by Layer-wise Relevance Propagation. In: PLoS ONE 10.7, e0130140.
- Muggleton, S., De Raedt, L. (1994) Inductive logic programming: theory and methods. J. Logic Program. 19–20, 629–679 (1994). Special Issue on 10 Years of Logic Programming
- Renz, J. (2002) Qualitative Spatial Reasoning with Topological Information. Springer-Verlag, Berlin, Heidelberg (2002)
- Rabold, J., Siebers, M., Schmid U. (2018) Explaining Black-Box Classifiers with ILP – Empowering LIME with Aleph to Approximate Non-linear Decisions with Relational Rules. ILP 2018. Springer-Verlag, Berlin, Heidelberg (2018)
- Srinivasan, A. (2001). The Aleph Manual (http://web.comlab.ox.ac.uk/oucl/research/ areas/machlearn/Aleph/)