



# **Explainable Artificial Intelligence (XAI) im Gesundheitswesen**

---

# Motivation der Literatur

---

Akzeptanz und  
Vertrauen in die  
Technik steigern

Klinische  
Entscheidungsfindung  
verbessern

Möglichkeit Patienten  
einzubeziehen

Modelle / Methoden  
verbessern



# Paper

Amann et al. *BMC Med Inform Decis Mak* (2020) 20:310  
<https://doi.org/10.1186/s12911-020-01332-6>

BMC Me

Methods of explainability

Outlook and results

Problem and hypothesis

Important reference papers

## RESEARCH ARTICLE

# Explainability for artificial intelligence in healthcare: a multidisciplinary perspective

Julia Amann<sup>1\*</sup>, Alessandro Blasimme<sup>1</sup>, Effy Vayena<sup>1</sup>, Dietmar Frey<sup>2</sup> and Vince I. Madigan<sup>1</sup>  
Precise4Q consortium



diagnostics



Review

## Applications of Explainable Artificial Intelligence in Diagnosis and Surgery

Yiming Zhang<sup>1,2</sup>, Ying Weng<sup>1,\*</sup> and Jonathan Lund<sup>2</sup>

- <sup>1</sup> School of Computer Science, Faculty of Science and Engineering, University of Nottingham Ningbo China, Ningbo 315100, China; yiming.zhang2@nottingham.edu.cn  
<sup>2</sup> School of Medicine, University of Nottingham, Nottingham NG7 2RD, UK; jon.lund@nottingham.ac.uk  
\* Correspondence: ying.weng@nottingham.edu.cn

**Abstract:** In recent years, artificial intelligence (AI) has shown great promise in medicine. However, explainability issues make AI applications in clinical usages difficult. Some research has been conducted into explainable artificial intelligence (XAI) to overcome the limitation of the black-box

## Open Access



ELSEVIER

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: [www.elsevier.com/locate/media](http://www.elsevier.com/locate/media)



## Explainable artificial intelligence (XAI) in deep learning-based medical image analysis



Bas H.M. van der Velden\*, Hugo J. Kuijff, Kenneth G.A. Gilhuijs, Max A. Viergever

Image Sciences Institute, University Medical Center Utrecht, Q.02.4.45, P.O. Box 85500, Utrecht, GA 3508, the Netherlands

### ARTICLE INFO

Article history:  
Received 22 July 2021  
Revised 15 March 2022  
Accepted 2 May 2022

### ABSTRACT

With an increase in deep learning-based methods, the call for explainability of such methods grows, especially in high-stakes decision making areas such as medical image analysis. This survey presents an overview of explainable artificial intelligence (XAI) used in deep learning-based medical image analysis. A framework of XAI criteria is introduced to classify deep learning-based medical image analysis meth-

# Explainability for artificial intelligence in healthcare: a multidisciplinary perspective

---

Journal	BMC medical informatics and decision making, 2020 <i>Peer-reviewed journal that publishes articles on medical informatics and decision-making.</i>
Einfluss	Q1, most influential
Jahr	2020
Zitierungen	1142

# Studie

---

Thematik wird aus unterschiedlichen Perspektiven betrachtet:

Technologisch

Rechtlich

Medizinisch

Patientensicht

Ethisch

# Ergebnisse: Technologisch

---

- “Explainability” ist entweder eine Charakteristik des Algorithmus oder kann durch andere Methoden approximiert werden
- Komplexe Modelle sind weniger transparent als ihre Alternativen, liefern jedoch oft bessere Ergebnisse
- Somit Trade-Off zwischen Erklärbarkeit und Performance
- Explainability kann helfen, das “Clever Hans“- Phänomen zu erkennen
- Trend: In der Entwicklung de facto Vorliebe für komplexe Algorithmen wie ANNs

# Ergebnisse: Rechtlich

---

## Rechtlich:

- Langsame Regulierung
- **Offene Fragen:**
  - Inwieweit Patienten über KI-gestützte Behandlungsentscheidungen informiert werden müssen
  - Haftungsfragen
- **Gefahren:**
  - Zu starke Einschränkung der Technologie
  - Indirekte Verpflichtung KI-Tools zu nutzen

→ Notwendigkeit klarer Regeln und rechtlicher Rahmenbedingungen für KI-gestützte Systeme zu finden

# Ergebnisse: Medizinisch

---

- „Clinical Decision Support Systems" (CDSS) haben das Potenzial, die Diagnose und Behandlung von Krankheiten zu verbessern
- Vorteil CDSS : Individuelle Erklärbarkeit ist möglich, bei traditionellen Tests nicht.
- **Fehlerraten:** KI-Systeme haben insgesamt geringere Fehlerraten als traditionelle Methoden, aber Modelle ohne Fehler oder Bias sind unmöglich.
- Bedeutung der Erklärbarkeit:
  - Besonders relevant bei Abweichungen zwischen CDSS-Empfehlungen und ärztlichen Erwartungen.
  - Entscheidend, um das Vertrauen in KI-gestützte CDSS im klinischen Alltag zu schaffen, da bisher noch fehlend.

# Ergebnisse: Patientenansicht

---

- Patientenzentrierte Versorgung sollte im Vordergrund stehen
  - Fokus auf die Werte und Bedürfnisse der Patienten.
- **Ohne Erklärbarkeit:** Ärzte können den Entscheidungsprozess nicht vollständig erklären, was mit den Werten der patientenzentrierten Versorgung in Konflikt stehen könnte.
- **Erklärbare KI:** Bietet eine visuelle Darstellung oder natürliche Sprach-Erklärung
  - Hilft Ärzten, Entscheidungen besser zu erklären und Patienten ein besseres Verständnis der Risiken und Optionen zu vermitteln.
  - Gibt Auskunft über die Ziele, welches das KI System verfolgt

# Diskussion und Qualität des Papers

---

- **Pro:**
  - Sehr gute Übersicht über die unterschiedlichen Facetten des Bereichs
  - Viele relevante Paper werden als Quellen angeführt
  - Für eine breite Leserschaft aus unterschiedlichen akademischen Feldern verständlich
- **Contra:**
  - Eher oberflächlich, jedoch ein guter Einstiegspunkt
- **Diskussionsfrage:**
  - Sollten genaue, aber nicht erklärbare KI-Systeme in der klinischen Praxis eingesetzt werden?

# Applications of Explainable Artificial Intelligence in Diagnosis and Surgery

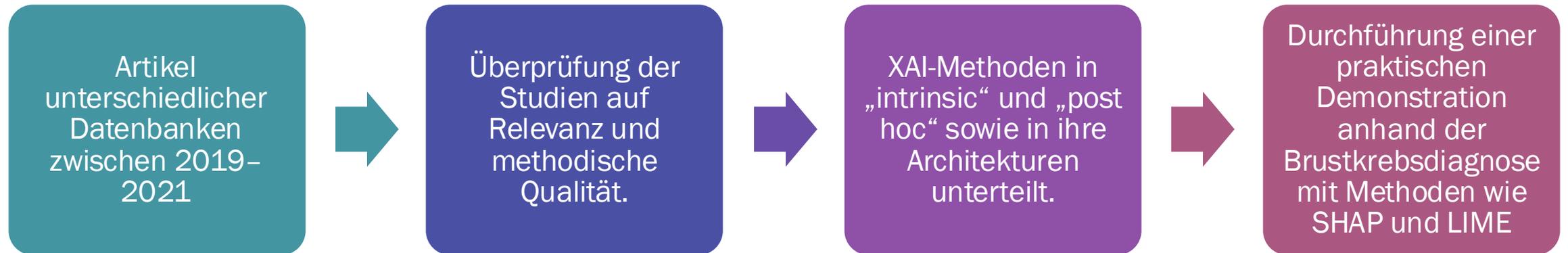
---

Journal	Diagnostics, 2022. <i>Diagnostics an international, peer-reviewed, open access journal focused on medical diagnosis.</i>
Einfluss	Q2, moderate influence
Jahr	2022
Zitierungen	236

# Studie: Ablauf

---

Systematische Literaturübersicht zur Analyse aktueller Trends in der (XAI) für medizinische Anwendungen.

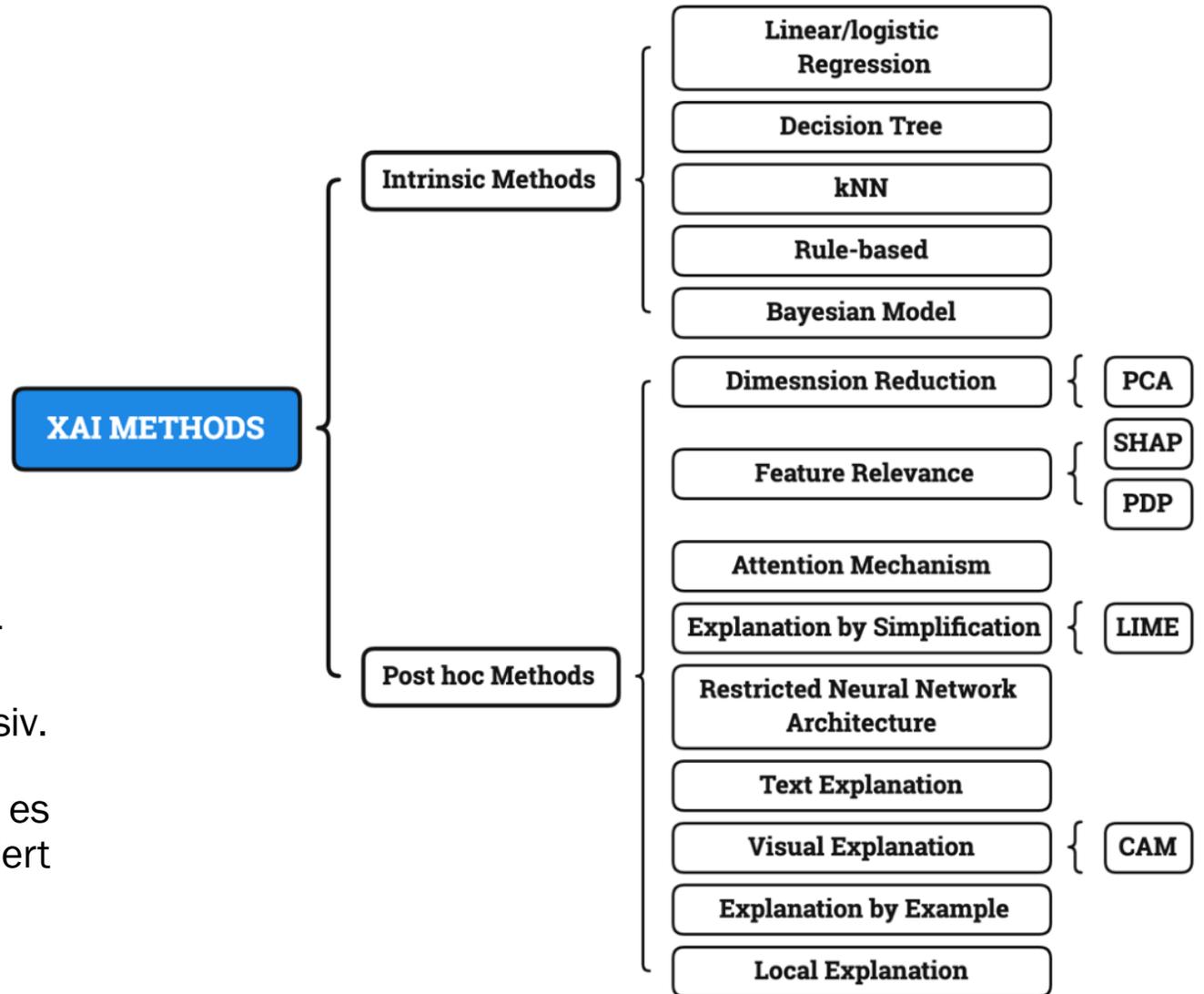


# Studie

**SHAP:** Berechnet die Bedeutung einzelner Merkmale für Vorhersagen mithilfe von Shapley-Werten aus der Spieltheorie. Liefert globale und lokale Erklärungen und ist für komplexe Modelle geeignet. Rechenintensiv.

**LIME:** Vereinfacht einzelne Vorhersagen lokal, indem es ein interpretierbares Ersatzmodell erstellt. Approximiert das Verhalten des Modells lokal, Ideal für schnelle Analysen.

[2]



# Studie

## Verwendete Maße:

- Accuracy
- Precision
- Recall
- F1-Score

## Demonstration:

- Kombination von Random-Forest-Modellen mit SHAP eine hohe Genauigkeit
- 95,9 % bei der Brustkrebsdiagnose
- Vergleich Regelbasiert: 60.81%

## Anwendungen in der Chirurgie

- Bewertung chirurgischer Fähigkeiten
- Instrumentenerkennung in der Laparoskopie

## Anwendungen in der Diagnose

- Allergiediagnose
- Alzheimer-Erkennung
- COVID-19-Erkennung

# Ergebnisse

---

Erklärbare KI (XAI) ist entscheidend für die Integration von KI in die Medizin.

**Wichtigkeit von XAI:** Sie stärkt das Vertrauen in KI-gestützte Medizinsysteme.

**Herausforderungen:** Kleine Datensätze und fehlende einheitliche Bewertungsstandards.

**Zukunftsperspektiven:** Bessere Datennutzung und stärkere Zusammenarbeit mit medizinischen Experten sind essenziell.

# Diskussion und Qualität des Papers

---

- **Pro:**
  - Grobe Übersicht über unterschiedliche Forschung in diesem Bereich
  - Es werden die primären Herausforderungen in dem Bereich aufgezeigt
- **Contra:**
  - **Probleme werden aufgezeigt, jedoch werden keine Lösungsansätze ausgearbeitet**
    - Fehlende interdisziplinärer Zusammenarbeit, jedoch keine konkreten Strategien dafür
    - Fehlenden einheitlichen Bewertungsmaßstab
  - Viele Begriffe, Methoden und Metriken sind nicht oder nicht ausreichend erklärt

# Explainable artificial intelligence (XAI) in deep learning-based medical image analysis

---

Journal	Medical Image Analysis, 2022 In 2022, <i>The journal is published bi-monthly and has been in publication since 1997. It is published by Elsevier</i>
Einfluss	Q1, most influential
Jahr	2022
Zitierungen	705

# Studie: Ablauf

---

**Methodik:** Systematische Literaturrecherche zu XAI-Techniken in der medizinischen Bildanalyse

- Strukturierte Suchstrategie mit SCOPUS (Datenbank für Paper)
- Nutzung spezifischer Keywords zu erklärbarem Deep Learning und medizinischer Bildgebung

**Ansatz:** Identifikation und Kategorisierung der Ansätze anhand eines Rahmens mit drei Kriterien:

- Modellbasiert (intrinsic) vs. Post-hoc-Erklärung
- Modellspezifisch vs. Modellagnostisch
- Global vs. Lokal

**Eingrenzung:**

- Berücksichtigung von Veröffentlichungen bis Oktober 2020
- Fokus auf Studien, die Deep-Learning-basierte XAI-Methoden einsetzen

# Studie: Problemstellung und Hypothese

---

**Problem:** "Black-Box"-Charakter von neuronalen Netzen erschwert Nachvollziehbarkeit

**Hypothese:** XAI-Techniken fördern Verständnis und Vertrauen, was den Einsatz in kritischen medizinischen Entscheidungsprozessen erleichtert

**Risiko:** Ohne XAI können Biases und Fehler unentdeckt bleiben, was potenziell schädliche Folgen im medizinischen Kontext haben könnte

**Kritik:** Mangel an Richtlinien und einem einheitlichen Bewertungsrahmen für XAI

# Ergebnisse

## Schwerpunkt:

- Starke Fokussie

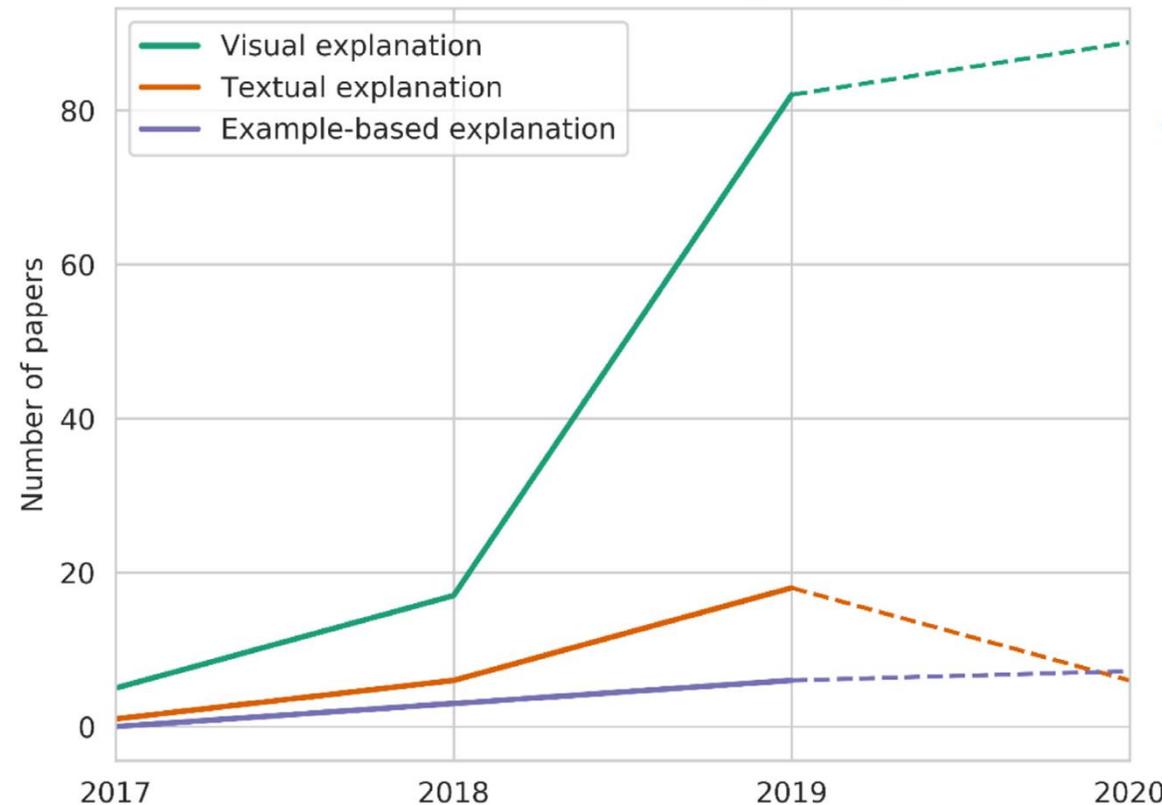
## Erkenntnis:

- Zunehmende In
- Herausforderun  
medizinische Ar

## Empfehlung:

- Sowohl visuelle  
werden, um kei
- Einführung univ  
Evaluierung

XAI used in medical image analysis



alyse  
ür

ickelt

; und

# Diskussion und Qualität des Papers

---

- **Pro:**
  - Begriffe, Methoden und Metriken sind gut erklärt
  - Paper ist sinnvoll gegliedert
  - Diskutiert Lücken in den aktuellen Methoden und schlägt zukünftige Richtungen vor.
  - Paper versucht die Vor- und Nachteile zwischen unterschiedlichen AI-Methoden aufzuschlüsseln
- **Contra:**
  - Primär eine Übersichtsstudie, die keine bahnbrechenden Methoden einführt
- **Diskussionsfrage:**
  - Welche Metriken sind potentiell sinnvoll, um (X)AI-Methoden miteinander zu vergleichen?

# Paper für den Langvortrag

Contents lists available at [ScienceDirect](#)

 **Medical Image Analysis** 

journal homepage: [www.elsevier.com/locate/media](http://www.elsevier.com/locate/media)

---

**Explainable artificial intelligence (XAI) in deep learning-based medical image analysis** 

Bas H.M. van der Velden\*, Hugo J. Kuijf, Kenneth G.A. Gilhuijs, Max A. Viergever

*Image Sciences Institute, University Medical Center Utrecht, Q.02.4.45, P.O. Box 85500, Utrecht, GA 3508, the Netherlands*

---

**ARTICLE INFO** **ABSTRACT**

*Article history:*  
Received 22 July 2021  
Revised 15 March 2022  
Accepted 2 May 2022

With an increase in deep learning-based methods, the call for explainability of such methods grows, especially in high-stakes decision making areas such as medical image analysis. This survey presents an overview of explainable artificial intelligence (XAI) used in deep learning-based medical image analysis. A framework of XAI criteria is introduced to classify deep learning-based medical image analysis meth-

# Quellen: Paper

---

Bas H.M. van der Velden, Hugo J. Kuijf, Kenneth G.A. Gilhuijs, Max A. Viergever,  
Explainable artificial intelligence (XAI) in deep learning-based medical image analysis,  
Medical Image Analysis,  
Volume 79, 2022, 102470, ISSN 1361-8415,  
<https://doi.org/10.1016/j.media.2022.102470>. (<https://www.sciencedirect.com/science/article/pii/S1361841522001177>)

Zhang, Y., Weng, Y., & Lund, J. (2022). Applications of Explainable Artificial Intelligence in Diagnosis and Surgery. *Diagnostics*, 12(2), 237.  
<https://doi.org/10.3390/diagnostics12020237>

Amann, J., Blasimme, A., Vayena, E. et al. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med Inform Decis Mak* 20, 310 (2020). <https://doi.org/10.1186/s12911-020-01332-6>

# Quellen: Andere

---

1. Tigerschiold, T. (2022, 17. November). What is Accuracy, Precision, Recall and F1 Score? Labelf-AI. <https://www.labelf.ai/blog/what-is-accuracy-precision-recall-and-f1-score>
2. LIME vs SHAP: A Comparative Analysis of Interpretability Tools. (o. D.). <https://www.markovml.com/blog/lime-vs-shap>
3. <https://pixabay.com/illustrations/ai-generated-stage-podium-gold-3d-8704010/>