
TraMeExCo

– Transparent Medical Expert Companion



Thomas Wittenberg¹, Michaela Benz¹, Matthias Bergler¹, Bettina Finzel³, Jens Garbas¹, Robert Koch¹, Sebastian Lapuschkin², Nadine Lang¹, Ines Rieger^{1,3}, Wojcieck Samek², Matthias Struck¹, Dominik Seuss^{1,3}, **Ute Schmid**³

¹ Fraunhofer Institut für Integrierte Schaltungen IIS, Erlangen

² Fraunhofer Heinrich-Hertz Institut HHI, Berlin

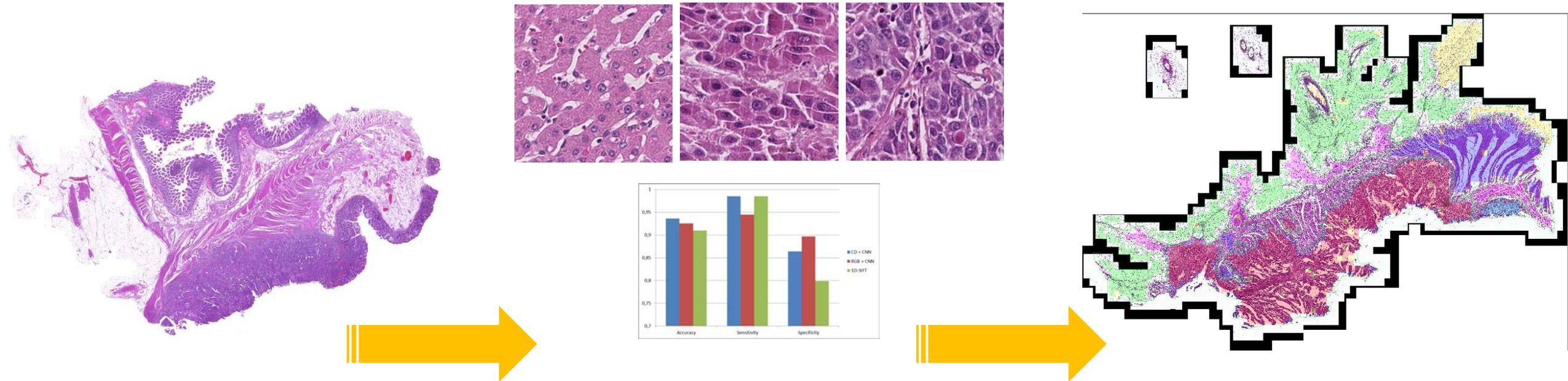
³ Universität Bamberg

Ausgangslage / Motivation

- **Klinische Assistenzsysteme** für die Datenauswertung, Diagnoseunterstützung und Therapieentscheidungen integrieren zunehmend maschinell gelernte Klassifikatoren
- Viele Klassifikatoren sind **Black-box**-Verfahren, die damit getroffenen Entscheidungen sind intransparent für das klinische Personal
- Für komplexe Entscheidungen bzgl. Diagnose und Therapie müssen die **Entscheidungswege nachvollziehbar** und **transparent** gestaltet werden

Anwendungsszenario 1: Digitale Pathologie

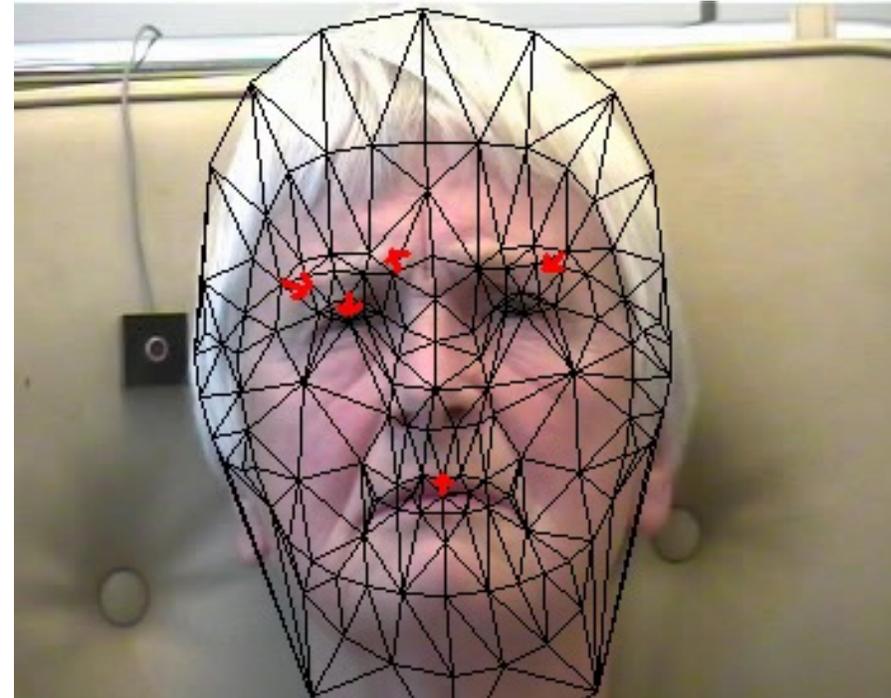
- Klassifikation verschiedener Gewebearten in großvolumigen, histologischen Mikroskopiedaten am Beispiel von Kolonkarzinomen



Aichinger, ..., Wittenberg, et al: Automated cancer stem cell recognition in HE stained tissue using convolutional neural networks and color deconvolution. Proc. SPIE 10140, Medical Imaging 2017: Digital Pathology, 101400N Conf. Vol. 10140, Orlando, FL, Feb. 11, 2017

Anwendungsszenario 2: Schmerzanalyse

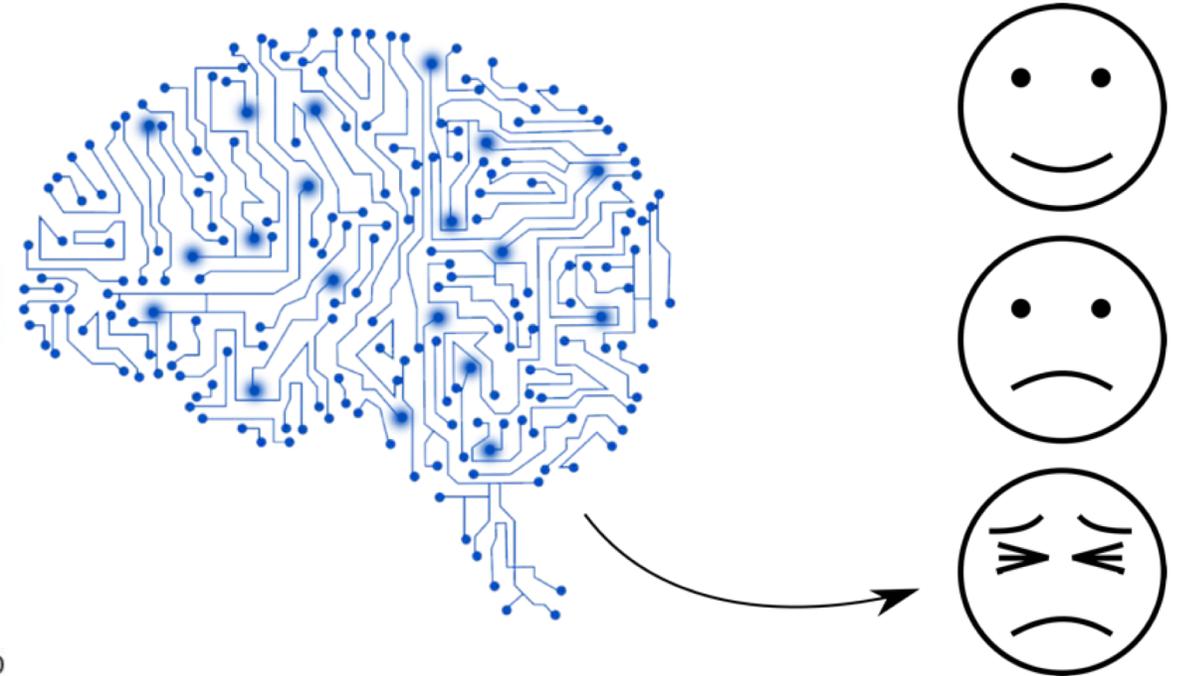
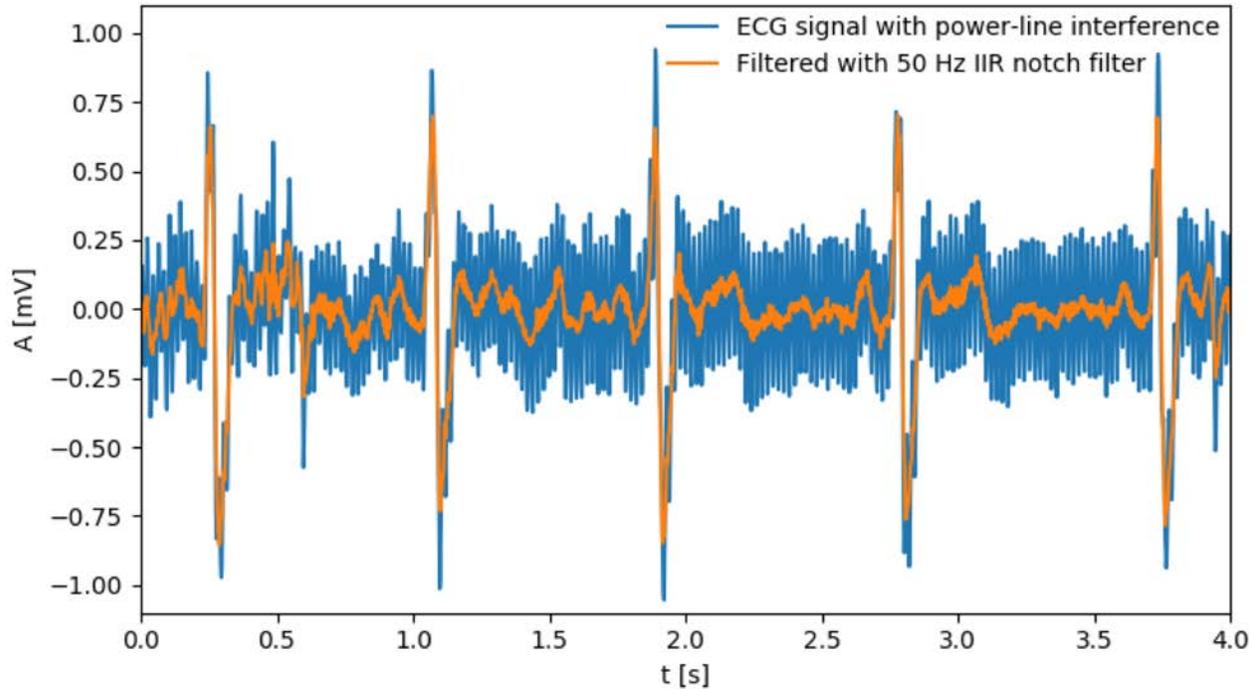
- Erkennung, Unterscheidung und Erklärbarkeit von *Action Units*
- im Kontext der FACS (*Facial Activity Coding System*) Klassifikation von Schmerz



Rieger I, Finzel B, Seuß D, Wittenberg T, Schmid U. Make Pain Estimation Transparent: A Roadmap to Fuse Bayesian Deep Learning and Inductive Logic Programming [2019].
To appear Proc's IEEE BMC

Anwendungsszenario 3: EKG / PGG Analyse

- Robuste Erkennung von R-Zacken in verrauschten Signalen



https://vheart.io/assets/uploads/2017/ecgnoise_figure2.png

Ziele

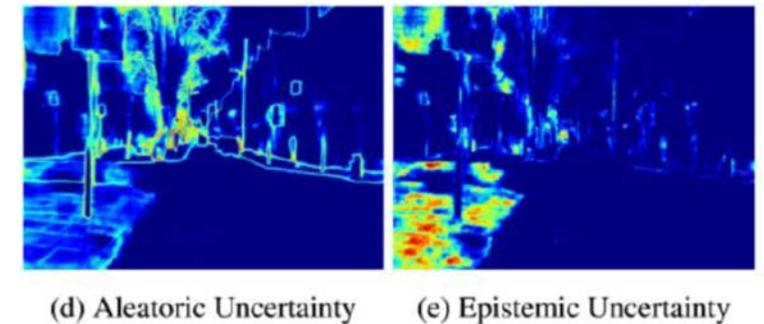
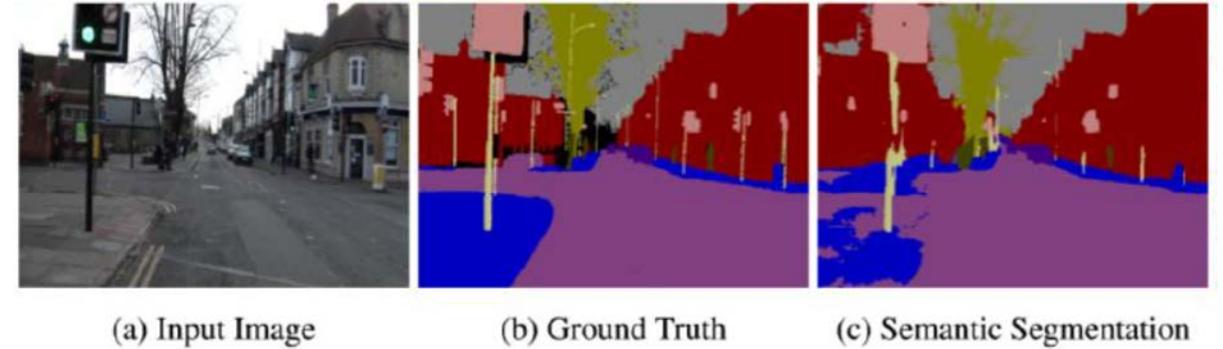
Kombination performanter (aber intransparenten) *Deep Learning* Ansätzen mit

- **XAI-Ansätzen zur Visualisierung** von Entscheidungsprozessen wie z.B.
 - **LIME** (Local Interpretable Model-agnostic Explanations),
 - **LRP** (Layerwise Relevance Propagation)
 - **Heat Maps**
- **ILP** (Inductive Logic Programming) zur Generierung **verbaler Erklärungen** über relevante zeitliche und räumliche Relationen

Weitz, K., Hassan, T., **Schmid, U.** & Garbas, J. (2018). Towards explaining deep learning networks to distinguish facial expressions of pain and emotions. Forum Bildverarbeitung (Karlsruhe KIT/Fraunhofer IOSB).
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. In *Int. Conf. on Inductive Logic Programming* (pp. 105-117). Springer.

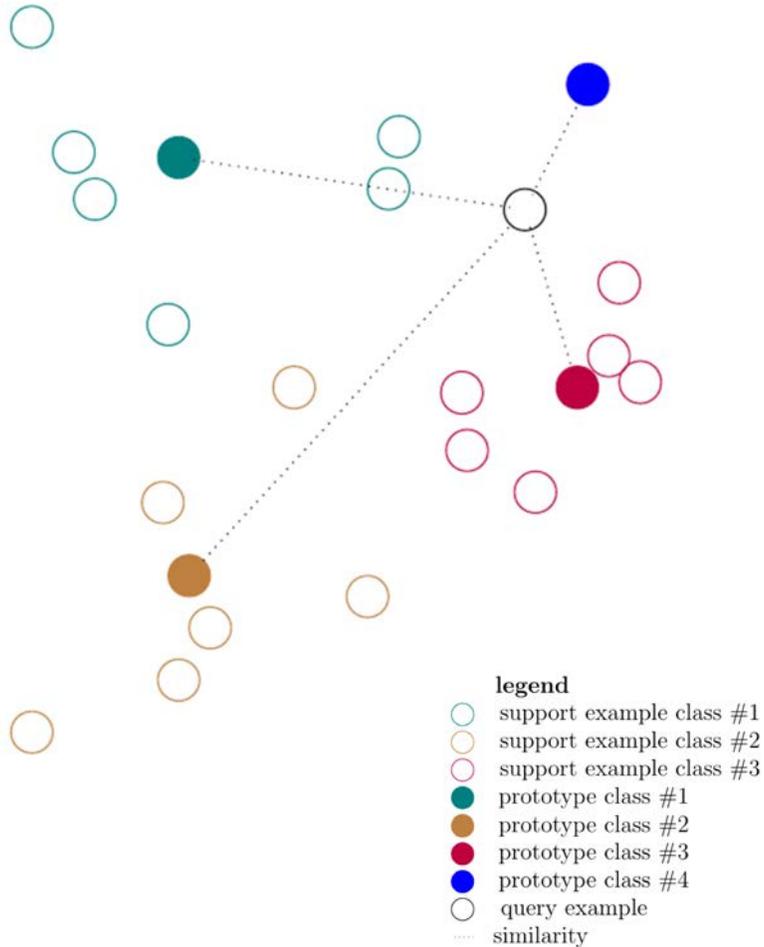
Spezifische Anforderungen

- Modellierung, Darstellung, und Kompensation **epistemischer** und **aleatorischer Unsicherheiten**
 - **Bayes'sches „Deep Learning“**
- Adaptierbare Klassifikation: Hinzufügen / Definition neuer Klassen zur Laufzeit
 - → **Few-Shot-Learning**



Kendall & Gal (2017) What uncertainties do we need in Bayesian Deep Learning for Computer Vision? *Advances in Neural Information Processing Systems*.

Beispiel Few-Shot Learning: Prototypical Networks



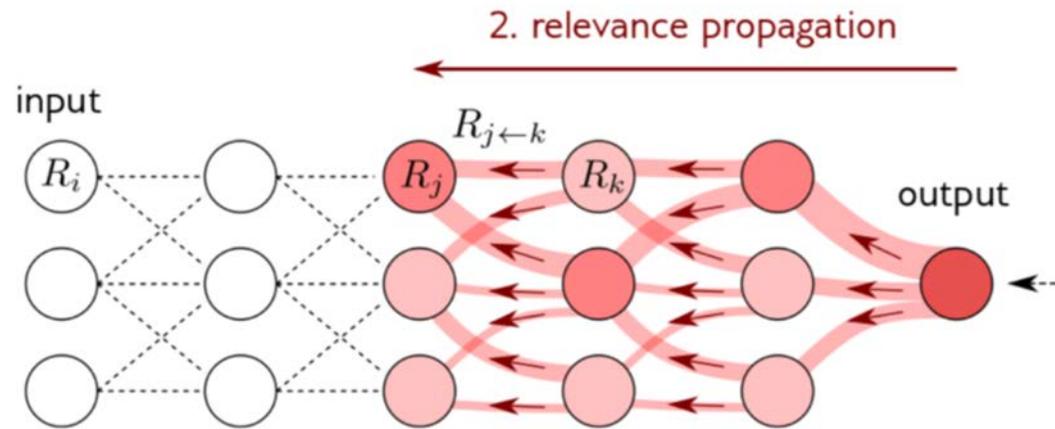
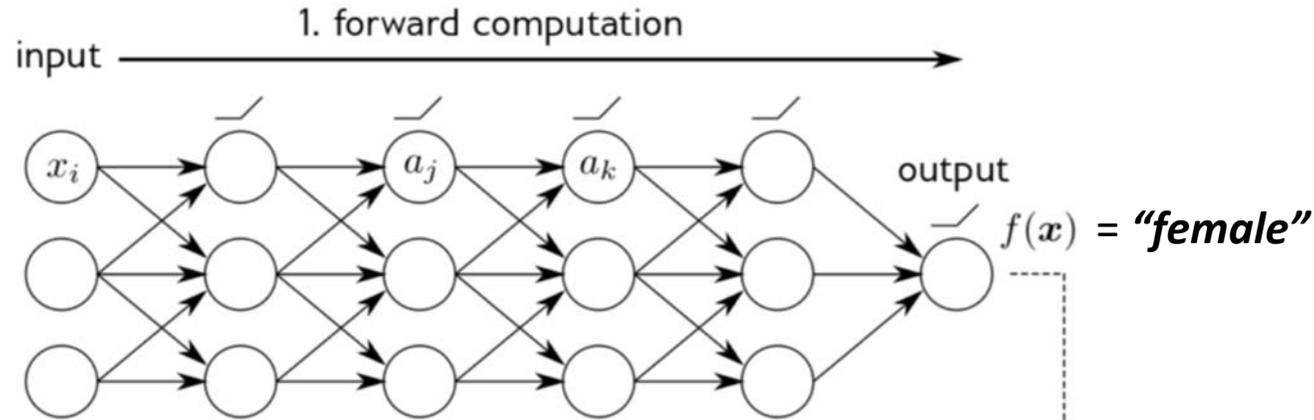
„Prototypical Network“

- Klassen werden durch sog. Prototypen repräsentiert
- Charakteristischer Merkmalsraum wird gelernt

Erste Ergebnisse

- Training mit 6 Gewebeklassen
- 3 zusätzliche (neue) Gewebeklassen bei der Anwendung

Beispiel Layerwise Relevance Propagation



LRP: Decomposition

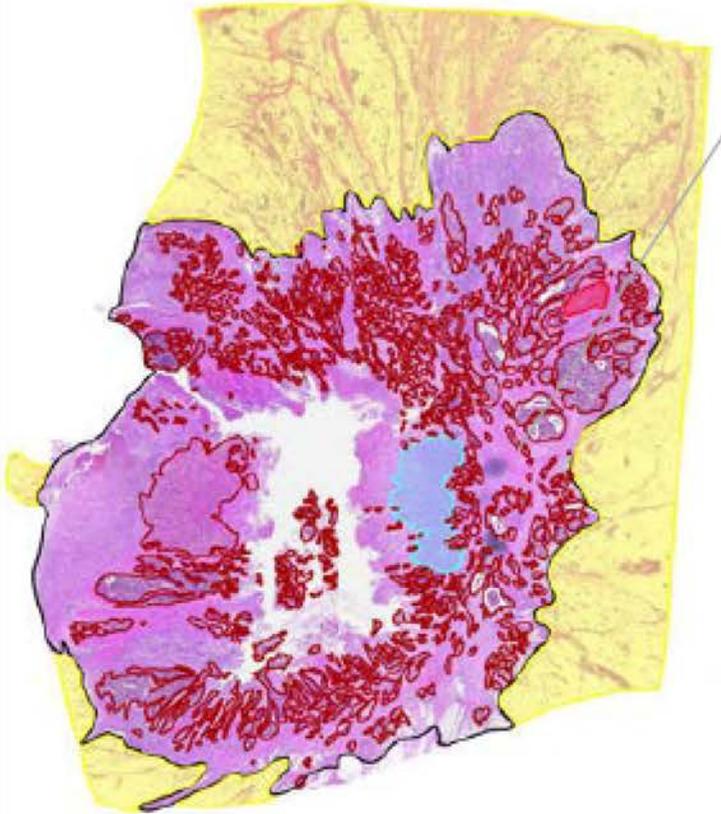
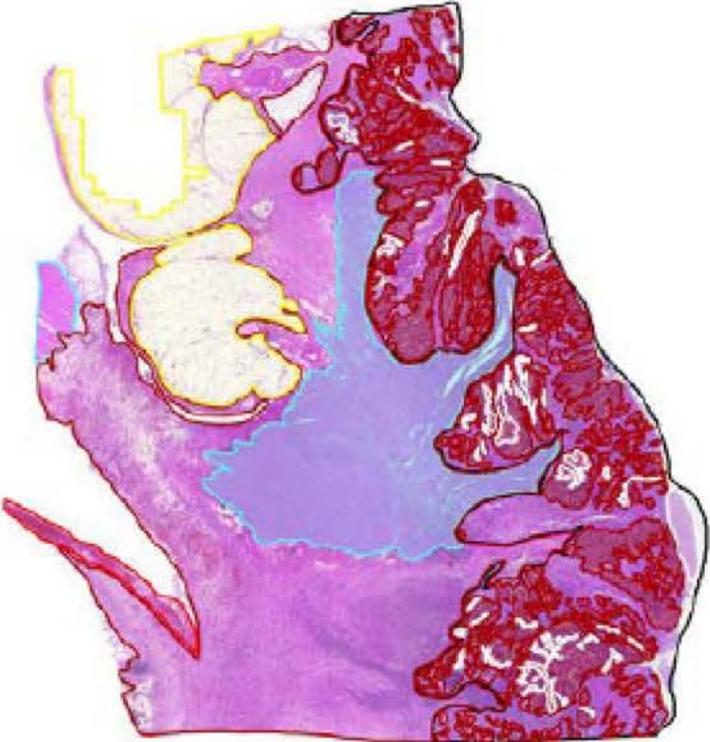
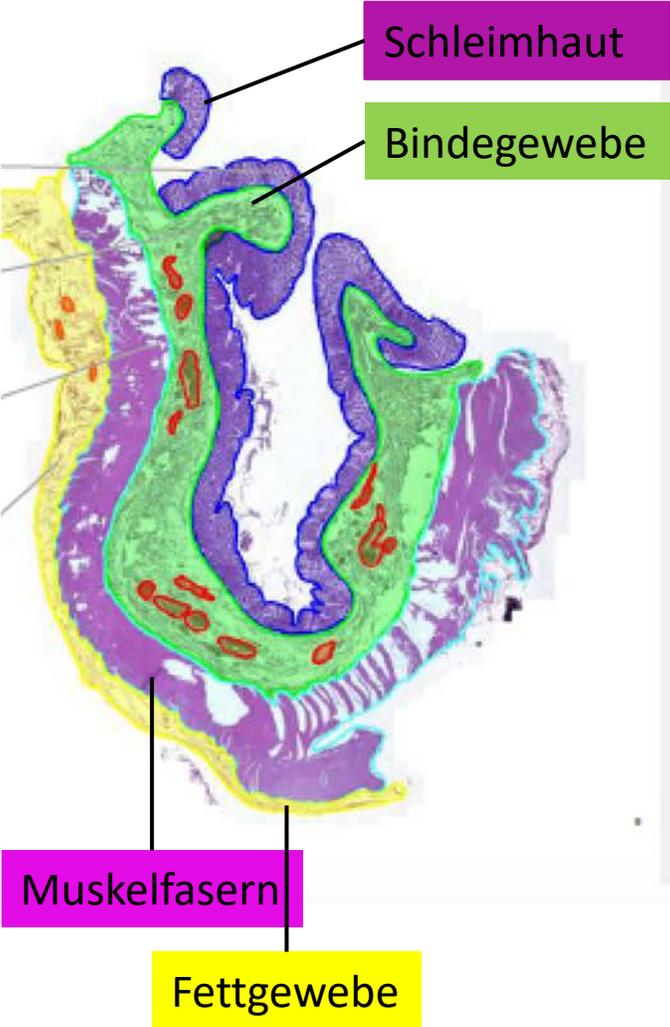
$$\sum_i R_i = f(x)$$

(how much does each pixel contribute to prediction)

$$R_j = \sum_k \frac{x_j w_{jk}}{\sum_j x_j w_{jk} + \epsilon} R_k$$

Samek, Binder, Montavon, Lapuschkin, Müller. (2017). Evaluating the visualization of what a deep neural network has learned. IEEE Trans. Neural networks and learning systems, 28(11), 2660-2673.

Beispiel Digitale Pathologie: Räumliche Relationen



Tumor

Beispiel Inductive Logic Programming

Datei Bearbeiten Einstellungen Hilfe



Details Erklärungen

Dieser Scan zeigt die Diagnose pT3.

Gründe:

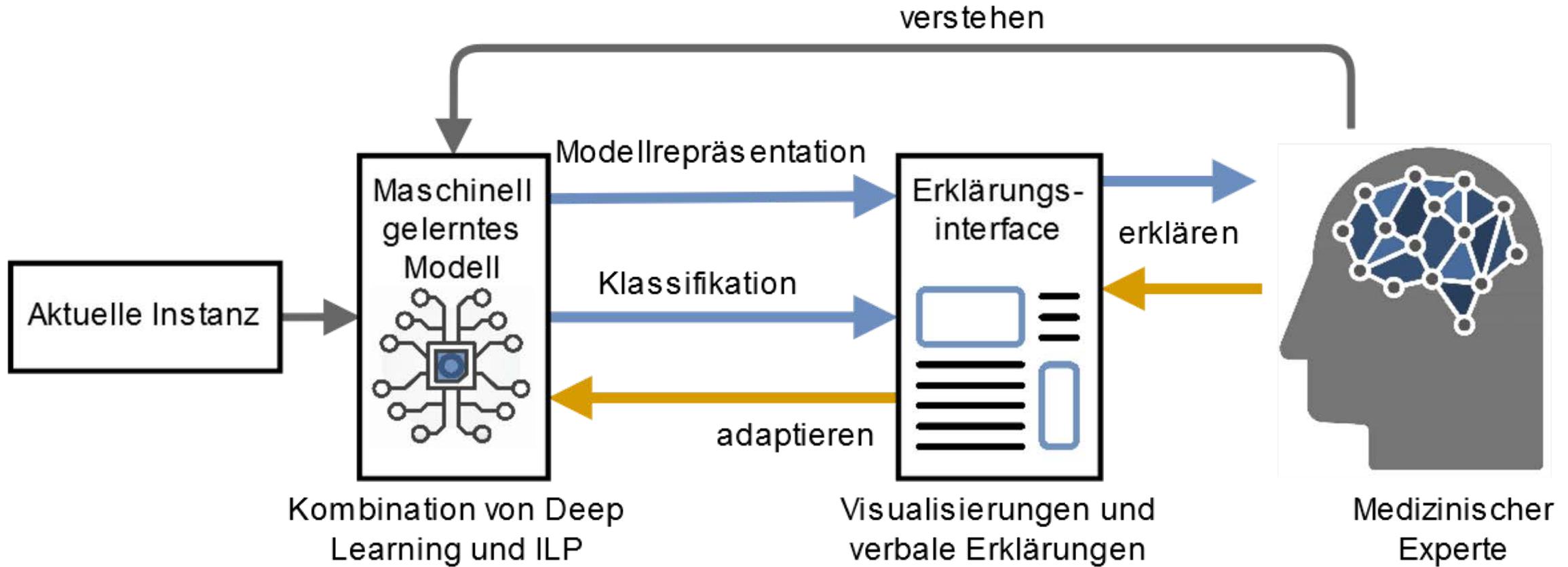
- Die größte Tumorausdehnung beträgt mehr als 5 cm.
- Der Tumor hat das Fettgewebe erreicht.
- Die Klassifikation ist sehr zuverlässig.

Interaktiv Erklärungen erweitern und korrigieren

Zeige Annotationen Zeige Kachel-Grid Zeige Koordinaten Suche Koordinate Wende an: LRP Wende an: CLRP Wende an: LIME Exportieren

Kacheln mit kontrastiven Beispielen	Ähnlichkeit: 87,2 %	Kacheln mit derselben Klassifizierung	Ähnlichkeit: 97,6 %
 1/18007201	Fläche "Tumor": 16.132 mm ² Koordinaten: (0,0),(227,227) Diagnose: pT2 Unsicherheit in Klassifizierung:	 1/2372	Fläche "Tumor": 26.457 mm ² Koordinaten: (0,454),(227,681) Diagnose: pT3 Unsicherheit in Klassifizierung:

Kooperatives Lernen durch wechselseitige Erklärungen



Herausforderungen

- Kombination von Deep Learning und ILP
- Fusion visueller und sprachlicher Erklärungen in ein System
- Korrigierbarkeit von Klassifikationsentscheidungen durch Experten
- Nutzung von Expertenrückmeldung bei der Adaptation von gelernten Modellen (inkrementelles, kooperatives Lernen)

- Übertragbarkeit auf andere Szenarien (Mammographie, Endoskopie,)
- Akzeptanz von XAI in der klinischen klinischen Praxis
- Spätere Zulassung als Medizinprodukt(e)

26.03.2019

Diagnose per Künstlicher Intelligenz soll transparenter

werden

