



inovex

KI-GESTÜTZTE BILDKLASSIFIKATION ZUR KREBSDIAGNOSE

Stefanie Marx





Stefanie Marx

Data & Machine Learning
Engineer @inovex

stefanie.marx@inovex.de

0174 7685550

www.inovex.de

Projekt entstanden in Kooperation mit:



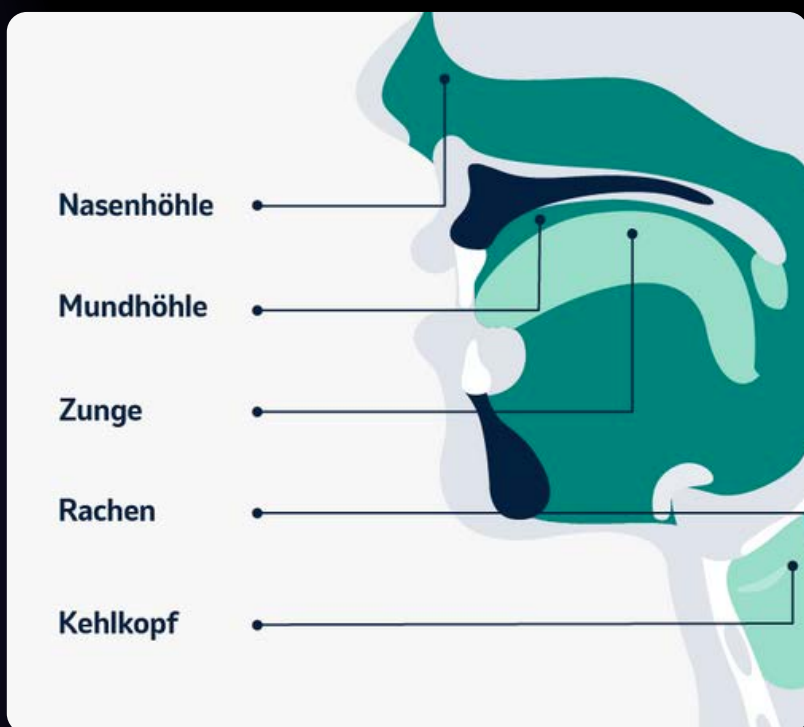
Dr. med Nina Wenda

Geschäftsführende Oberärztin,
Koordinatorin des Kopf-Hals-
Tumorzentrums
Helios-Klinik Wiesbaden

Agenda

- 01 Kopf-Hals-Krebs
- 02 Konfokales Laserendomikroskop
- 03 Bild-Datensatz
- 04 Data Preprocessing
- 05 Modelle
- 06 Ergebnisse & Ausblick

Kopf-Hals Krebs



890.000

Neue Fälle jährlich,
weltweit

7.

Häufigster Tumor
weltweit

1/3

Diagnosen in frühem
Tumorstadium

Folgen einer späten Diagnose

Geringere Heilungschance und umfassenderer
chirurgischer Eingriff
→ Risiko von nachhaltigen Einschränkungen bei
wichtigen Funktionen wie Sprechen oder Schlucken

Goldstandard der Diagnose

Biopsie und histopathologische Untersuchung
→ invasiv, erfordert mehrere Operationen

Konfokale Laserendomikroskopie (CLE)



Echtzeitübertragung von
Gewebe auf Zellebene

Intraoperative, **non-invasive**
Bestimmung der Tumorgrenze

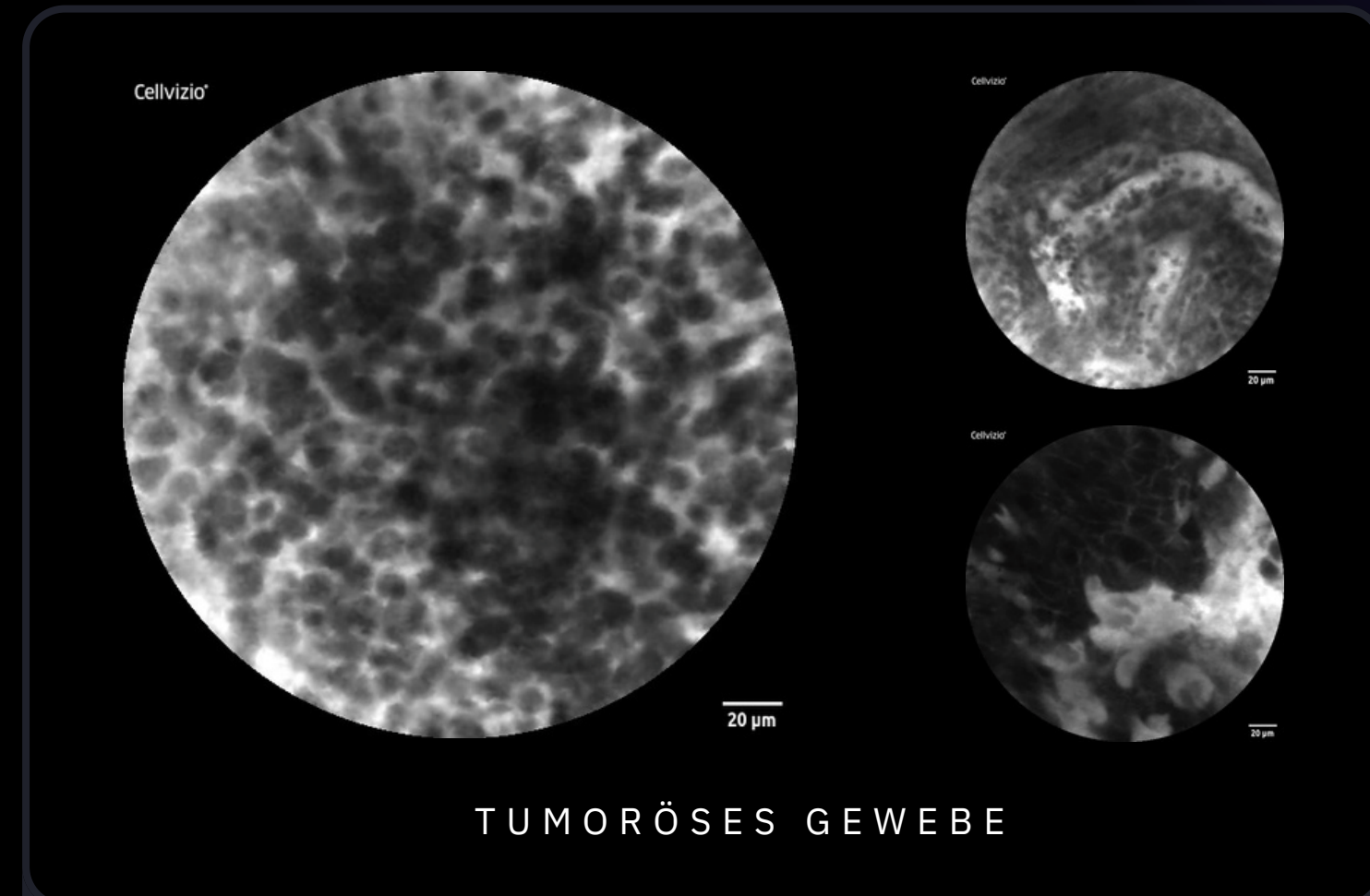
Ursprünglich entwickelt für die
Gastroenterologie

Seit ein paar Jahren auch in der
Kopf-Hals Onkologie eingesetzt

Interpretation von CLE-Bildern

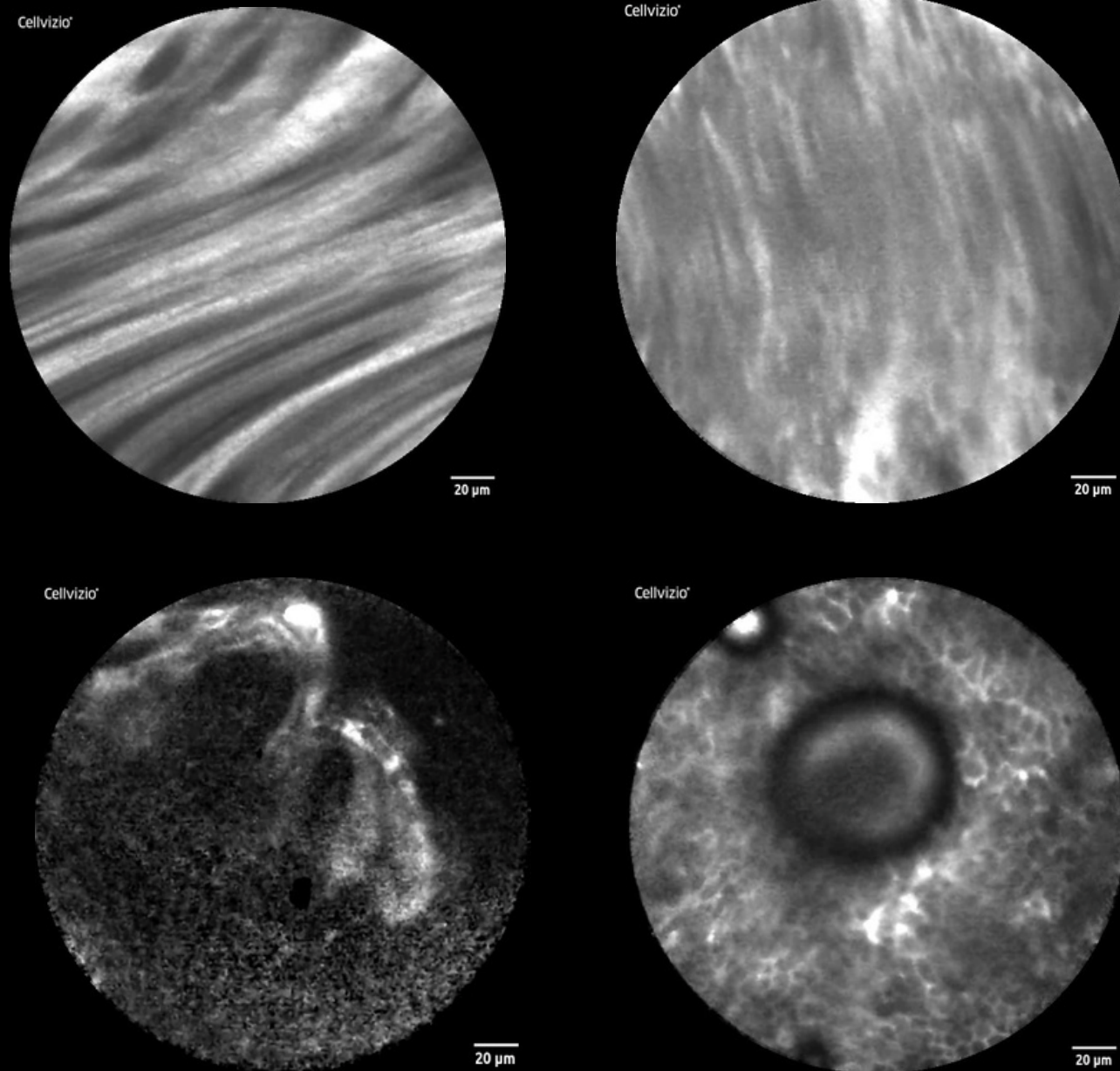


- Homogene, wabenartige Zellarchitektur
- ‘Kopfsteinpflaster’
- Klar definierte Zellränder, interzelluläre Räume
- Konsistente Helligkeit
- Kontrastmittel flutet gleichmäßig an
- Kapillarschlingen



- Irreguläre, ‘chaotische’ Zellstruktur
- Zellen variieren in Form und Größe
- Bereiche mit unterschiedlichen Graustufen
- Kontrastmittel tritt aus
- Vergößerte Kapillare und atypische Blutgefäße

Interpretation von CLE-Bildern



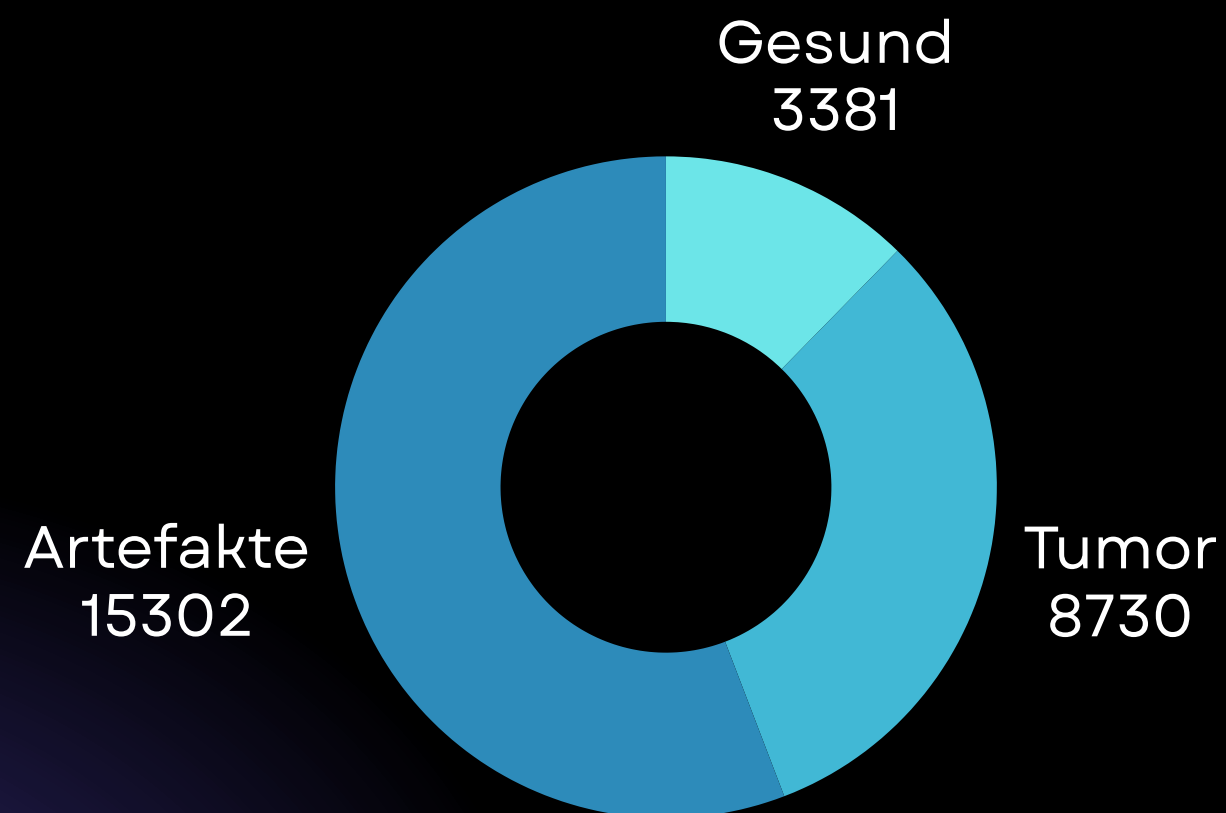
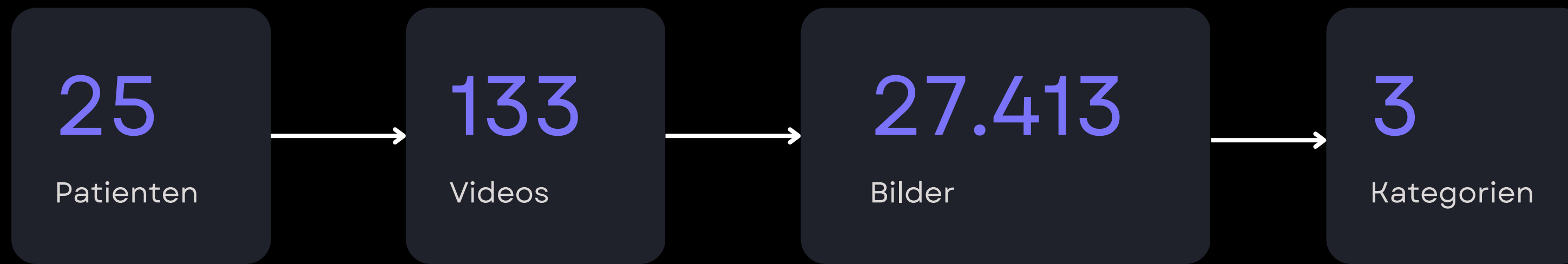
- Bildqualität kann durch Bewegungsartefakte, unebenes Gewebe oder einen schlechten Sondenkontakt beeinträchtigt werden
- Geschickte Handhabung der Sonde und zuverlässige Diagnose erfordern viel Training und **jahrelange Erfahrung**

→ Das limitiert die breite Einsetzbarkeit von CLE

Ziel

- Automatische Klassifikation von gesunden und tumorösen Bildern mittels Deep Learning
- Unterstützung präziserer und effizienterer intraoperativer Entscheidungen

Data Collection



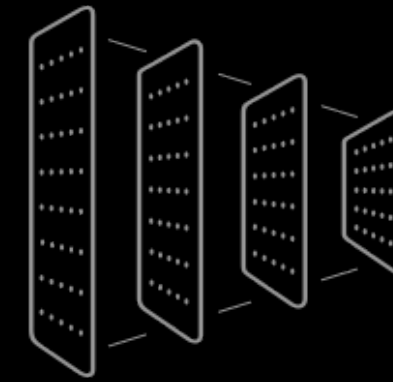
- Patienten mit bestätigter Tumordiagnose im Kopf-Hals-Bereich
- Alle Bilder wurden annotiert in 3 Kategorien von CLE-Expertin → Ground Truth
- Bilder haben eine Auflösung von 1 μm & meist eine Größe von 512x512 Pixeln

Klassifikationsansätze



Machine Learning

- Nutzt manuelle **Feature Extraction**: relevante Bildmerkmale (z. B. Textur, Kontrast, Struktur) werden durch vordefinierte Algorithmen berechnet
- Anschließend Klassifikation mit Modellen wie SVM oder Random Forest
- Gut bei kleinen, homogenen Datensätzen, aber anfällig für Bildvariabilität und begrenzt durch Expertenwissen



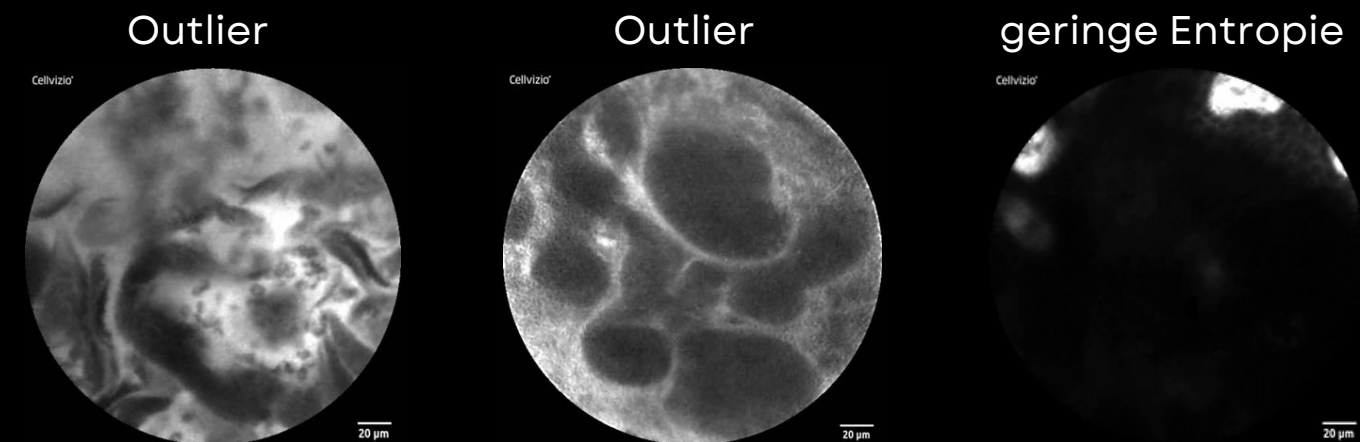
Deep Learning

- **Convolutional Neural Networks (CNNs)** erkennen hierarchische Strukturen wie Kanten, Zellen und Gewebemuster
- Lernt Merkmale automatisch direkt aus den Rohbildern, kein Vorwissen nötig
- Robuster gegenüber Bildvariationen, benötigt aber große Datenmengen
- Übertrifft klassische ML-Ansätze auf komplexen, vielfältigen Datensätzen

Data Preprocessing

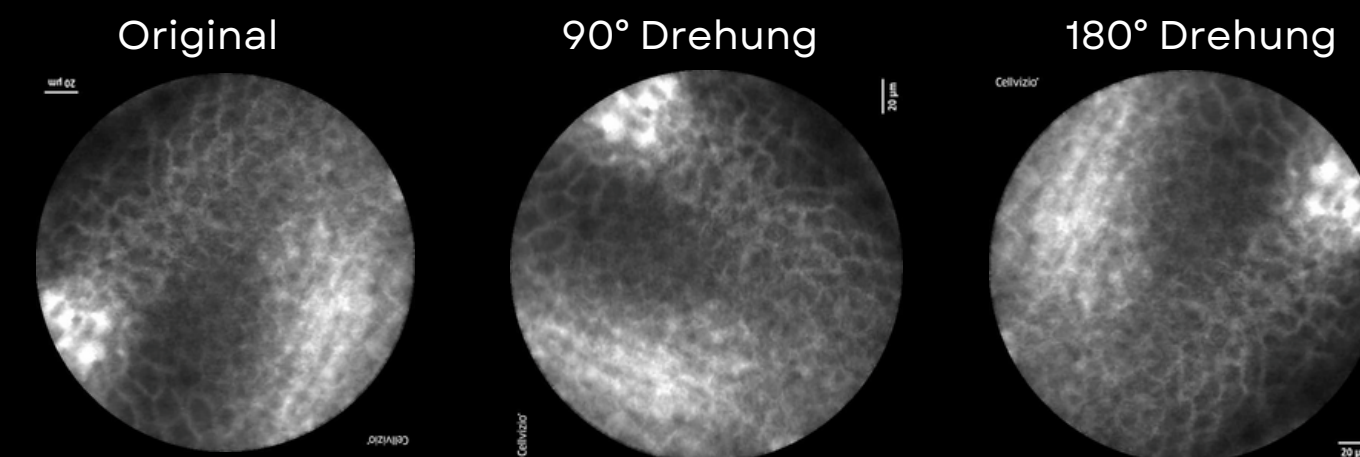
01 Data Cleaning

Bilder mit ungewöhnlichen Strukturen & niedriger Entropie wurden nachträglich aussortiert



02 Data Augmentation

Gesunde Bilder wurden rotiert, um ausgewogene Klassen zu erreichen
→ je 2 zusätzliche gesunde Bilder



03 Normalization

Einheitliche Größe & Pixelwerte skalieren, um konsistente Intensitäten zu gewährleisten

Finale Datenverteilung

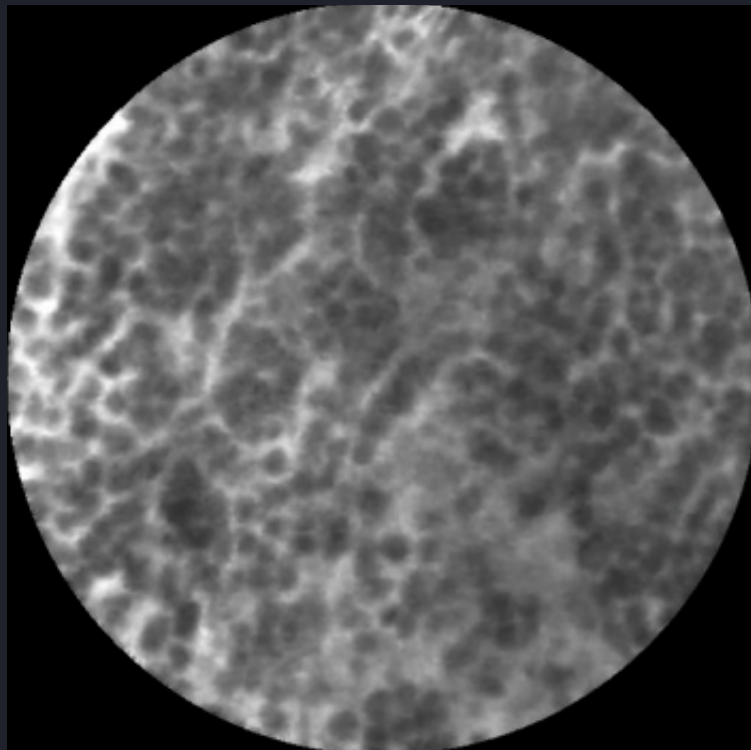
8.142 gesunde Bilder & 7.997 Tumor Bilder

Image Transformation Approaches

Herausforderung: CLE hat kreisförmiges Sichtfeld → CNNs erfordern rechteckige Eingabebilder

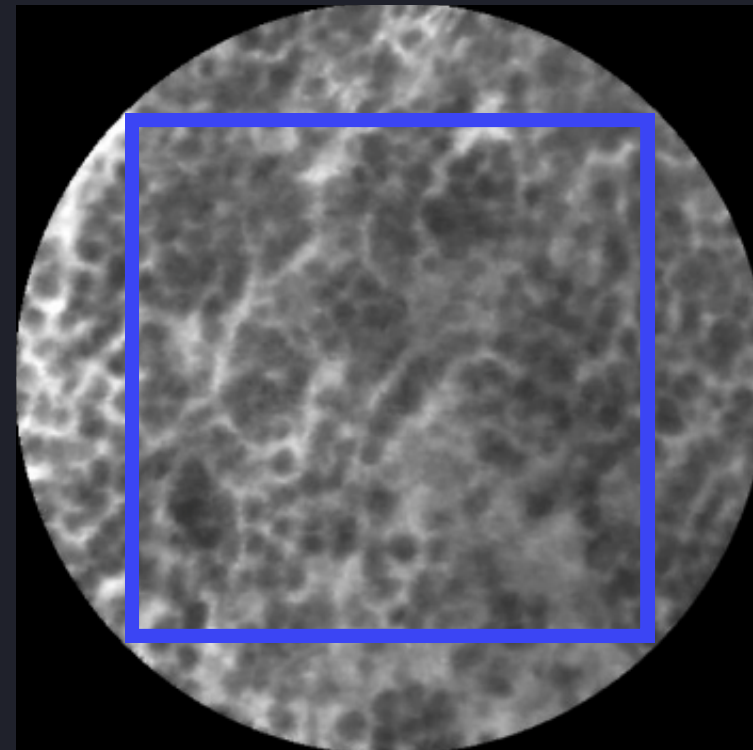
01 Full Image

komplettes Bild inklusive Ränder



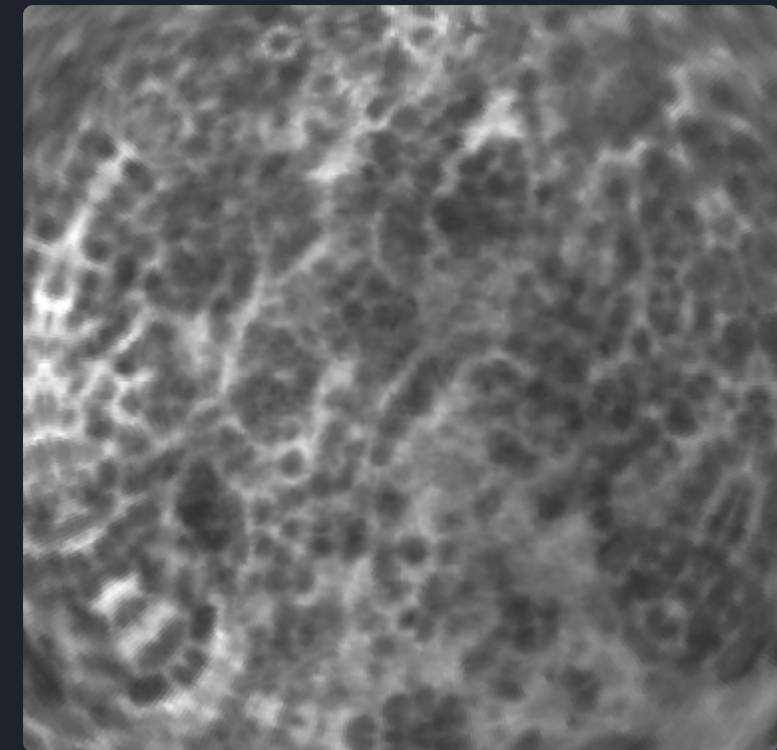
02 Cropped Square

Sichtfeld in größtmöglichen rechteckigen Bereich zuschneiden



03 Circular Extrapolation

Kreisförmiger Bildinhalt wird nach außen gespiegelt erweitert, um schwarzen Rand zu entfernen



Model Selection

01 Trained from scratch

- LeNet-5, AlexNet
- Lernen CLE-spezifische Merkmale ohne Vorwissen
- Dienen als Baseline

02 Pre-trained

- ResNet, InceptionV3, EfficientNet
- Vortrainiert auf ImageNet → bereits generische Bildmerkmale gelernt
- Feinabstimmung (Fine-Tuning) auf CLE-Daten

Transfer Learning:

- Nutzung vorhandenen Wissens über Bildstrukturen (z. B. Kanten, Texturen)
- Schnellere Konvergenz und höhere Genauigkeit
- Letzte Klassifizierungsschicht wird ersetzt um binäre Klassifikation zu erhalten

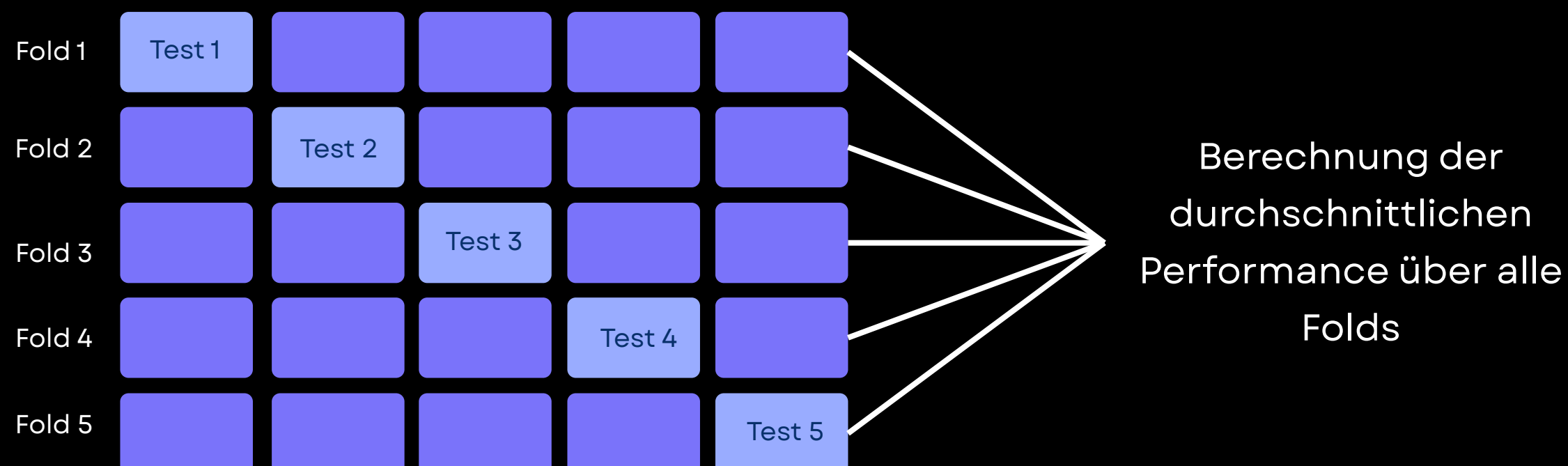
Training

Group-Stratified Train/Test Split

- Sorgt für ausgeglichene Klassenverteilung (~50 % gesund / ~50 % Tumor)
- Kein Patient erscheint in beiden Sets
- Realistische klinische Bedingung: Modell sieht neue Patienten erst im Test

Cross-Validation

- Mehrfache Aufteilung des Datensatzes, sodass jedes Teilset einmal als Testset dient
- 5-Fold Cross-Validation für Modellvergleich & schnellere Optimierung
- Leave-One-Patient-Out (LOPO) Validation zur Untersuchung patientenspezifischer Unterschiede



Training

Training Configuration

Beim Training werden Hyperparameter festgelegt, die das Lernverhalten des Modells steuern, unter anderem:

- **Learning Rate:** Gibt an, wie groß die Schritte beim Lernen sind (zu hoch = instabil, zu niedrig = langsam)
- **Epochen:** Wie oft das Modell alle Trainingsdaten sieht, also die Trainingsdauer
- **Dropout/Regularisierung:** Verhindert, dass das Modell sich zu stark an Trainingsdaten anpasst (Overfitting)

Strategie: Zunächst experimentelle Abstimmung der Hyperparameter, anschließend gezieltes Hyperparameter-Tuning des besten Modells zur weiteren Optimierung

Ergebnisse

Model	Accuracy	Sensitivity	Specificity
LeNet	65.8 %	65.7 %	66.8 %
AlexNet	75.6 %	82.7 %	70.2 %
Inception	80.5 %	88.9 %	73.8 %
ResNet	79.9 %	88.4 %	73.5 %
EfficientNet	84.3 %	88.7 %	81.1 %

Accuracy

→ Wie viele Bilder **insgesamt** richtig erkannt wurden

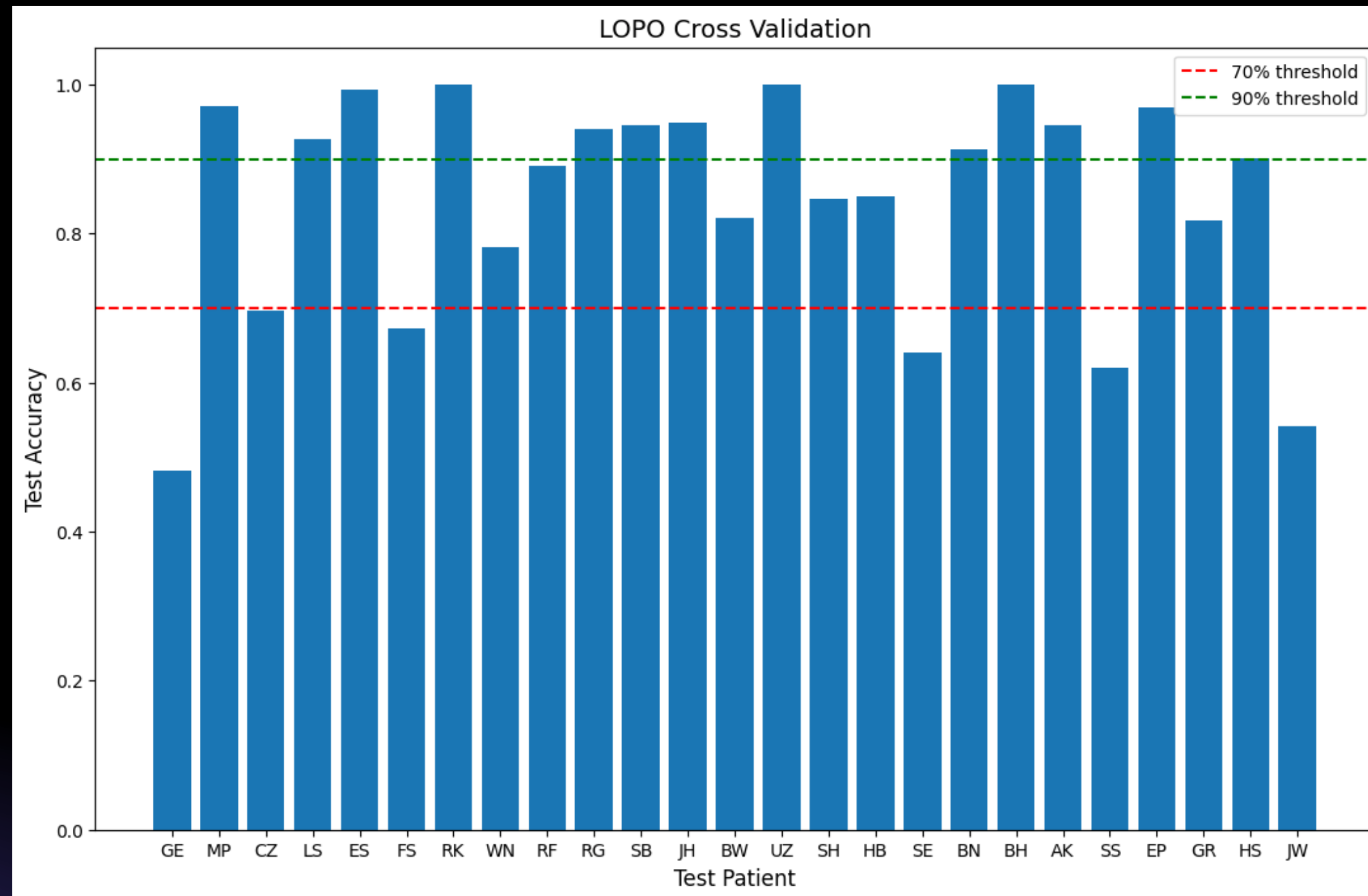
Sensitivity

→ Wie viele **Tumor** Bilder richtig erkannt wurden

Specificity

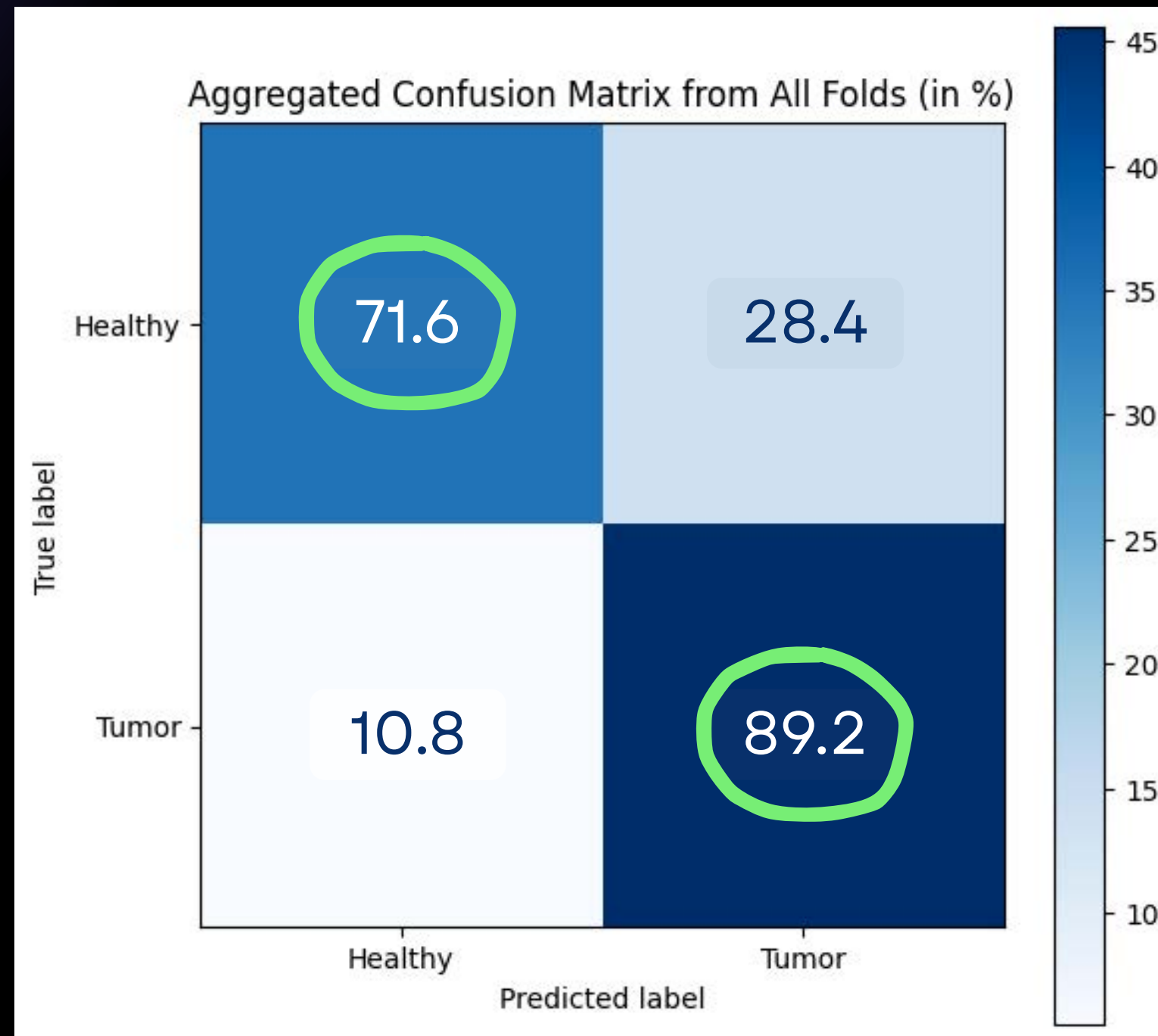
→ Wie viele **gesunde** Bilder richtig erkannt wurden

Leave-One-Patient-Out Validation



- Starke Varianz der Modellleistung zwischen den Patienten
- 13 Patienten mit Accuracy **> 90 %**, davon 3 mit **100 %**
- aber auch 5 Patienten **< 70 %**

Ergebnisse

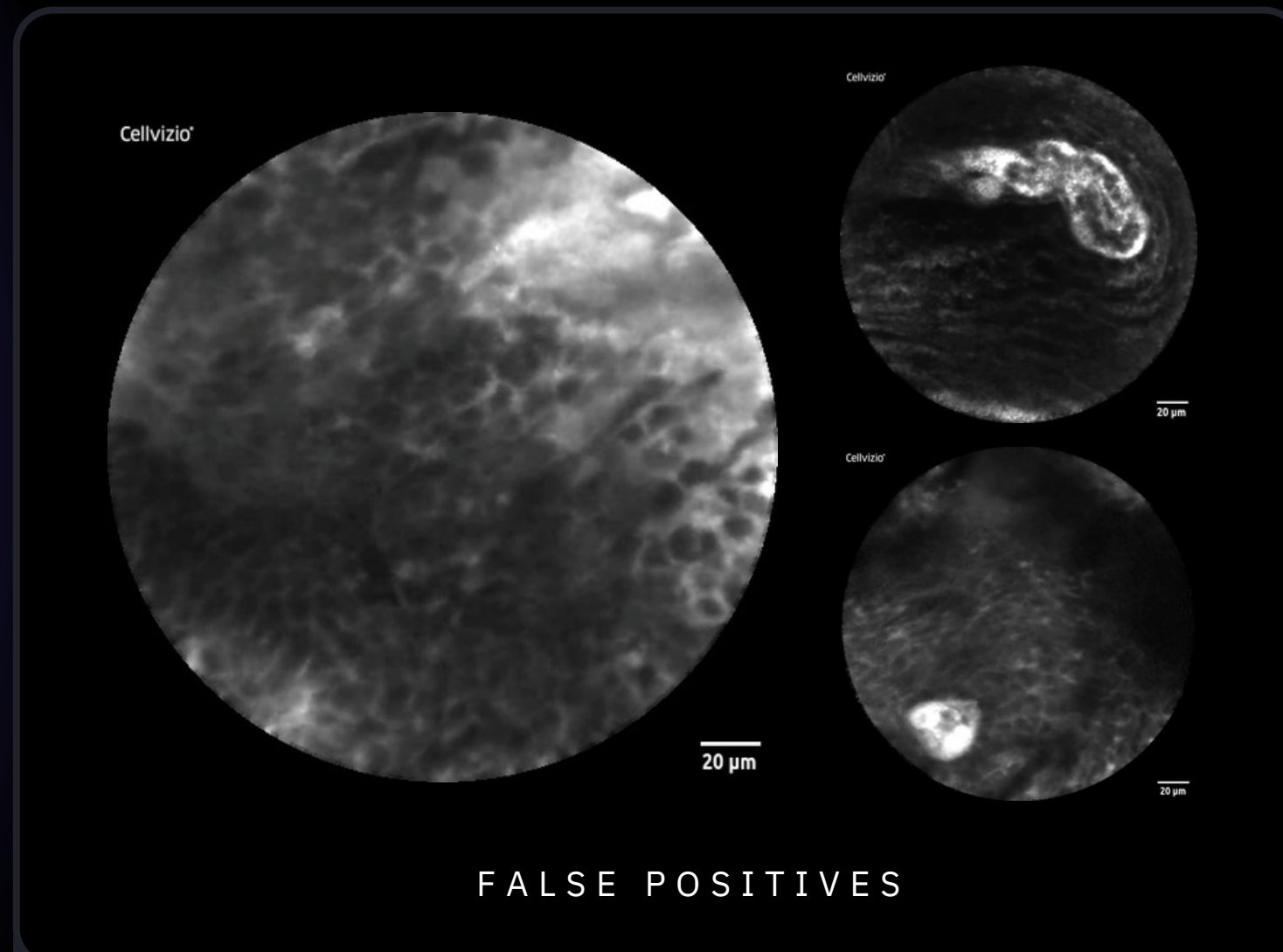


- Modell hat Schwierigkeiten, gesundes Gewebe korrekt zu klassifizieren
- Viele **False Positives** → gesundes Gewebe im Datensatz unzureichend oder atypisch repräsentiert

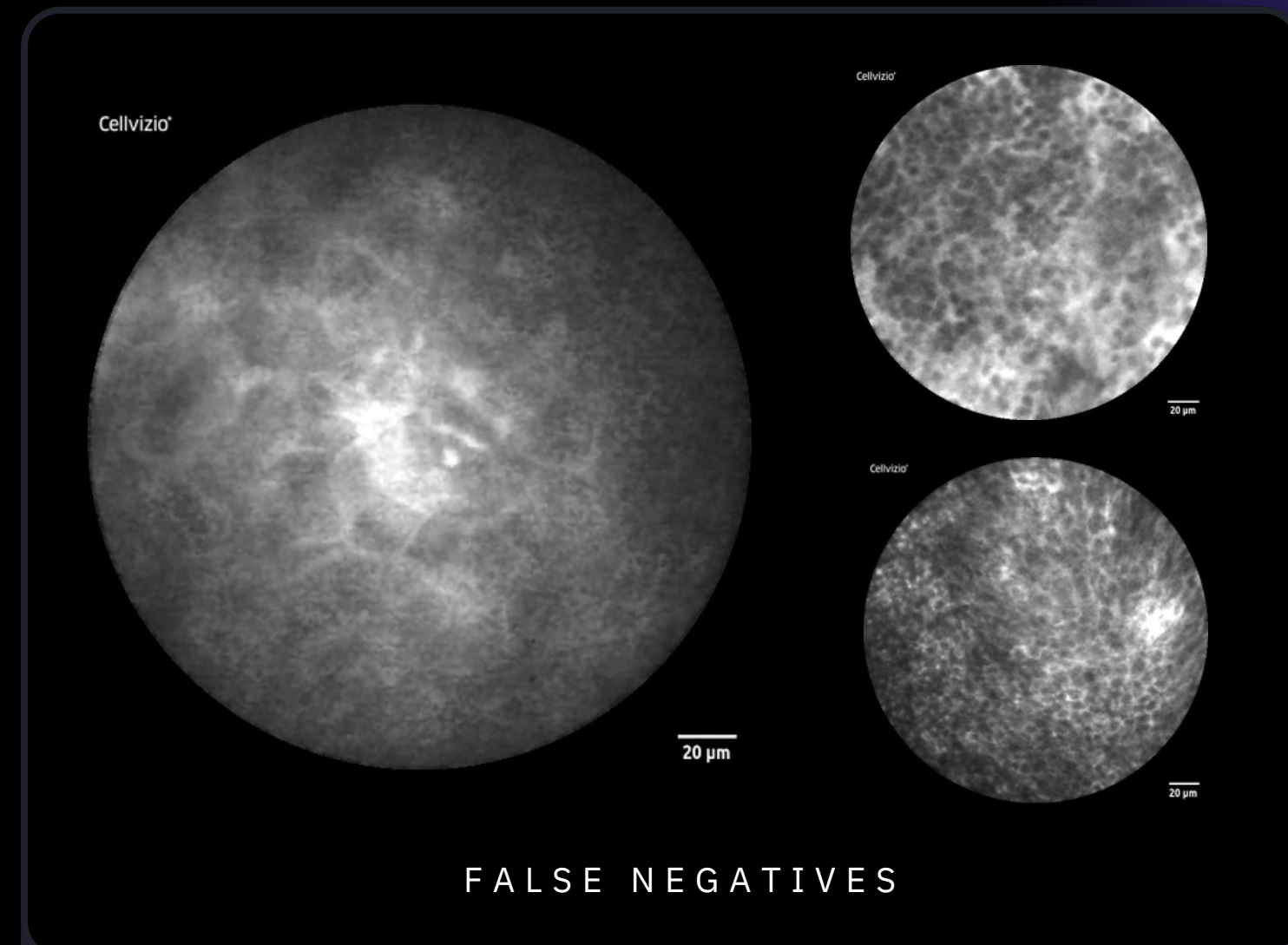
Zu beachten:

- Gesunde Bilder waren von Anfang an unterrepräsentiert, Augmentation nur künstlicher Ausgleich
- „Gesunde“ Proben stammen aus nicht-tumorösen Bereichen von Tumorpatienten, nicht von gesunden Personen

Fehlklassifikationen



- Gesundes Gewebe fälschlich als Tumor klassifiziert
- Tritt oft bei dunklen Bildern auf



- Tumoröses Gewebe fälschlich als gesund klassifiziert
- Tritt oft bei niedriger Bildqualität auf

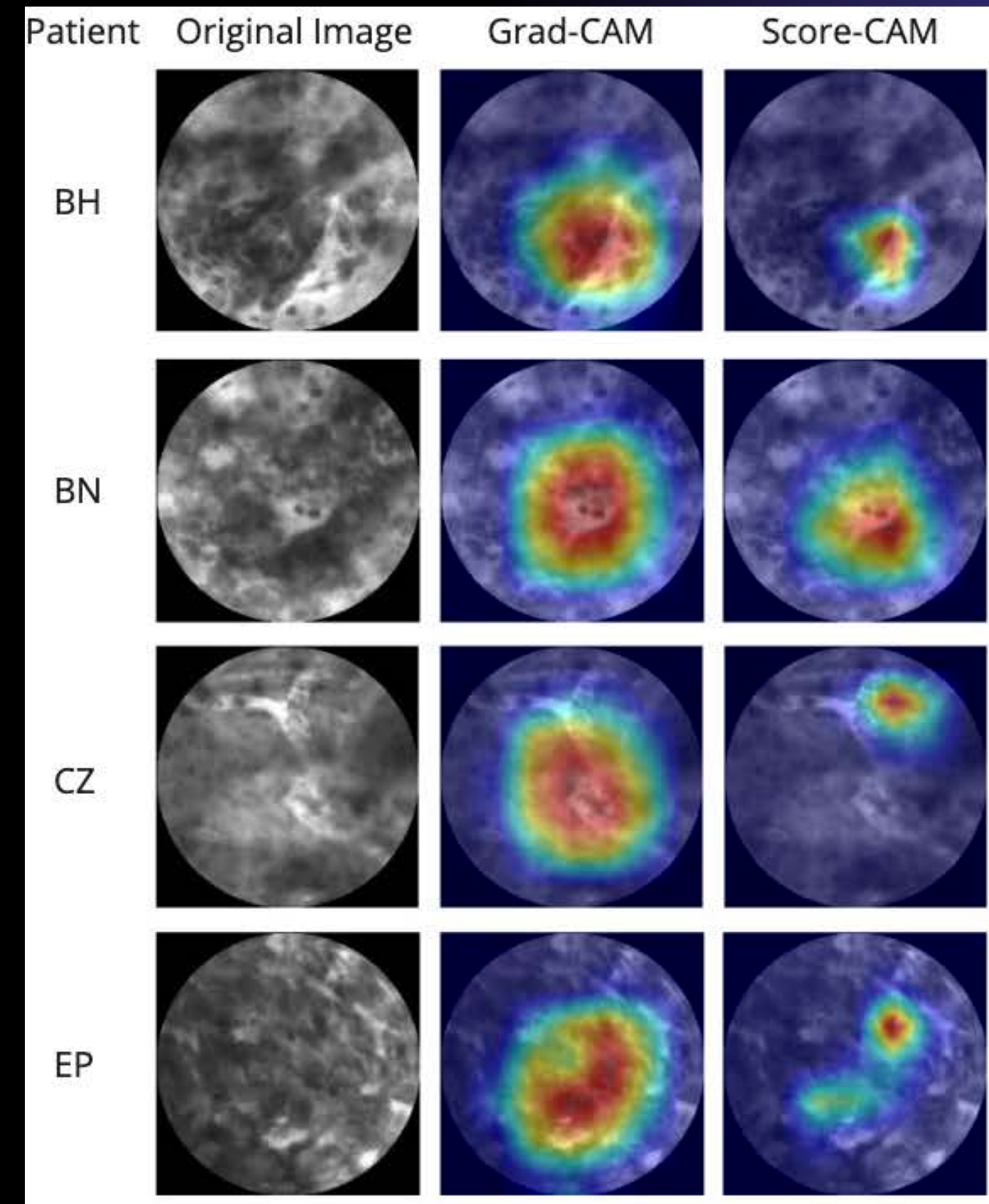
Class Activation Maps

Funktion

- Zeigen, welche Bildbereiche das Modell für seine Entscheidung nutzt
- Hervorgehobene Bereiche → besonders relevant für die Klassifikation

Erkenntnisse

- Modell fokussiert auf klinisch relevante Tumorregionen, jedoch werden oft ähnliche Regionen aktiviert
- Expertenvalidierung nötig



Future Work

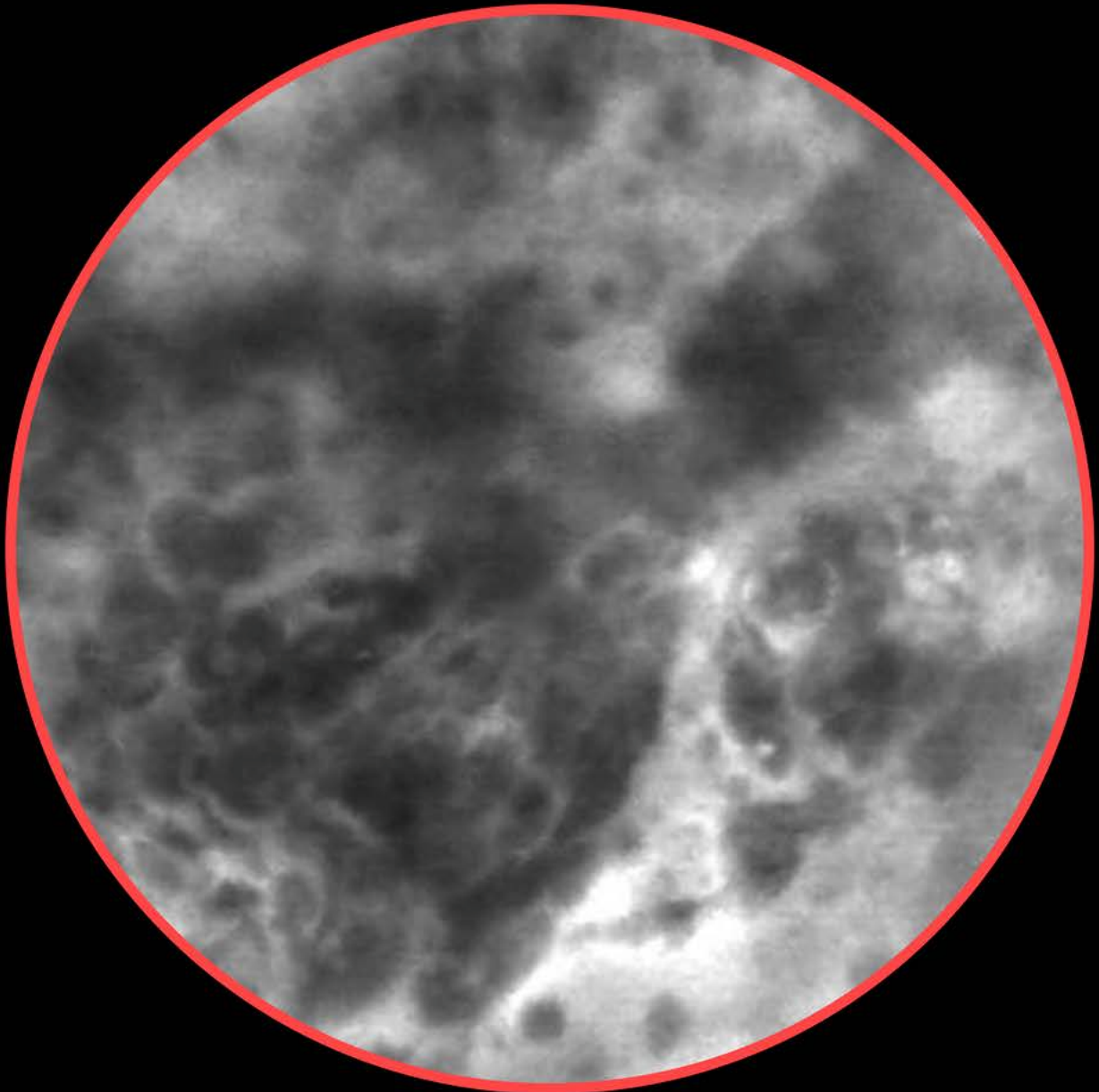
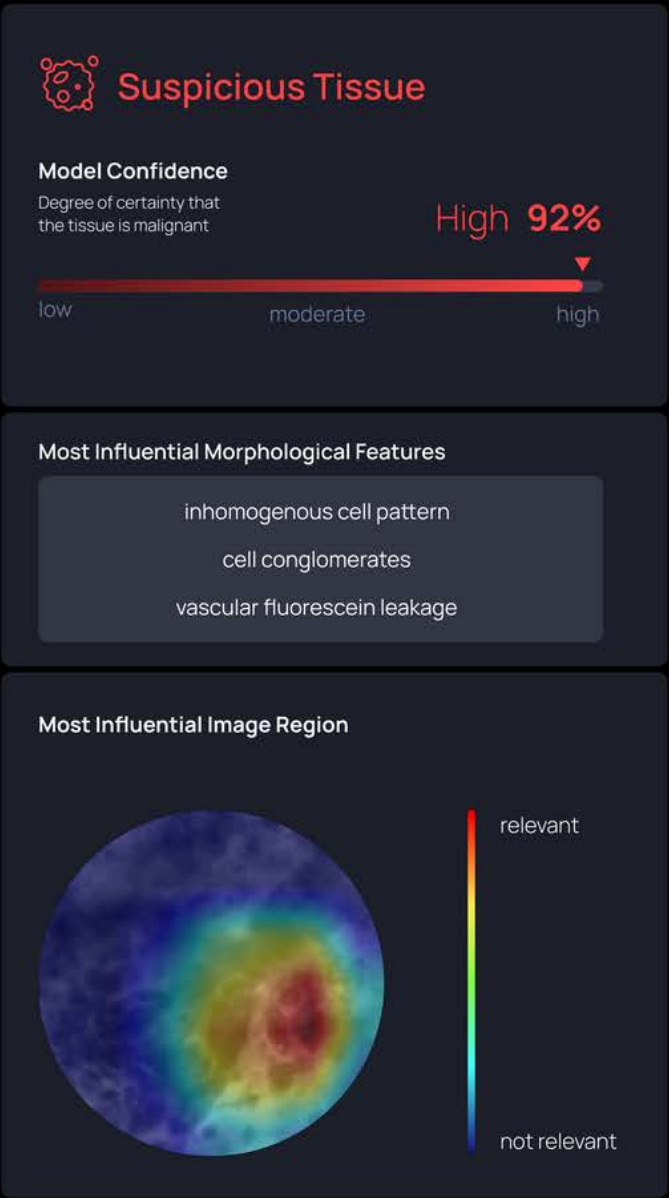
Potenzielle Weiterentwicklung

- Two-Stage CNN Pipeline:
 - Diagnostische vs. nicht-diagnostische Bilder klassifizieren
 - Gesunde vs. Tumor-Bilder klassifizieren
- Segmentierung der Tumorränder

Mensch-KI Kollaboration:

- Interpretierbarkeit der Modelle ist entscheidend für die klinische Akzeptanz
- ML-Modelle sollten nicht nur anhand ihrer Vorhersageleistung bewertet werden, sondern auch anhand ihrer Interaktion mit menschlichen Benutzern

User Interface



@Masterarbeit von
Alina Döring
Enabling AI Integration
in Surgery

Vielen Dank!



Stefanie Marx

Data & Machine Learning
Engineer

stefanie.marx@inovex.de

0174 7685550

www.inovex.de

