

Automation Bias bei der KI-gestützten Detektion zerebraler Aneurysmen mittels Time-of-Flight (TOF) MR-Angiografie

Su Hwan Kim¹, Severin Schramm¹, Evamaria Olga Riedel¹, Lena Schmitzer¹, Enrike Rosenkranz¹, Olivia Kertels¹, Jannis Bodden¹, Karolin Paprottka¹, Dominik Sepp¹, Martin Renz¹, Jan Kirschke¹, Thomas Baum¹, Christian Maegerlein¹, Tobias Boeckh-Behrens¹, Claus Zimmer¹, Benedikt Wiestler¹, Dennis M. Hedderich¹

¹ Institut für diagnostische und interventionelle Neuroradiologie, TUM-Universitätsklinikum, München

Hintergrund

KI-Systeme haben das Potenzial, die Detektion zerebraler Aneurysmen zu unterstützen. Doch die Rolle des Automation Bias (Tendenz von Menschen, sich übermäßig auf automatisierte Systeme zu verlassen) in diesem Kontext ist noch unklar.

Methoden

Neun Ärztinnen und Ärzte mit unterschiedlichem neuroradiologischem Erfahrungsgrad bewerteten 20 Time-of-Flight (TOF)-MR-Angiographien auf das Vorhandensein von Aneurysmen und erteilten Follow-Up-Empfehlungen. Jeder Fall wurde mit und ohne KI-Unterstützung bewertet, wobei zwischen den beiden Sitzungen mindestens vier Wochen lagen. Es wurden absichtlich zehn Fälle mit min. einem falsch-positiven KI-Befund eingeschlossen. Der Referenzstandard wurde durch zwei neuroradiologische Experten festgelegt. Die Aneurysma-Bewertungen, Follow-Up-Empfehlungen und die Interpretationsdauer wurden mithilfe von Wilcoxon-Signed-Rank-Tests ausgewertet.

Ergebnisse

Die Exposition der Teilnehmer mit falsch-positiven KI-Befunden führte zu signifikant höheren Aneurysma-Bewertungen ($p = 0,01$). Unerfahrene Teilnehmer wurden durch falsch-positive KI-Befunde besonders stark beeinflusst. Die unerfahrenen Teilnehmer empfahlen bei falsch-positiven KI-Befunden zudem umfangreichere Follow-Up-Untersuchungen ($p = 0,005$). Die Interpretationsdauer war mit KI-Unterstützung bei Teilnehmern aller Erfahrungsgrade signifikant kürzer ($p < 0,001$).

Schlussfolgerung

Falsch-positive Aneurysma-Befunde von KI-Systemen können Fehleinschätzungen und potenziell unnötige Follow-Up-Untersuchungen verursachen, die zu Mehrkosten und zur unbegründeten Besorgnis von Patienten führen können. Herausforderungen bei der Mensch-KI-Interaktion müssen berücksichtigt werden, um eine sichere Anwendung von KI-Systemen zu gewährleisten.

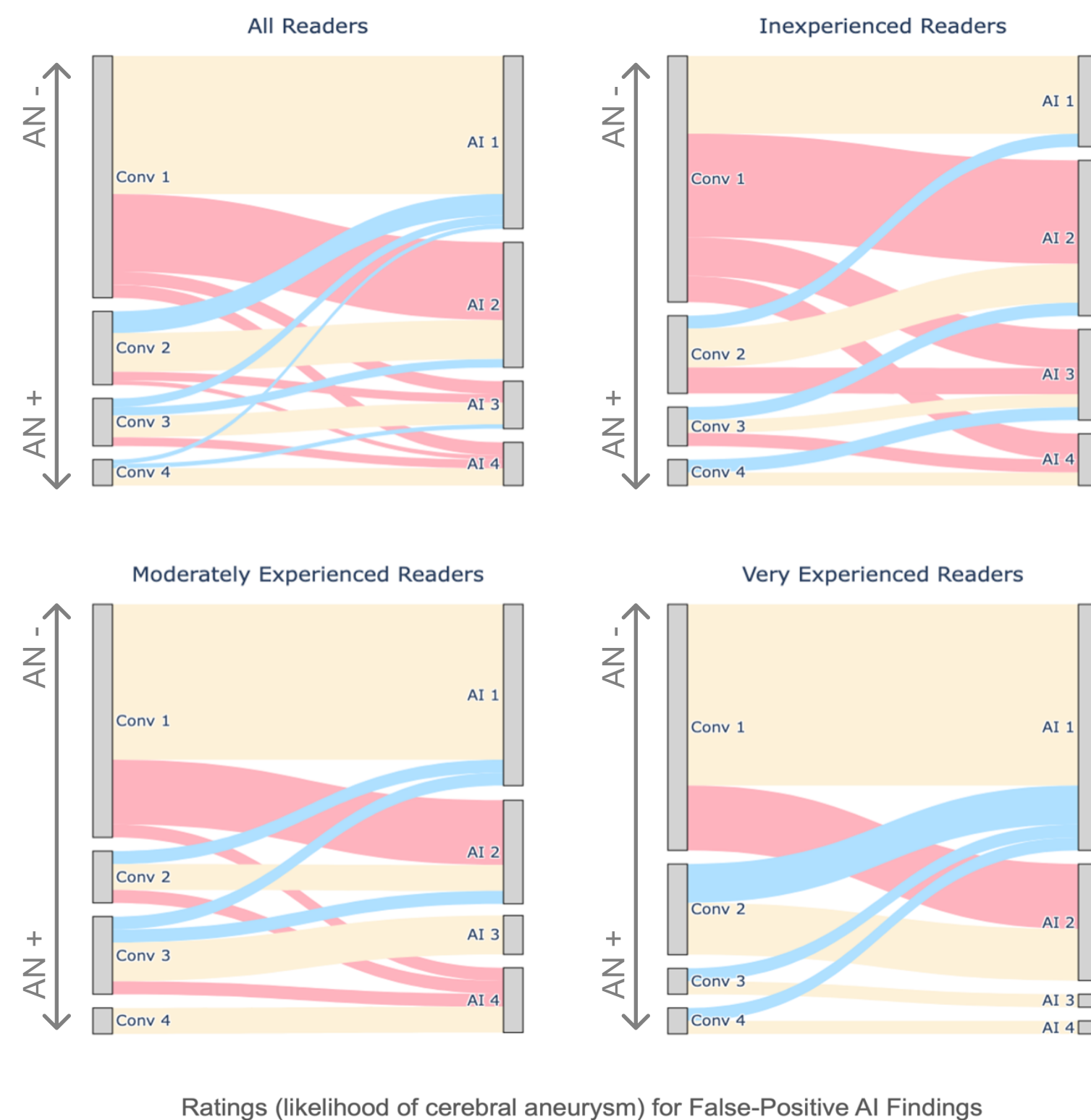


Abbildung 1: Bewertungen für falsch-positive KI-Befunde. Die Diagramme veranschaulichen, wie identische Gefäßsegmente in Kenntnis ("AI") und Unkenntnis ("Conv") falsch-positiver KI-Ergebnisse auf einer Likert-Skala von 1-4 bewertet wurden (1: sicher nicht vorhanden, 4: sicher vorhanden). **Rot:** Bewertung höher mit KI. **Blau:** Bewertung niedriger mit KI. **Gelb:** Bewertung identisch mit KI.

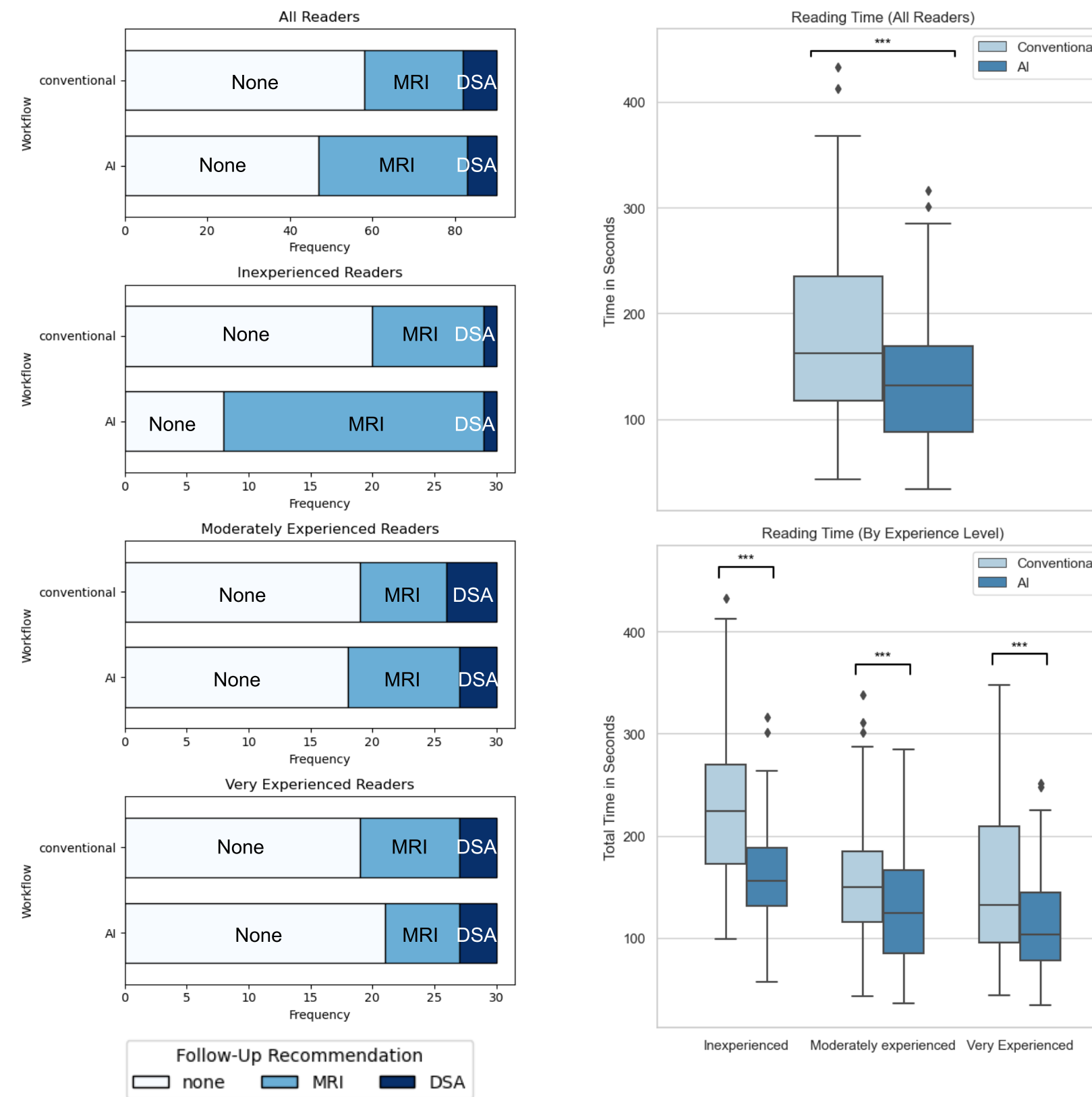


Abbildung 2: Follow-Up-Empfehlungen für falsch-positive KI-Befunde. Unerfahrene Teilnehmer empfahlen signifikant umfangreichere Verlaufskontrollen in Kenntnis falsch-positiver KI-Empfehlungen ($p = 0,005$).

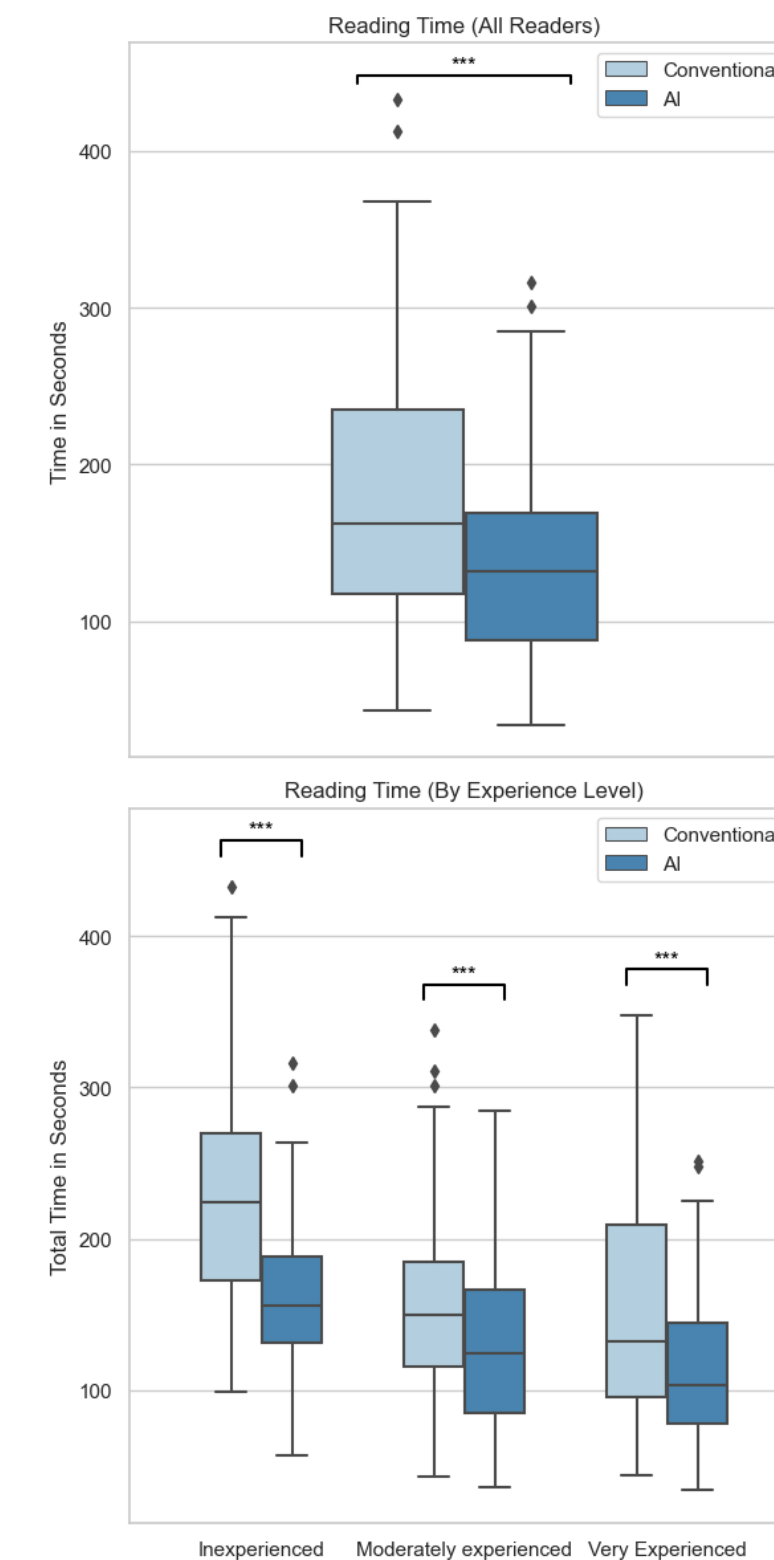


Abbildung 3: Interpretationsdauer. Die Interpretationsdauer war mit KI-Unterstützung bei Teilnehmern aller Erfahrungsgrade signifikant kürzer ($p < 0,001$).