

Von Deep Learning zu
vertrauenswürdigen Entscheidungen:
Erklärbare Tumorerkennung im OP



Eine kurze Vorstellung



Stefanie Marx

Data & Machine Learning
Engineer @inovex

stefanie.marx@inovex.de
www.inovex.de



Alina Döring

Human-Centered AI @inovex

alina.doering@inovex.de
www.inovex.de

Es geht um



Dr. med Nina Wenda

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

Wie viel ist *genug*?

Nina



Onkologische
Sicherheit

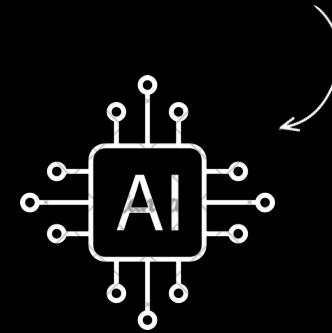
Lebensqualität von
Patient:innen

Wie viel ist *genug*?

Nina



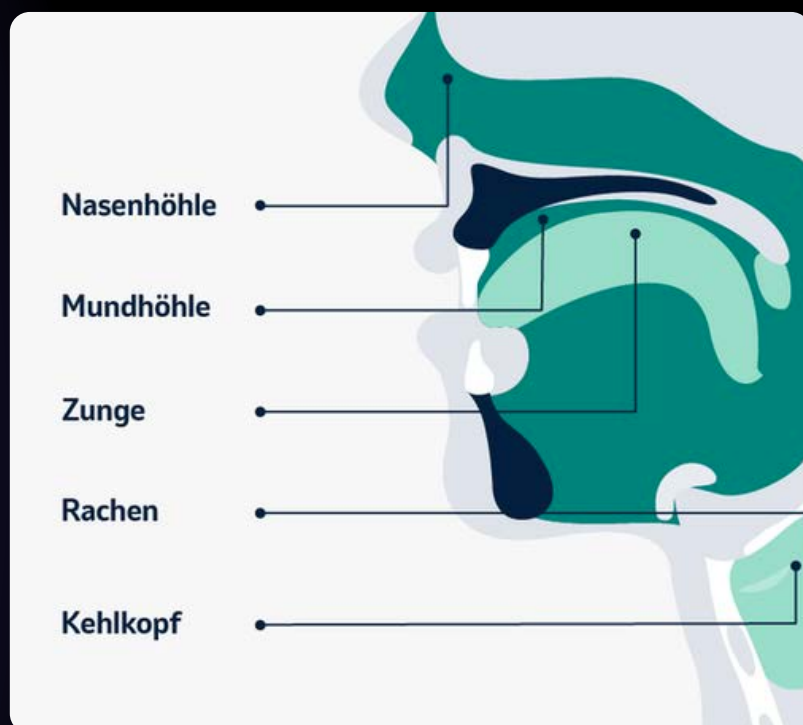
KI-Assistenz



Onkologische
Sicherheit

Lebensqualität von
Patient:innen

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)



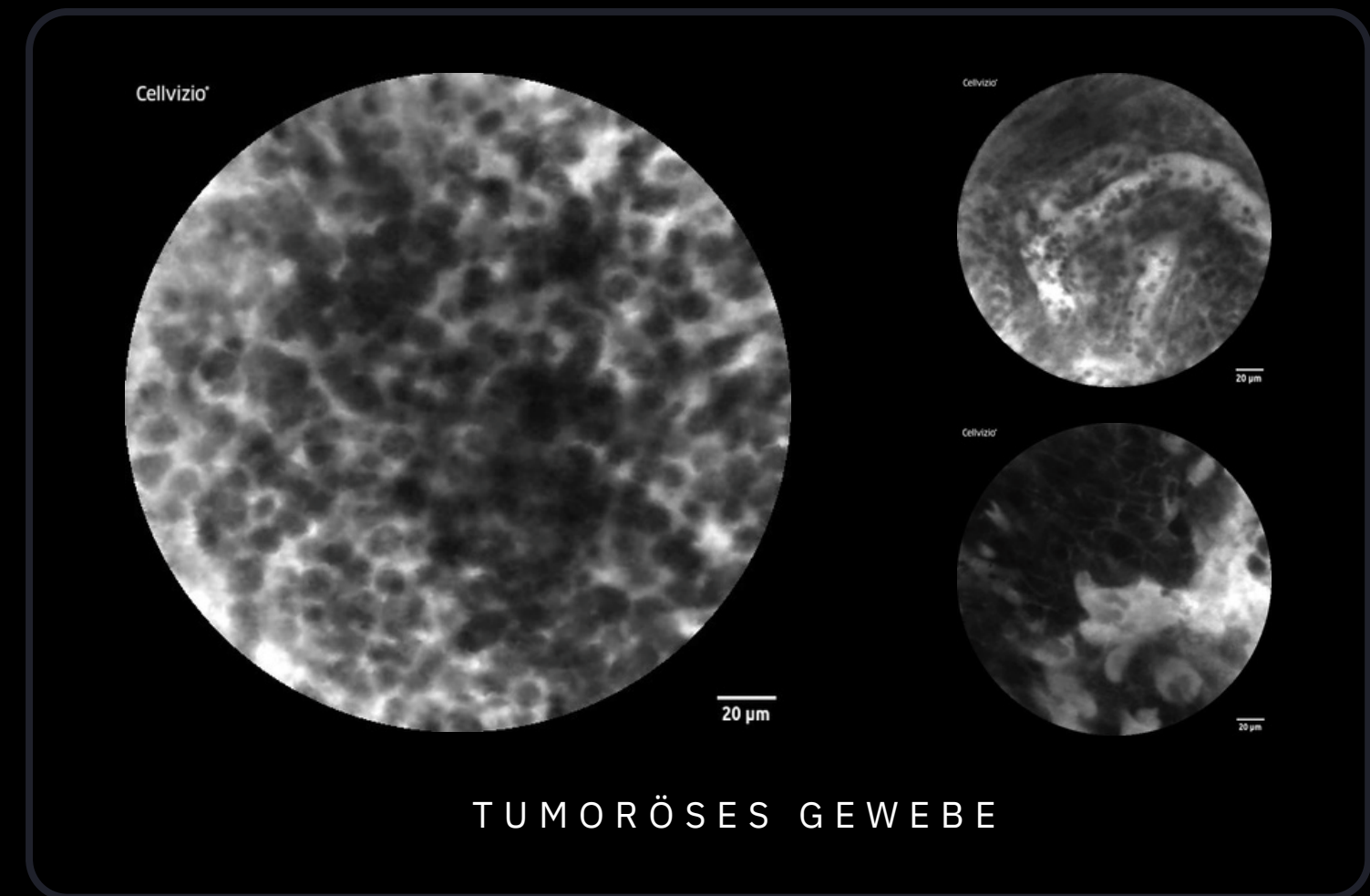
Echtzeitübertragung von
Gewebe auf Zellebene

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

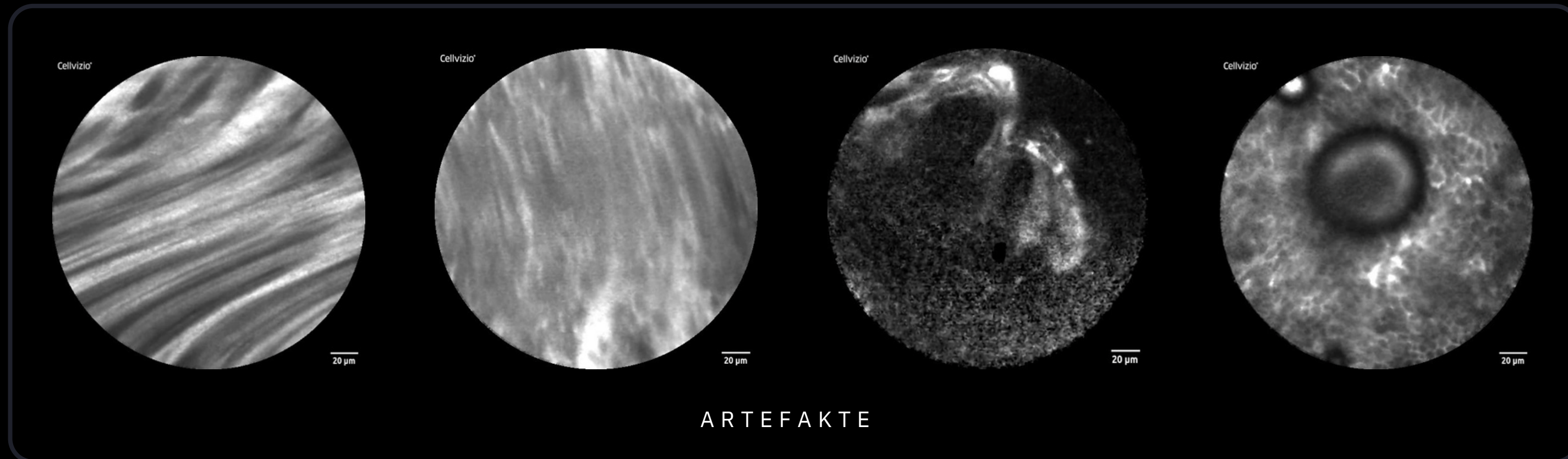
Ursprünglich entwickelt für die
Gastroenterologie

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

Interpretation von CLE-Bildern

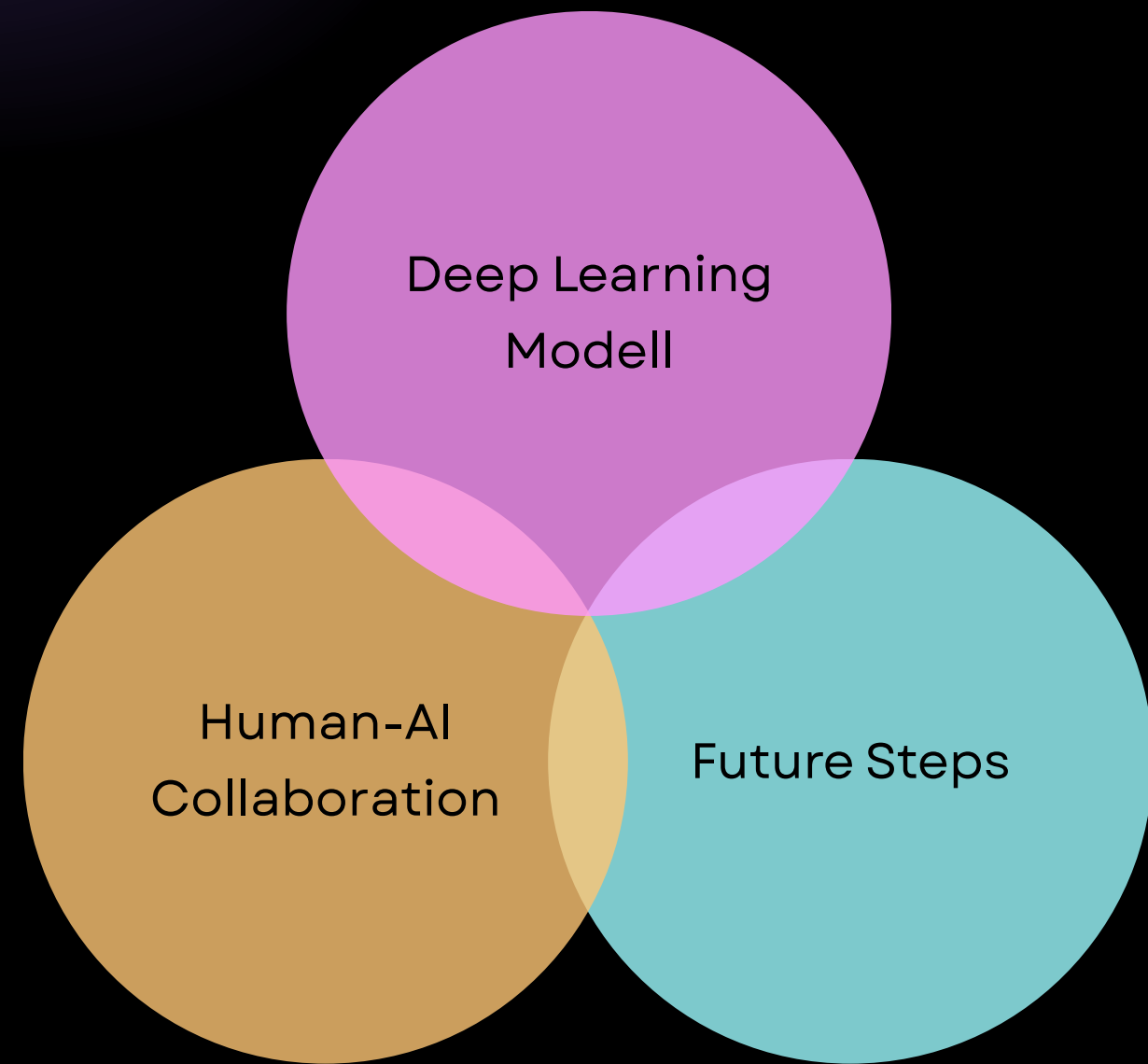


Interpretation von CLE-Bildern



- 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



1

Binäre Klassifikation mit Deep Learning (CNNs)

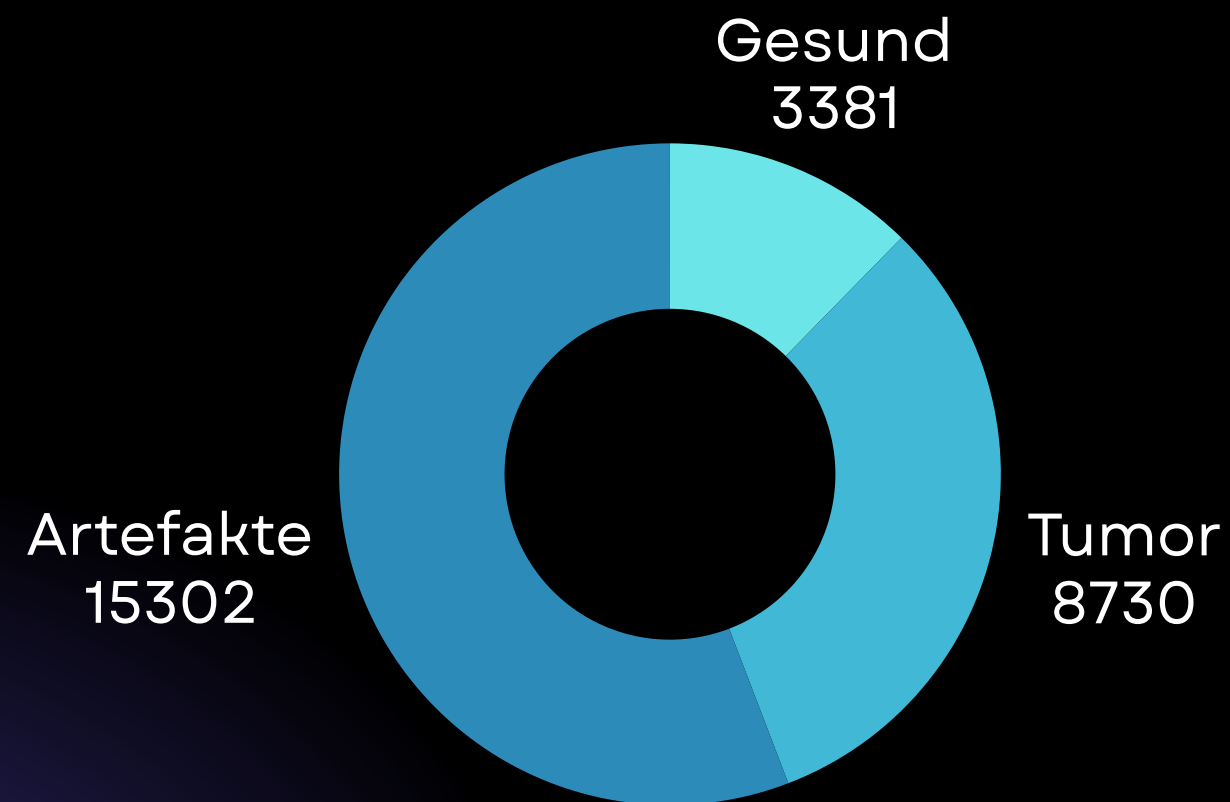
2

Human-AI Collaboration,
Erklärbare User Interfaces

3

Priorisierte Weiterentwicklung

Data Collection

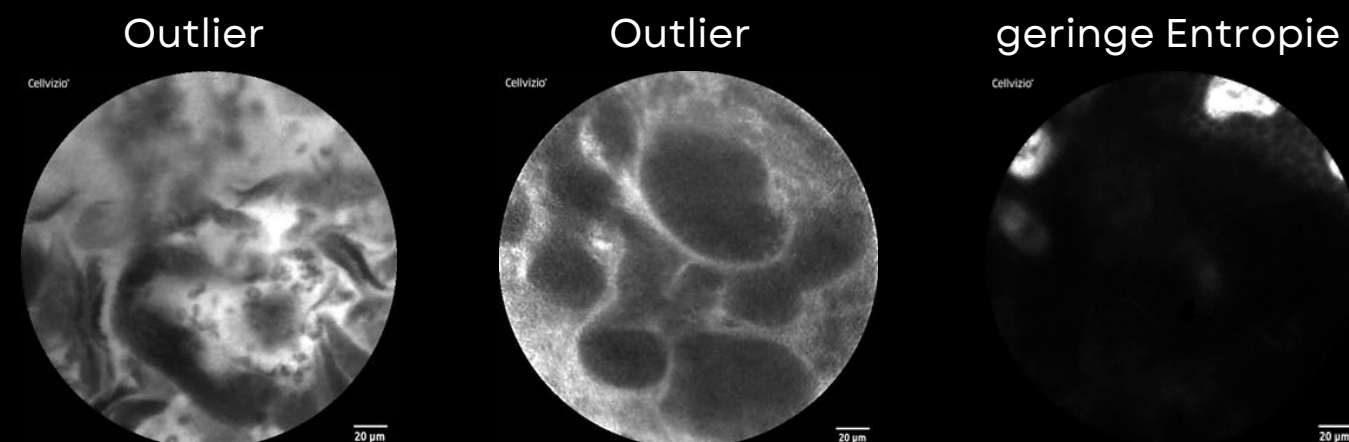


- Patient:innen mit bestätigter Tumordiagnose im Kopf-Hals-Bereich
- Alle Bilder wurden annotiert in 3 Kategorien von CLE-Expertin → Ground Truth

Data Preprocessing

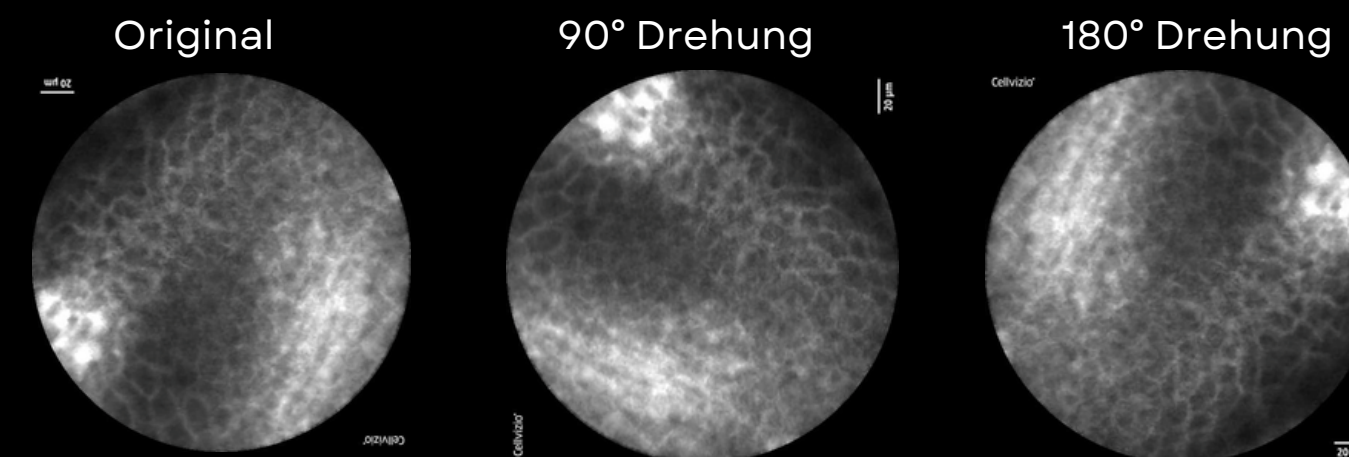
01 Data Cleaning

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

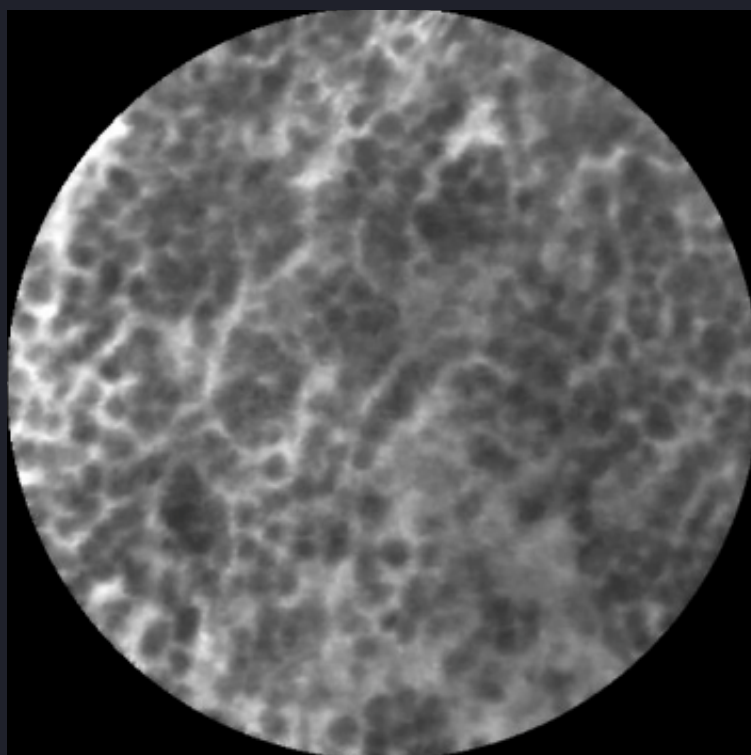
8.142 gesunde Bilder & 7.997 Tumor Bilder

Image Transformation Approaches

Herausforderung: CLE hat kreisförmiges Sichtfeld → CNNs erfordern rechteckige Eingabebilder

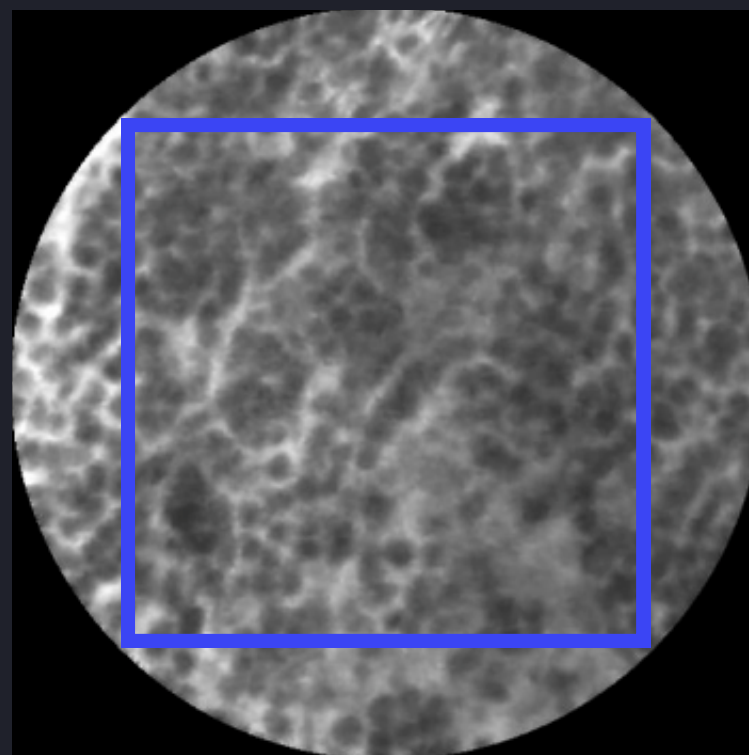
01 Full Image

komplettes Bild inklusive Ränder



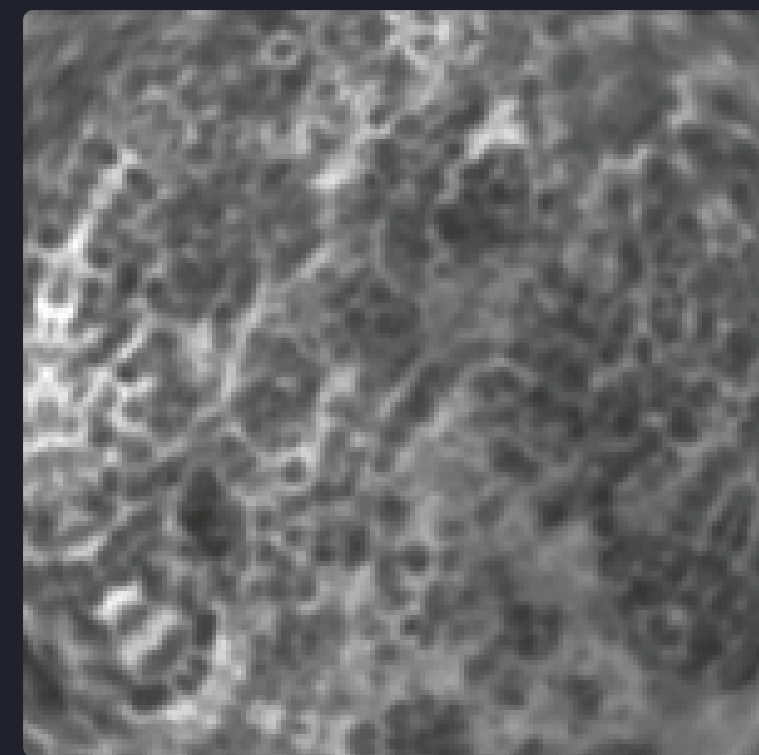
02 Cropped Square

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 (Transfer Learning)
- Feinabstimmung (Fine-Tuning) auf CLE-Daten

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

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

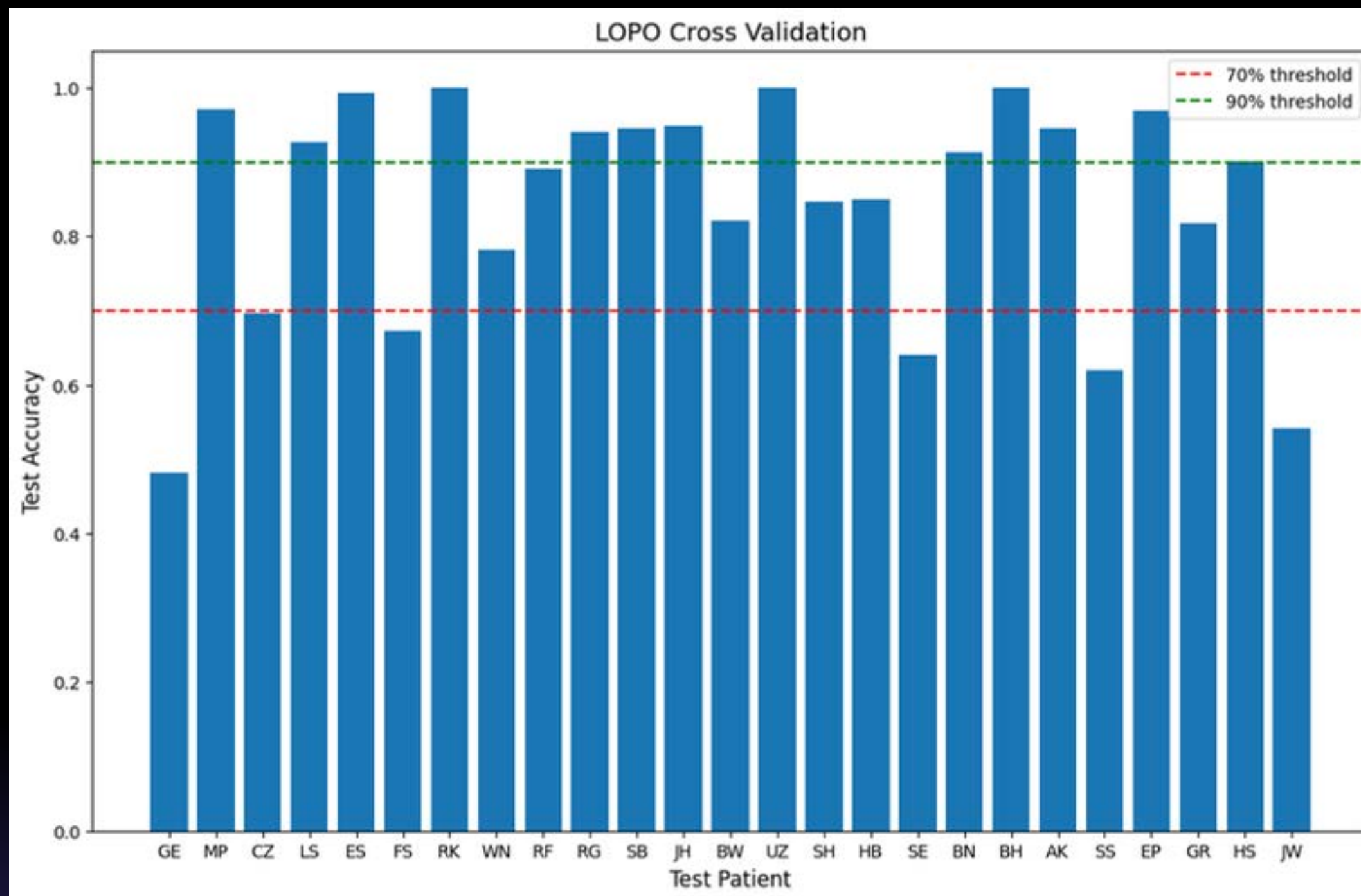
Sensitivity

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

Specificity

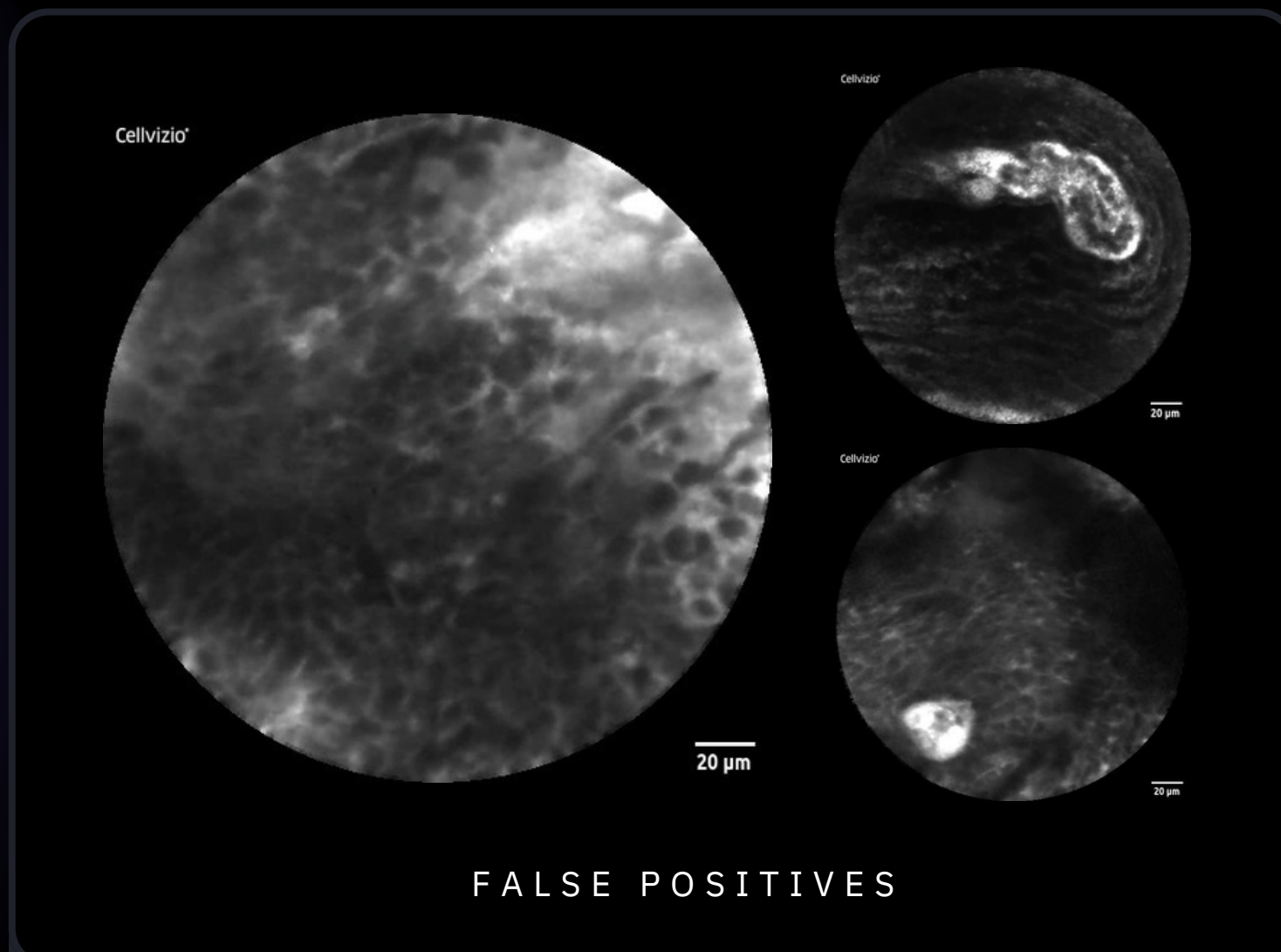
→ Wie viele **gesunde** Bilder richtig erkannt wurden

Leave-One-Patient-Out Validation

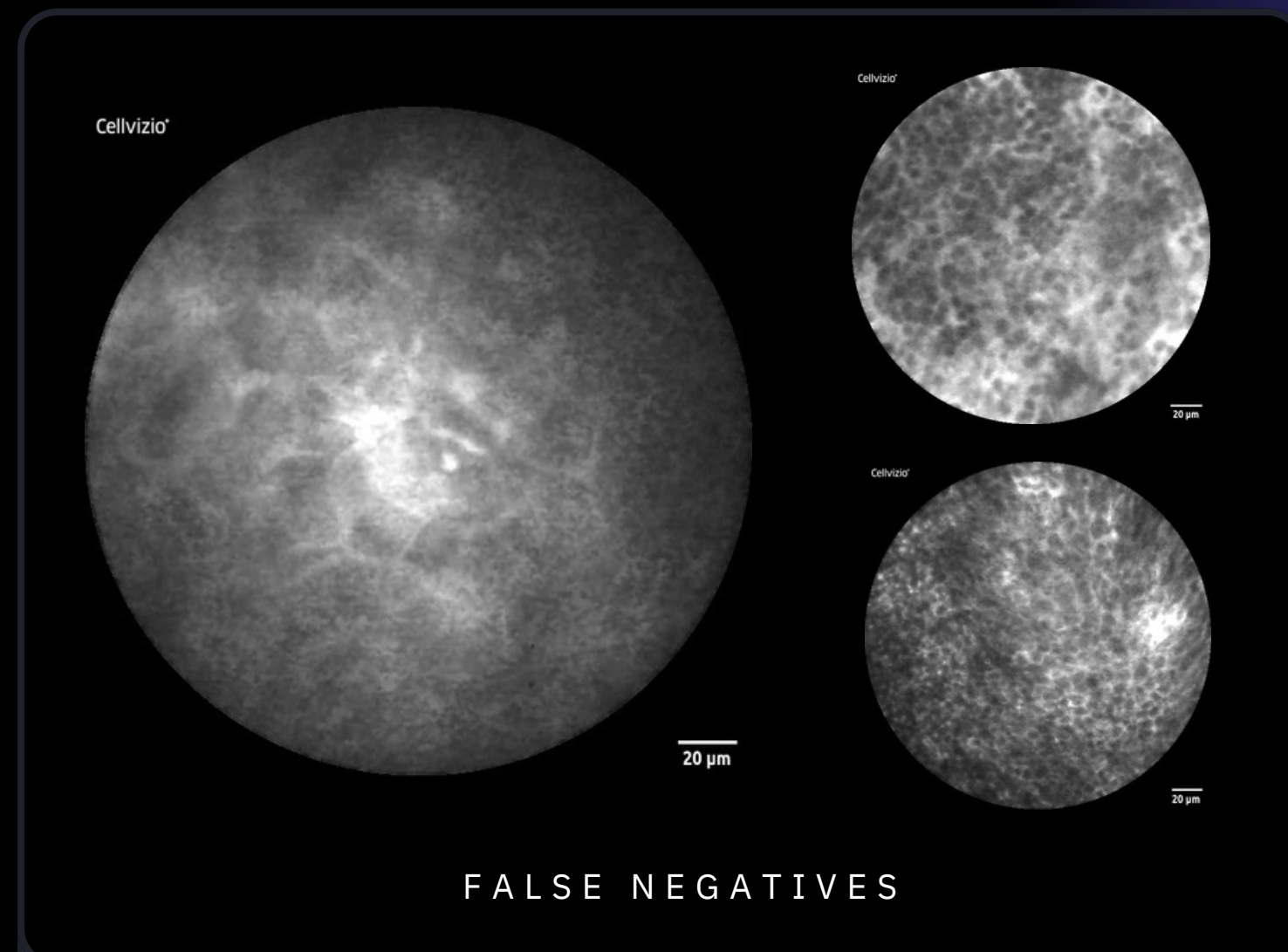


- Starke Varianz der Modellleistung zwischen den Patienten
- 13 Patienten mit Accuracy **> 90 %**, davon 3 mit **100 %**
- aber auch 5 Patienten **< 70 %**

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

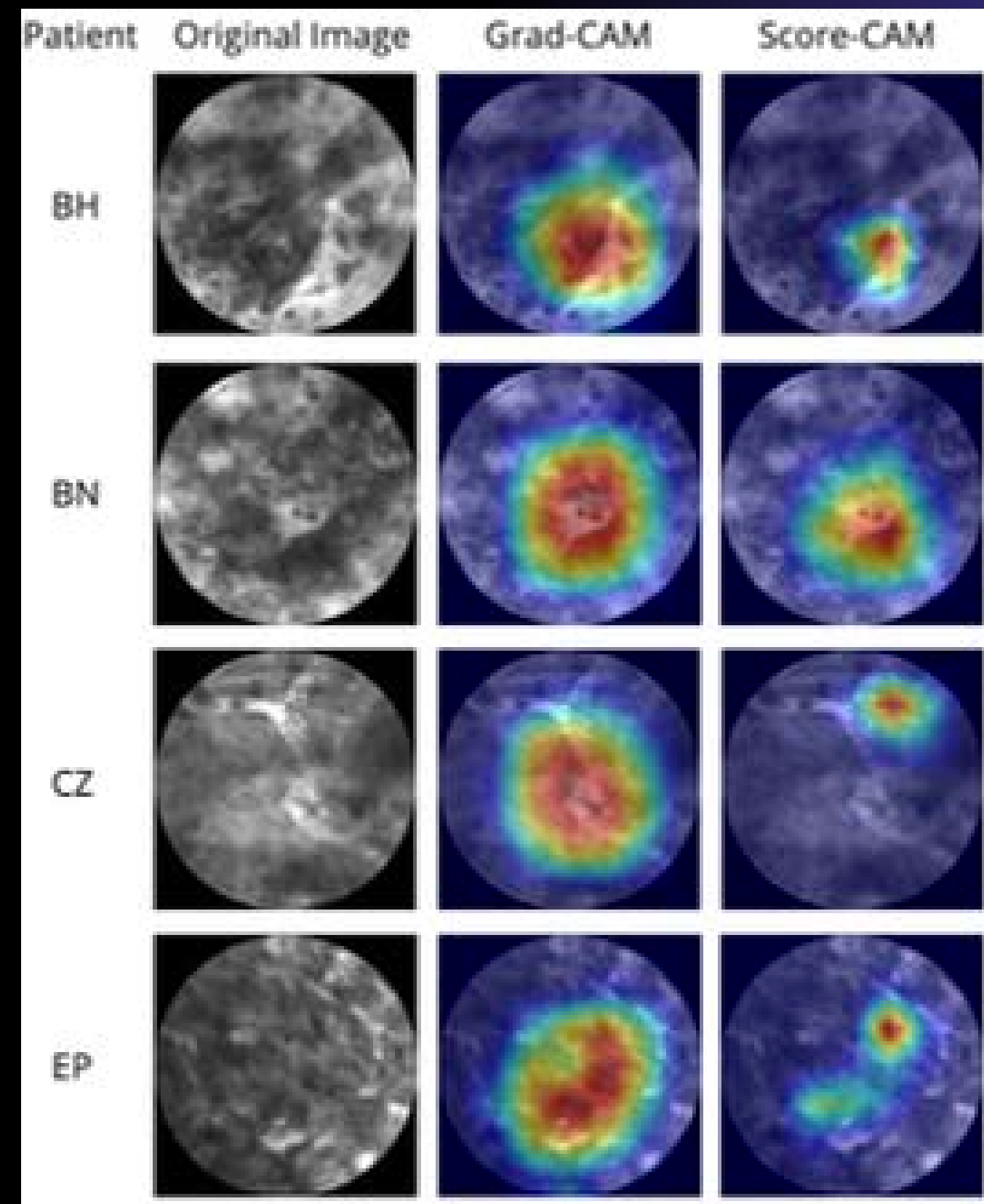
Class Activation Maps

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



Wir gestalten KI

Wir gestalten KI-Kollaboration

Human-Centered AI

Verständnis | Vertrauen* | Befähigung

Was ist der “*Job to be Done*”?

technisch-zentriert

Mit möglichst hoher Accuracy malignes von gesundem Gewebe unterscheiden können.

Was ist der “*Job to be Done*”?

technisch-zentriert

Mit möglichst hoher Accuracy malignes von gesundem Gewebe unterscheiden können.

menschen-zentriert

Sicherheit schaffen bei einer Entscheidung, die unter Ungewissheit, Zeitdruck und sehr hoher Verantwortung getroffen wird.

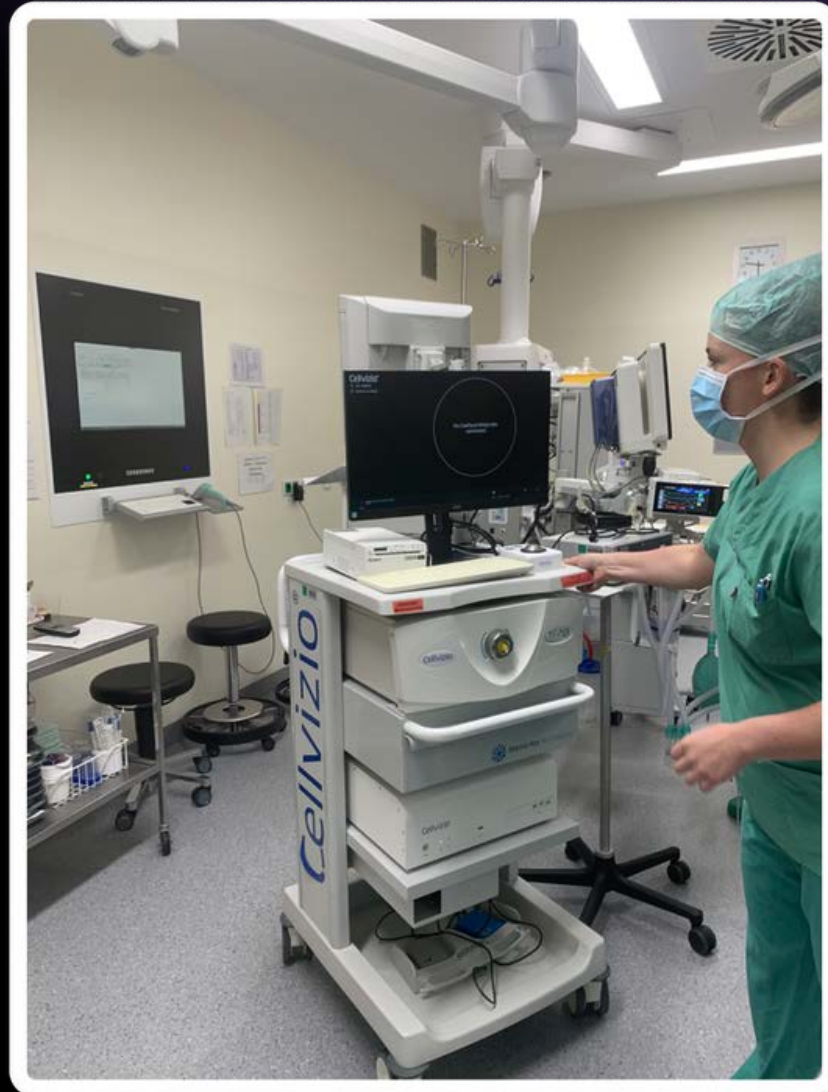


WIE SIEHT HUMAN-CENTERED AI IN DER PRAXIS AUS?

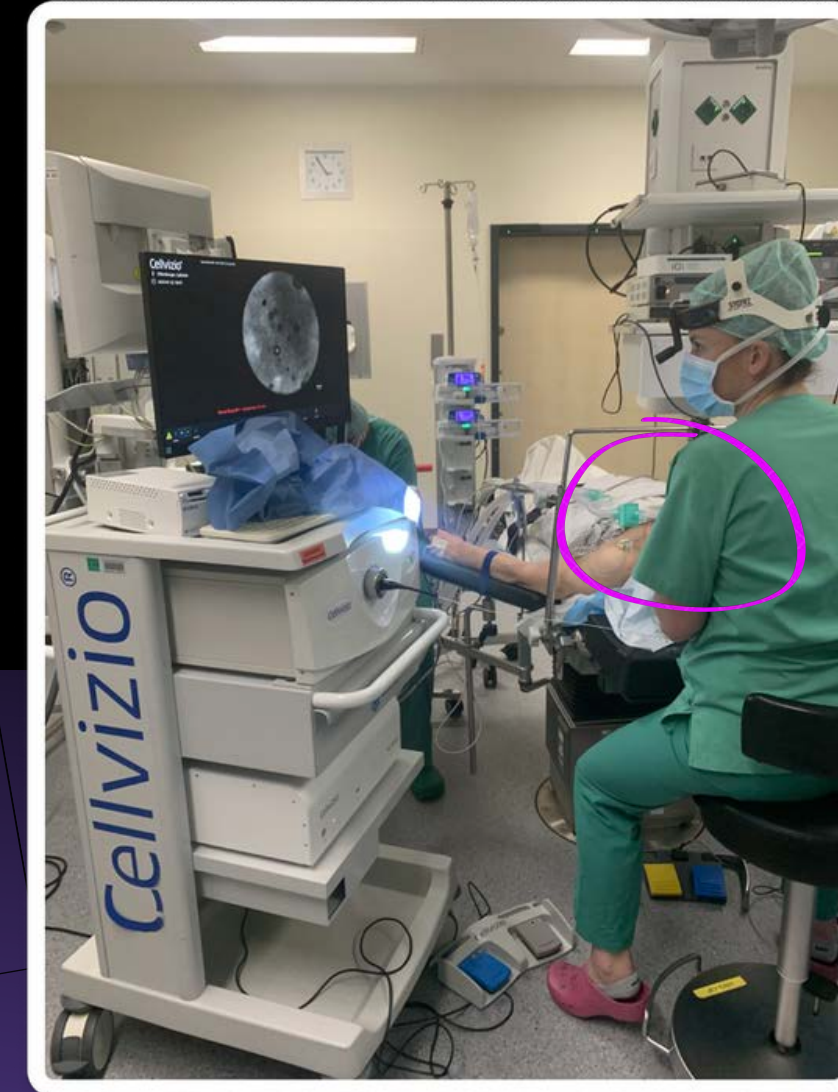
So zum Beispiel ;)



Pre-Study: User Shadowing



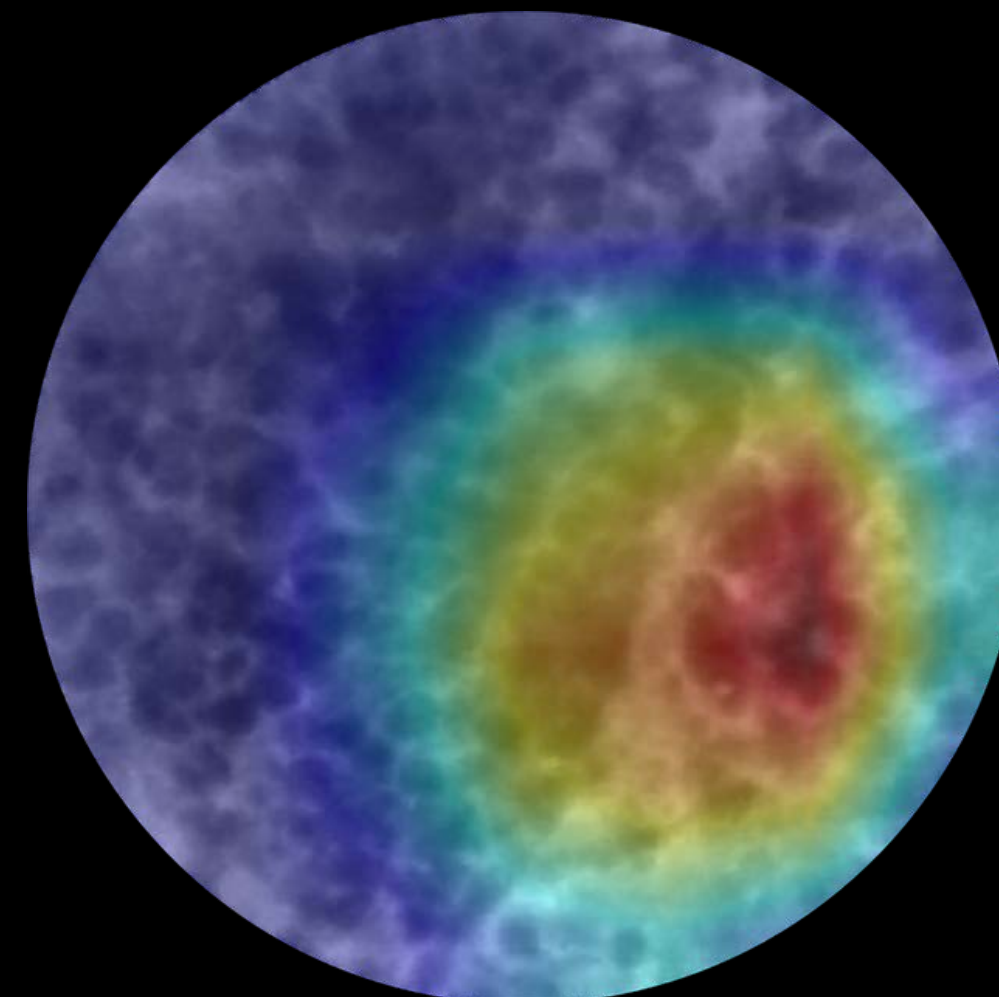
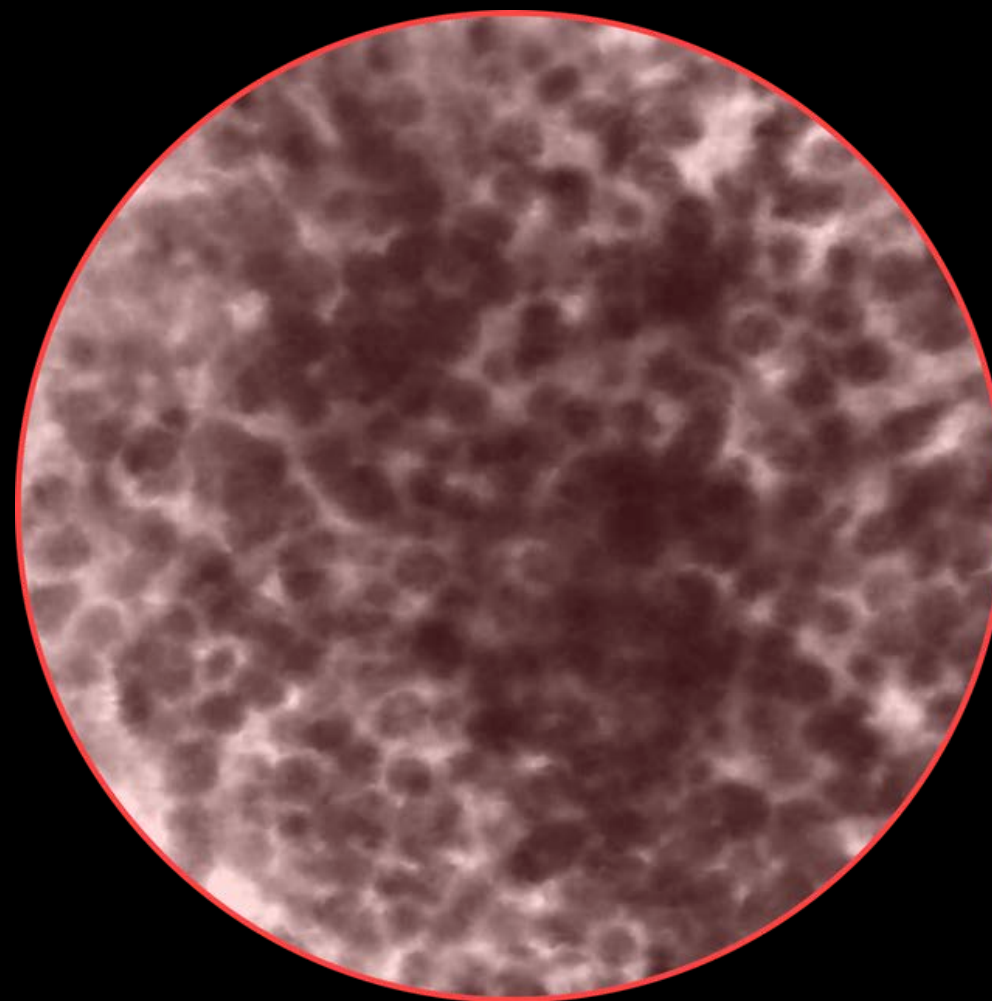
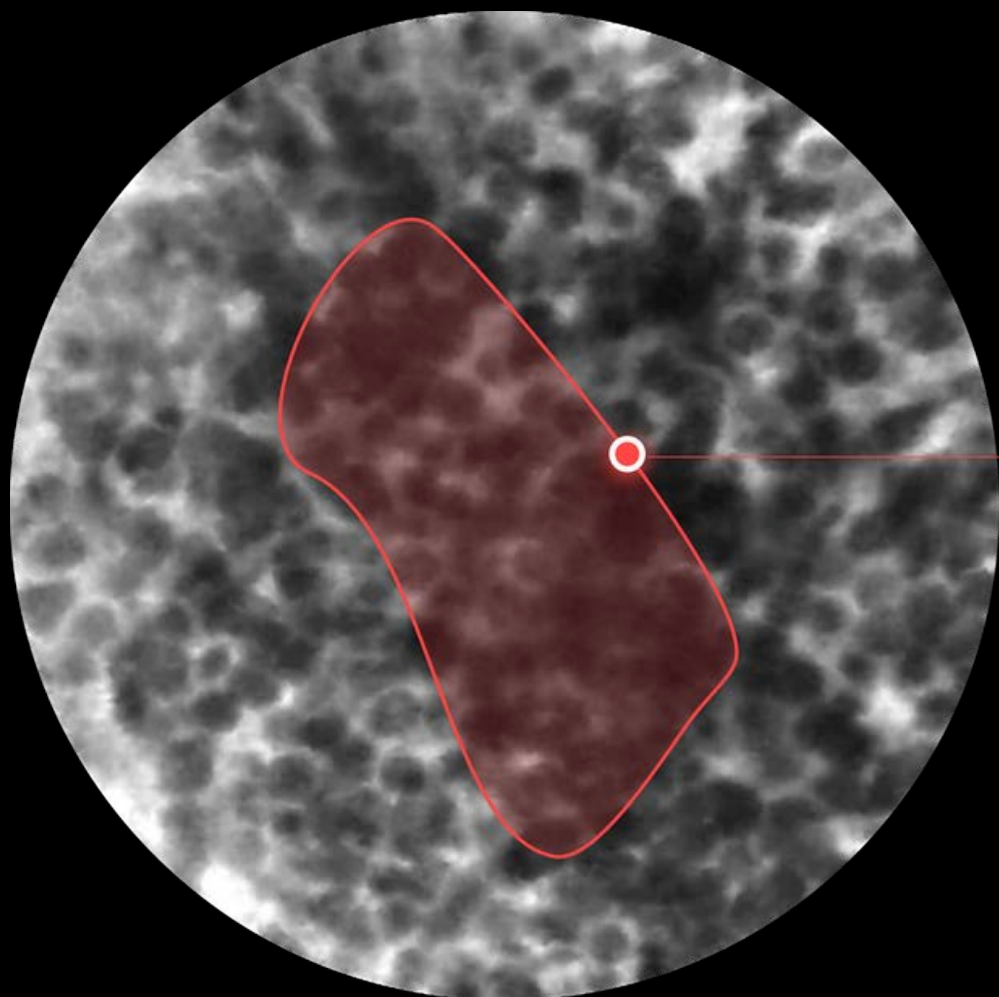
OPERATEUR:INNEN TRAGEN
HANDSCHUHE



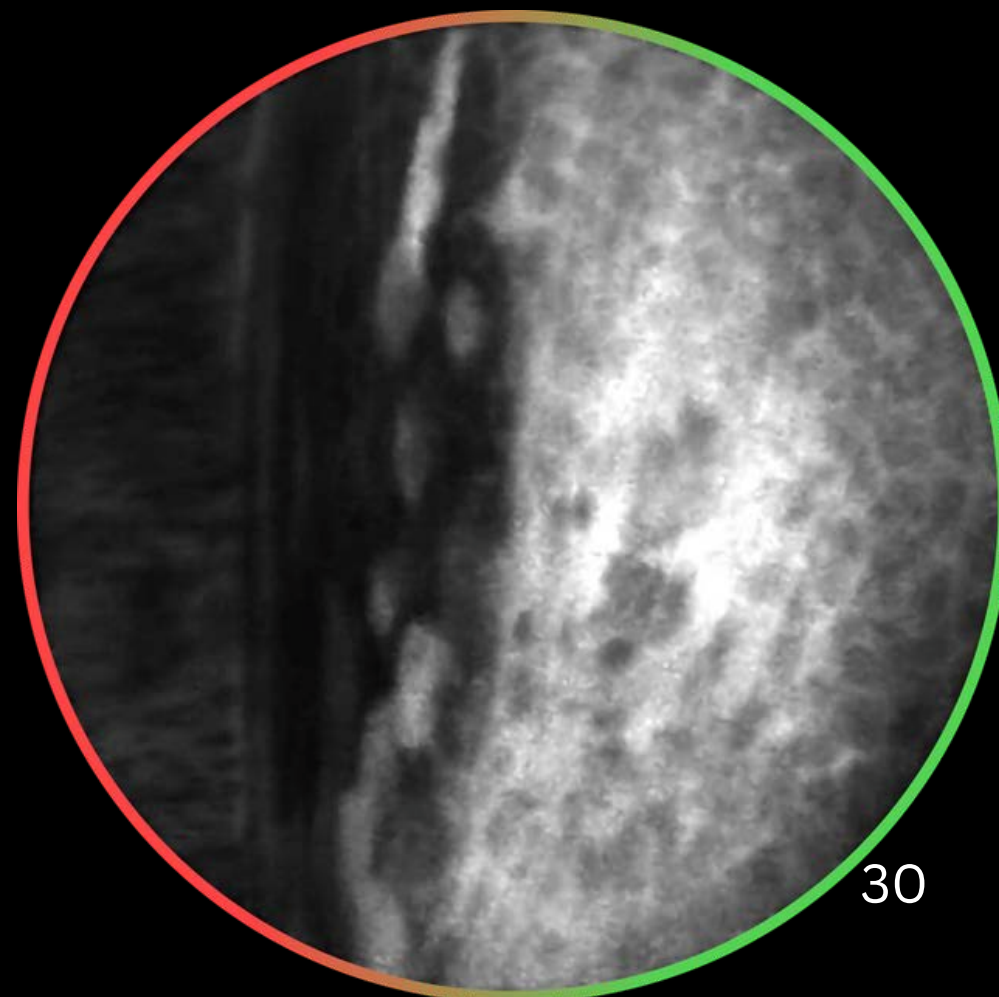
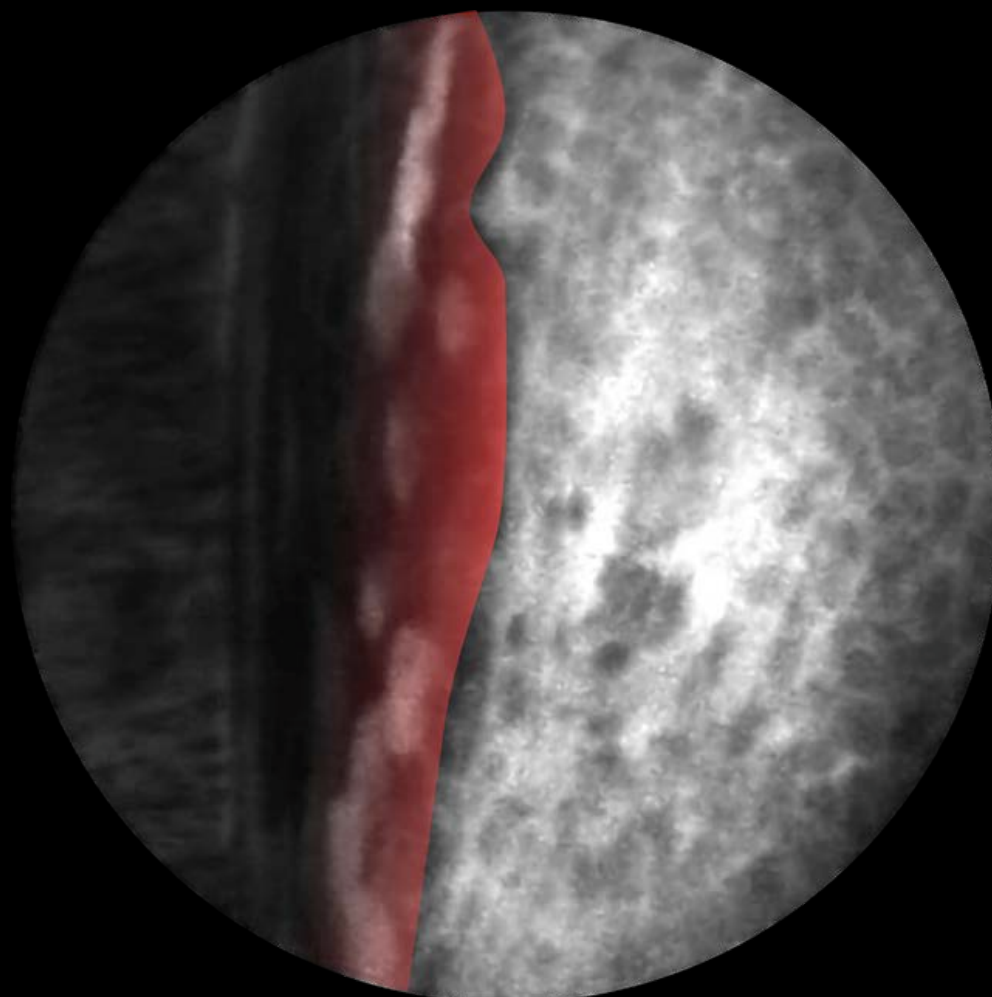
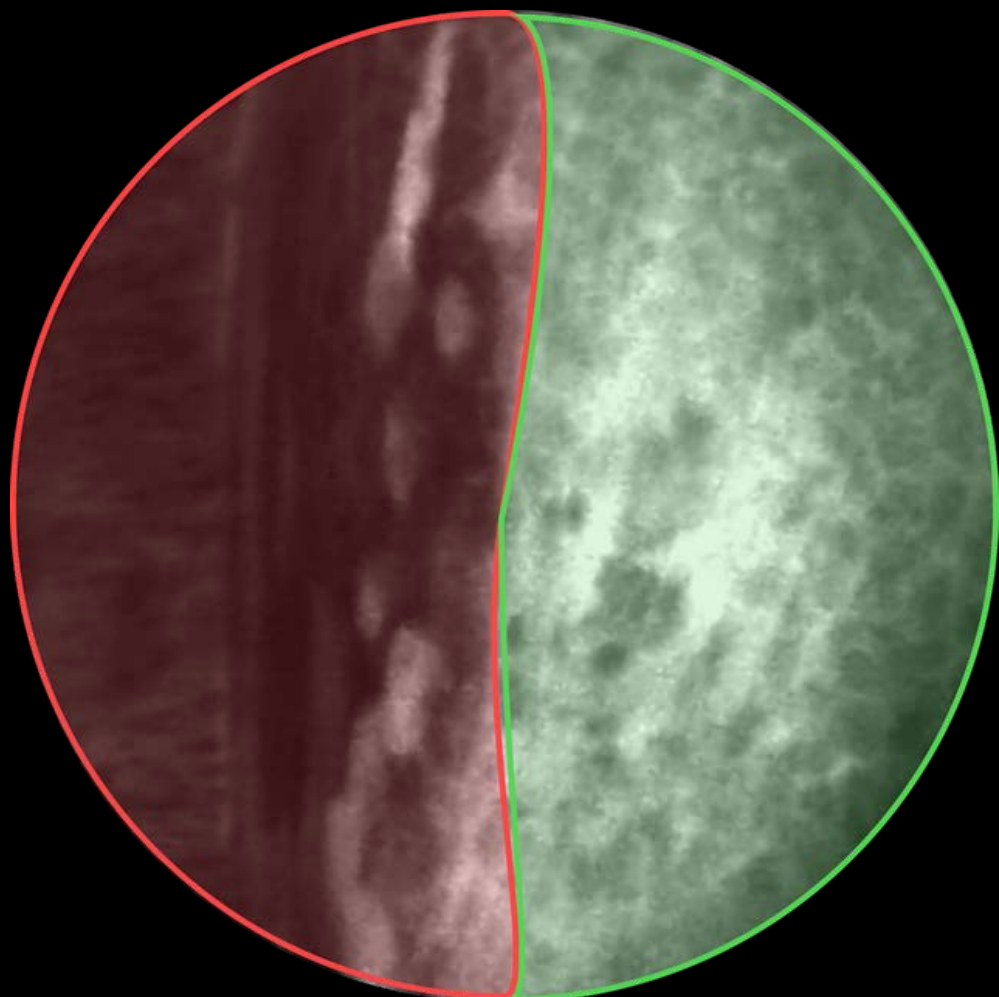
HÄNDE WERDEN WÄHREND DER
OP GEBRAUCHT

**Klassische Interaktionsmuster wie Klicken,
Zoomen, oder die Nutzung von Maus/Keyboard
sind aktuell weder möglich noch effektiv!**

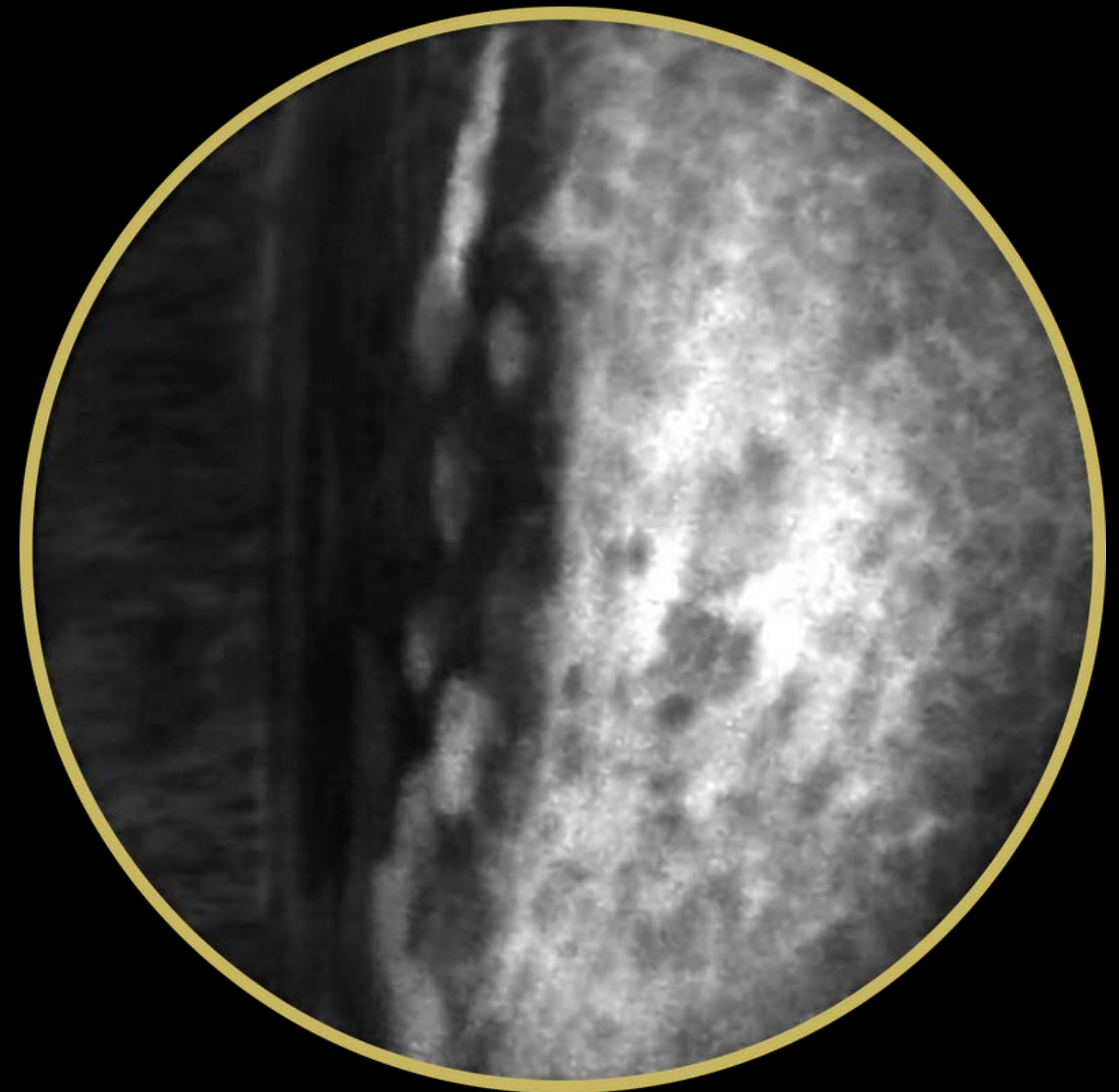
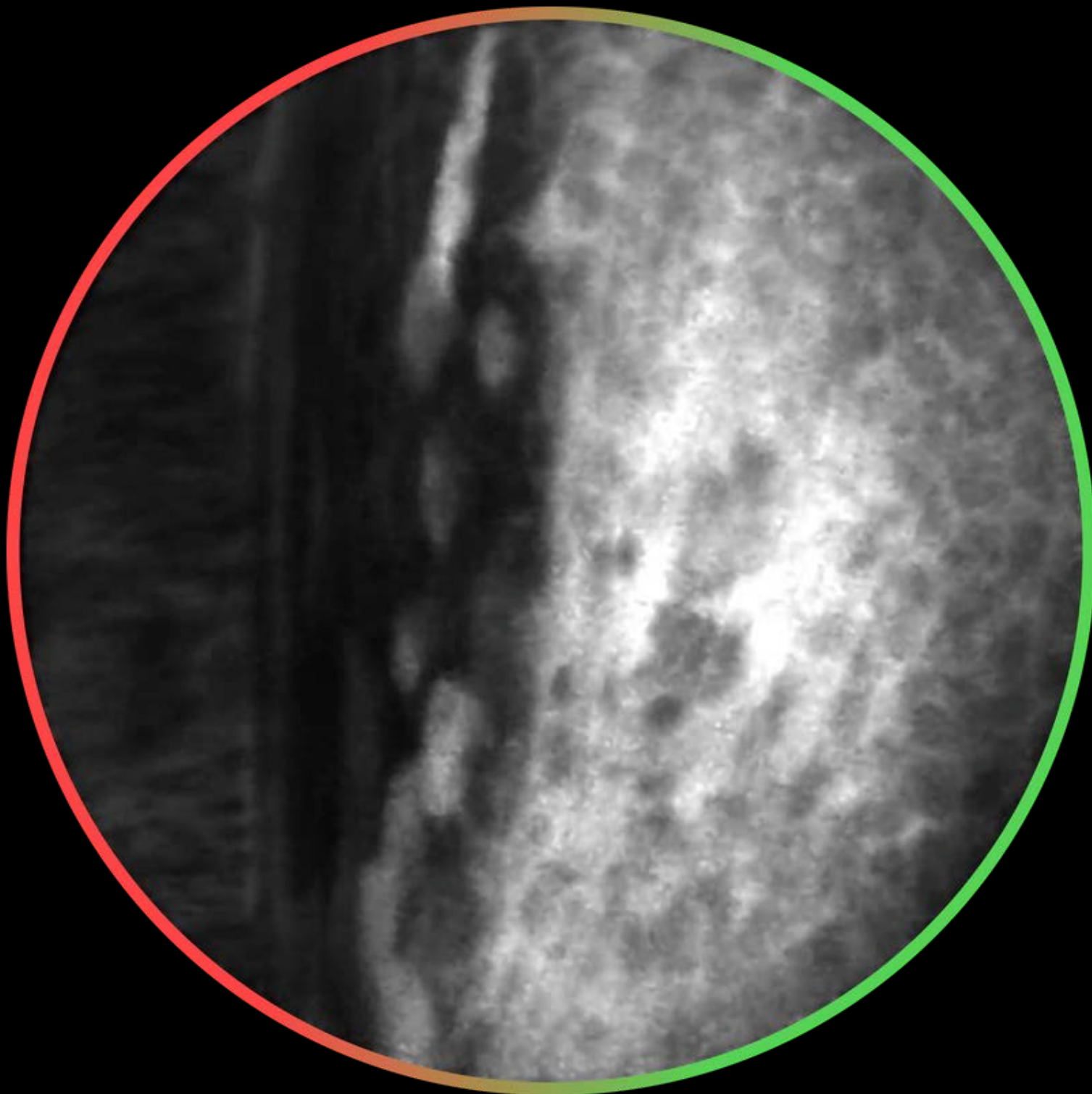
TUMOR



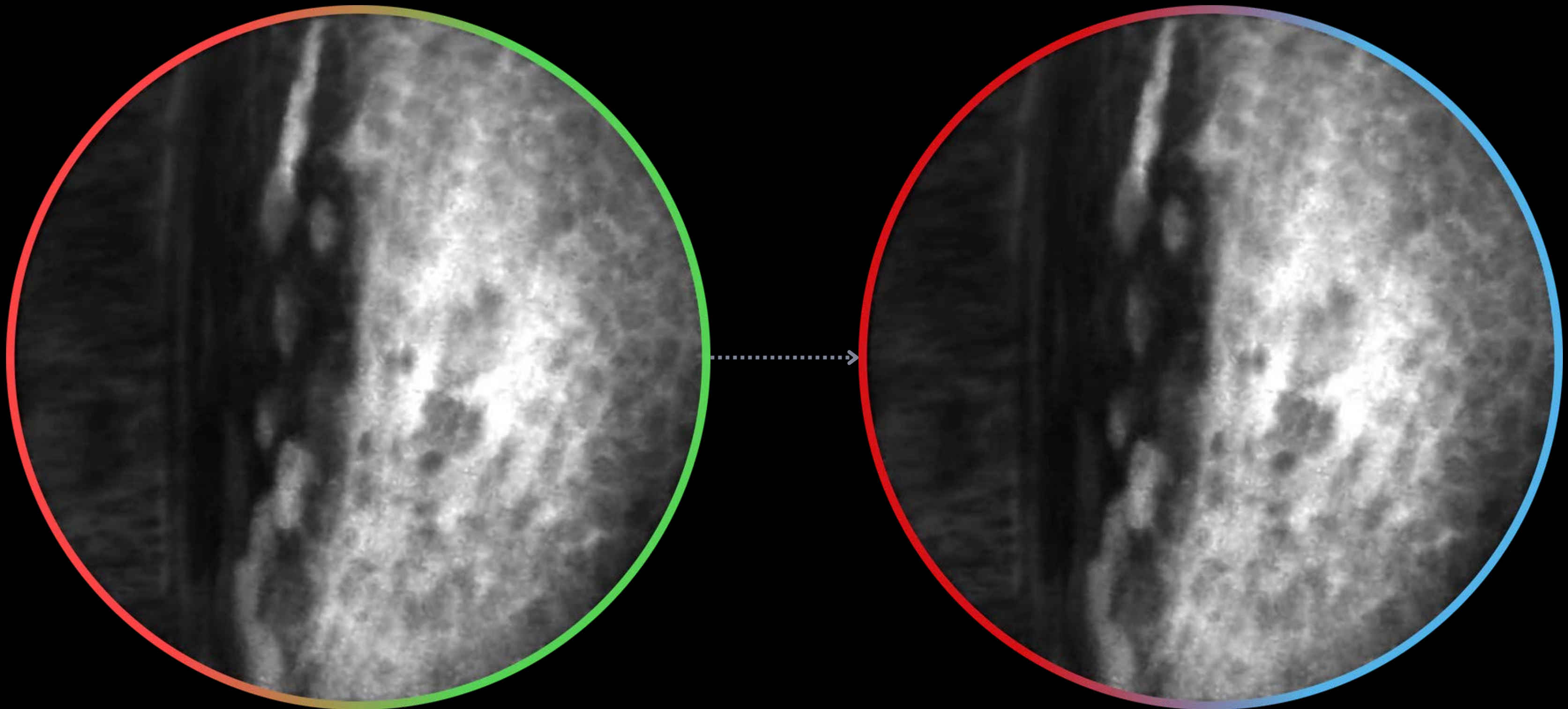
RANDBEREICH



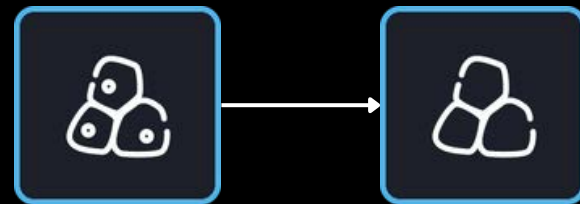
Ampel-Farbschema ist nicht barrierefrei



Ampel-Farbschema ist nicht barrierefrei



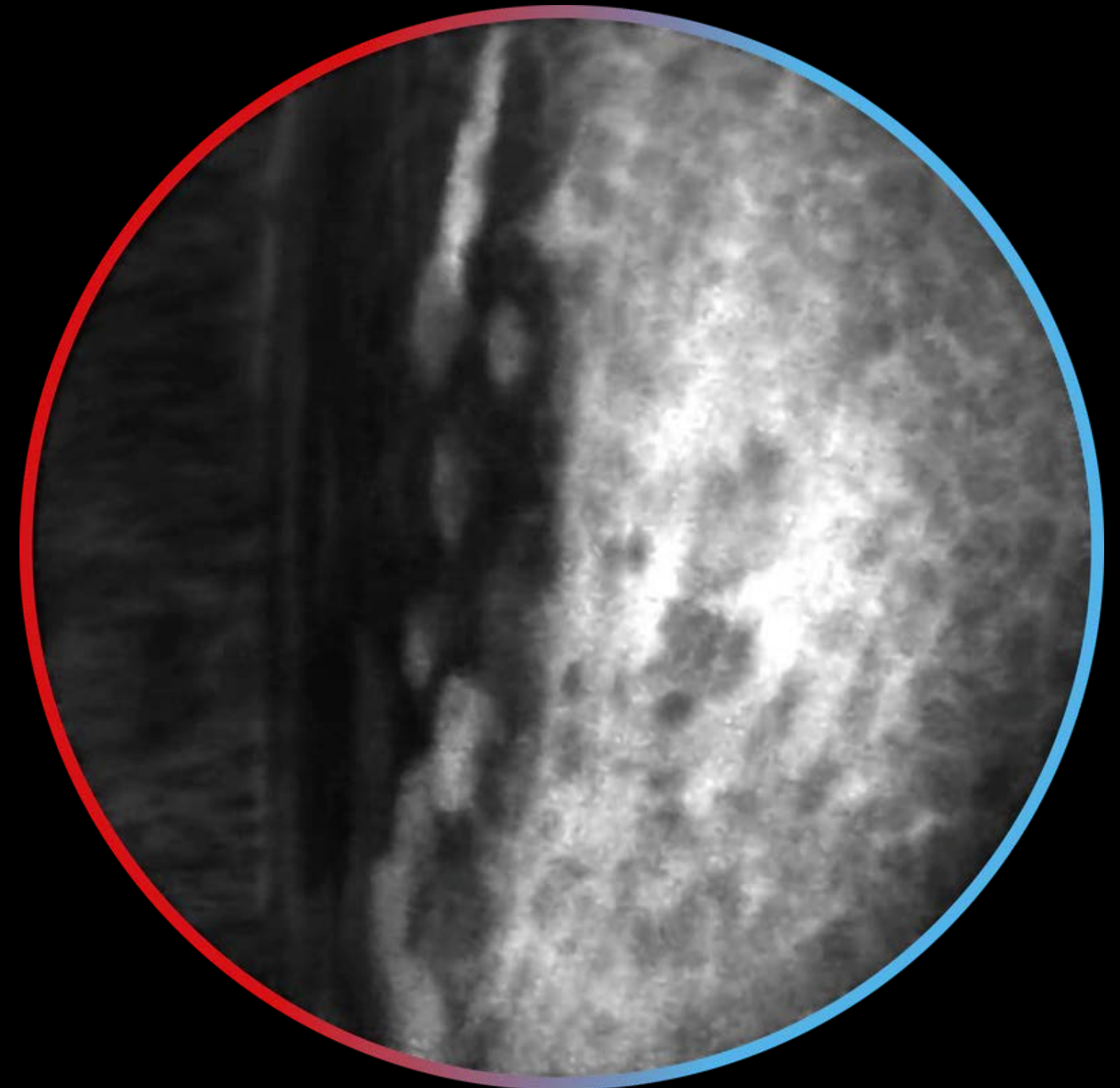
Mentale Modelle in Icons übersetzen



 healthy tissue



 suspicious tissue



Interface Ist-Zustand

The screenshot displays the Cellvizio software interface. At the top left, the 'Cellvizio' logo is visible, along with user information 'Doe, Jane' and a timestamp '2025-06-20, 09:39'. The top right shows the device status 'GastroFlex UHD - CM - 13247 (18 Users left)'. The central feature is a large circular 'In vivo Echtzeit CLE-Bilder' (live CLE image). To the left of this image is a 'Gespeicherte Video Sequenzen' (saved video sequences) list containing two entries: '001 - SEQUENCE' with a duration of '01:52' and '002 - SEQUENCE' with a duration of '01:59'. At the bottom, a 'Vorqualifizierung der Bildqualität' (pre-qualification of image quality) bar is shown, consisting of alternating green and grey segments. On the right side, three labels with dotted lines point to their respective UI elements: 'Gespeicherte Video Sequenzen' points to the sequence list, 'In vivo Echtzeit CLE-Bilder' points to the central circular image, and 'Vorqualifizierung der Bildqualität' points to the quality bar.

Suspicious Tissue

Model Confidence

Degree of certainty that
the tissue is malignant

High **92%**



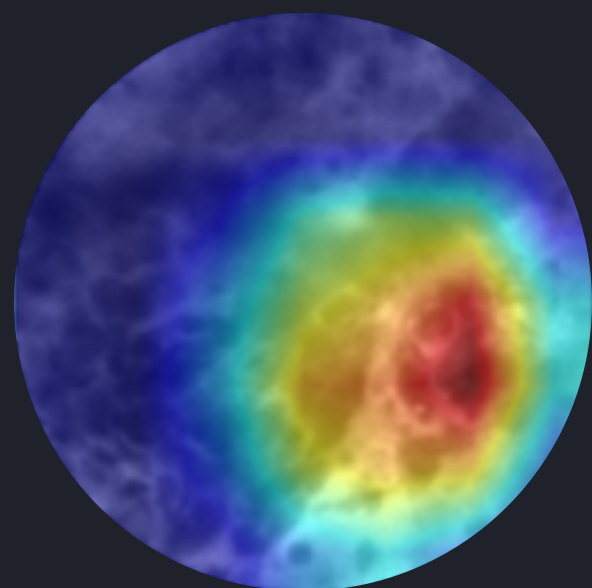
Most Influential Morphological Features

inhomogenous cell pattern

cell conglomerates

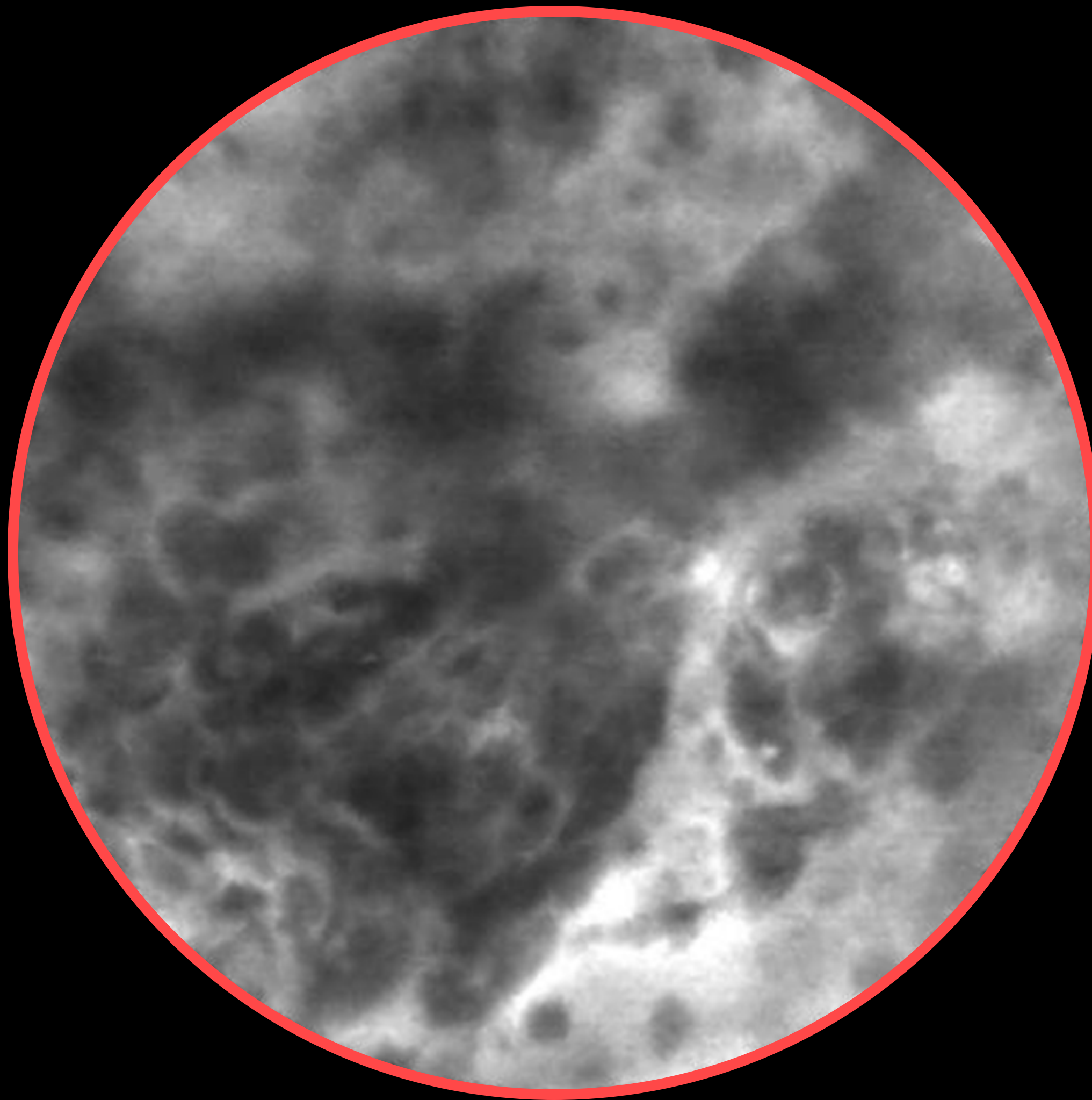
vascular fluorescein leakage

Most Influential Image Region



relevant

not relevant



Current Score

15



Classification

Please classify this case

 Suspicious Tissue

 Healthy Tissue

Confidence

Please rate how confident you feel about your classification

75%

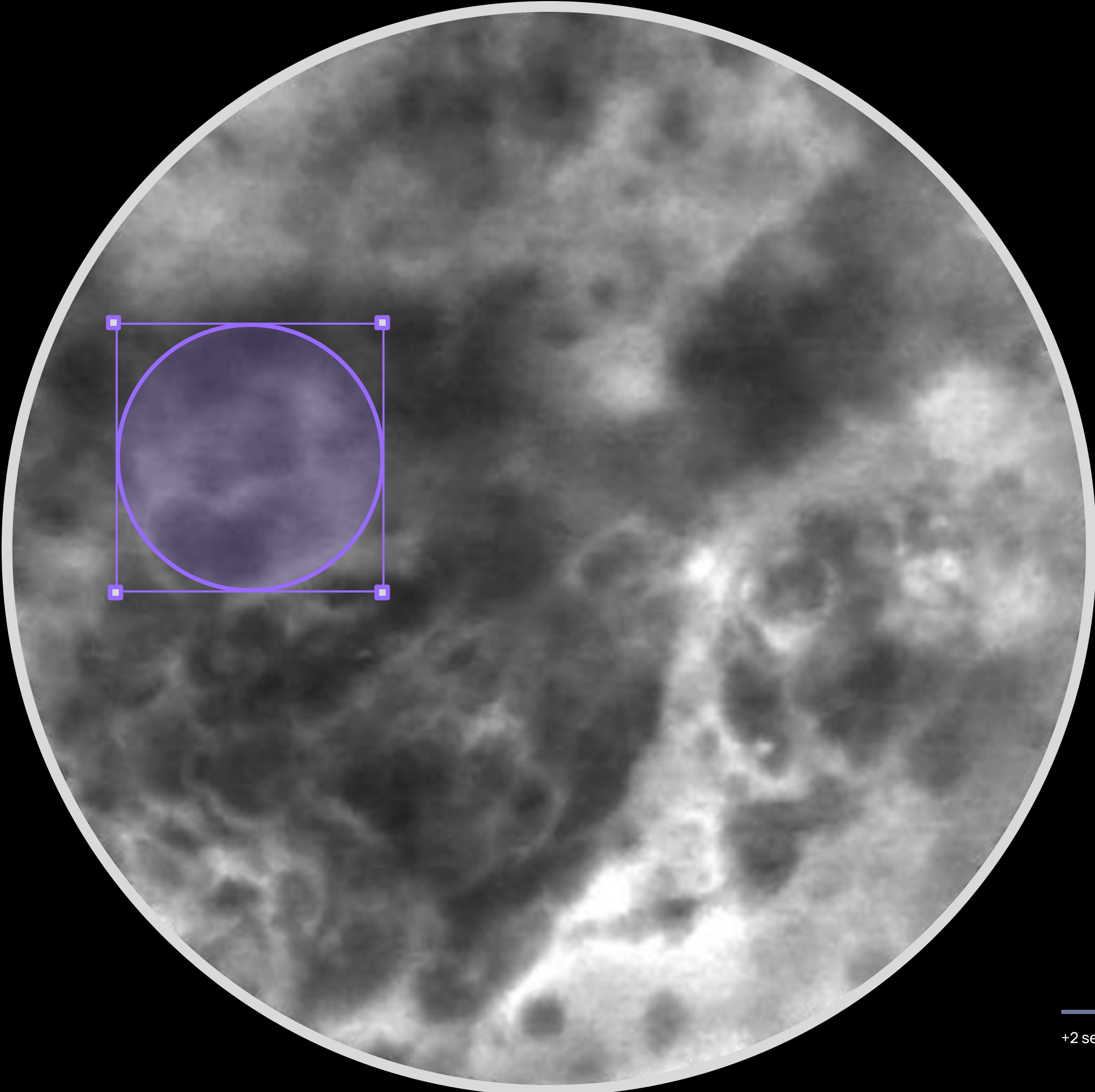
Most Influential Image Region

Select the region that contributed most to your decision



Select

See result



 Zoom

 Move

 Contrast

 Reset

+2 sec

-2 sec

Human-AI Collaboration Principles



Chirurg:innen nutzen Erklärungen um die KI-Entscheidung auf Plausibilität zu überprüfen



Das Kommunizieren von Modell-Unsicherheit stärkt die Kalibrierung von Vertrauen



Verfügbarkeit von KI-Unterstützung ist eine subjektive Präferenz



Informationsdichte: Die richtigen Erklärungen anzeigen bewahrt den Fokus

→ Daraus ergeben sich die Roadmap für eine nutzerzentrierte Weiterentwicklung

Modellperformance verbessern

Vision Transformer

Weitere Modellklassen evaluieren

Spatio-temporale Analyse

Videosequenzen statt Einzelbilder

Erklärbarkeit verbessern

Uncertainty Quantification

Kalibrierte Unsicherheit anzeigen

Morphologische Feature-Extraktion

Klinisch interpretierbare Merkmale

**Nutzerzentrierte Evaluation bestimmt
unsere Priorisierung**

Wir sind keine Chirurg:innen.

Vielen Dank!



Alina Döring



Stefanie Marx

