

## Mehr Transparenz bei KI in der Diagnostik

*Informatiker Alessandro Wollek und Ethikforscherin Theresa Willem im Interview*

**Künstliche Intelligenz (KI) könnte Diagnosen in der Radiologie unterstützen, allerdings ist es bisher oft schwierig, Empfehlungen der KI nachzuvollziehen. Forschende haben nun untersucht, ob und wie visuelle Darstellungen der KI-Bildauswertung - sogenannte Saliency Maps - helfen können. Im Interview erklären Alessandro Wollek, Doktorand im Bereich Informatik, und Theresa Willem, Doktorandin in der Medizinethik, wie ihre Ergebnisse dazu beitragen können, KI transparenter zu machen.**

### **Herr Wollek, welche Rolle spielt KI in der radiologischen Diagnostik?**

Wollek: Durch die Unterstützung von KI könnten Radiolog:innen schneller und sicherer Diagnosen treffen. Dies wäre besonders hilfreich in Stresssituationen, für Radiolog:innen mit noch wenig Praxiserfahrung und um mehr Zeit für die Beratung zu schaffen.

### **Wo liegt dabei die Schwierigkeit?**

Willem: Der Entscheidungsprozess der KI muss für die Ärzt:innen soweit nachvollziehbar sein, dass sie bestmöglich entscheiden können, inwieweit sie der KI-Empfehlung vertrauen sollten. Das ist bisher schwierig. Hier haben wir angesetzt, wir wollten herausfinden, wie sich der Prozess transparenter gestalten lässt.

### **Wie war die Ausgangslage für Ihre Studie?**

Wollek: Bisher werden in der Radiologie sogenannte Convolutional Neural Network (CNN) Algorithmen eingesetzt. In anderen Bereichen verwendet man dagegen bereits die neueren Vision Transformer Algorithmen (ViT). Diese sind zwar in der Regel besser, machen also weniger Fehler bei der Diagnose, brauchen jedoch riesige Datenmengen. Da derart große Datensätze in der Radiologie nicht zur Verfügung stehen, kamen ViT Algorithmen dort bisher nicht zum Einsatz.

Wir haben nun für die radiologische Bildauswertung anhand des Beispiels Pneumothorax untersucht, ob ViT Algorithmen trotz der verhältnismäßig kleinen Datensätze genauso gut wie CNN Algorithmen funktionieren. Dies hat sich bestätigt, bei all unseren gemessenen Parametern haben ViTs vergleichbar wie CNNs abgeschnitten.

### **Worin liegt der Vorteil, dass nun Vision Transformers ebenso verwendet werden können?**

Wollek: Man kann für jede Empfehlung der KI eine Art Karte erstellen, die zeigt, welche Bildausschnitte die KI für ihren Entscheidungsprozess mit einbezogen hat. Für ViTs kann man dafür einen bestimmten Kartentyp verwenden, den es bei CNNs nicht gibt und von dem wir vermutet haben, dass er hilfreicher ist.

Die „Karten“ heißen Saliency Maps - wörtlich übersetzt Saliency-Karten; sie stellen dar, auf welche Bereiche eines Bildes das neuronale Netz die Vorhersage stützt. Es gibt verschiedene Kartentypen: Gradient-weighted class activation mapping (GradCAM) und Transformer-Multi-Modal-Explainability (TMME). Für CNN Algorithmen ist nur das GradCAM Modell möglich, bei ViT Algorithmen auch

TMME. Wir konnten nun erstmalig TMME Saliency Maps in der Radiologie verwenden und dadurch beide Arten von Saliency Maps in ihrer Qualität vergleichen.

### **Welcher Typ war besser? Und nach welchen Kriterien?**

Wollek: Wir sind von der Frage ausgegangen: Zeigen die Saliency Maps auch wirklich zuverlässig das, was die KI für die Empfehlung nutzt? Möglich wäre auch, dass die Karten fehlerhaft sind und zwar Ausschnitte hervorheben, die uns als Betrachter zunächst plausibel erscheinen, aber nicht wirklich dem entsprechen, worauf die KI ihre Aussage gestützt hat. Um das zu untersuchen haben wir beide Saliency Maps nach quantitativen Kriterien ausgewertet. Bei allen von uns gemessenen Kriterien schnitten die TMME Saliency Maps besser ab, waren also vereinfacht gesagt zuverlässiger darin zu zeigen, welche Bestandteile eines Bildes die KI ihrer Diagnoseempfehlung zugrunde gelegt hat.

### **Wie helfen diese Ergebnisse den Radiolog:innen?**

Willem: In unserer Pilotstudie haben drei Radiolog:innen je 70 Röntgenbilder mit TMME und 70 Röntgenbilder mit GradCAM Saliency Maps analysiert. Dabei beurteilten sie 47 Prozent der TMME Saliency Maps als hilfreich bei der Diagnose, dagegen nur 39 Prozent der GradCAM Saliency Maps.

### **Wie können die Erkenntnisse zukünftig genutzt werden?**

Willem: Wir starten demnächst eine weitere Studie mit deutlich mehr Radiolog:innen. Sollte sich das Ergebnis bestätigen, könnten Saliency Maps zukünftig in der KI-gestützten Diagnostik eingesetzt werden. Wir möchten die Diagnoseempfehlungen von KI für die klinische Praxis damit ein Stück weit transparenter machen.

### **PUBLIKATION**

Alessandro Wollek, Robert Graf, Saša Čečatka, Nicola Fink, Theresa Willem, Bastian O. Sabel, and Tobias Lasser: Attention-based Saliency Maps Improve Interpretability of Pneumothorax Classification, *Radiology: Artificial Intelligence* (2023), DOI <https://doi.org/10.1148/ryai.220187>

### **WEITERE INFORMATIONEN**

Der Leiter der Studie PD Dr. Tobias Lasser forscht am Munich Institute of Biomedical Engineering (MIBE), einem Integrativen Forschungsinstitut der Technischen Universität München (TUM). Am MIBE entwickeln und verbessern Forschende aus der Medizin, den Natur- und Ingenieurwissenschaften und der Informatik gemeinsam Verfahren zur Prävention, Diagnose und Behandlung von Krankheiten. Die Aktivitäten reichen dabei von der Untersuchung grundlegender wissenschaftlicher Prinzipien bis zu deren Anwendung in medizinischen Geräten, Medikamenten oder Computerprogrammen.

An der Studie beteiligt waren Forschende der Technischen Universität München (TUM) und der Ludwig-Maximilians-Universität (LMU).

Die Forschungsarbeit wurde durch das Bundesministerium für Gesundheit gefördert.