

Aus der Hautklinik und Poliklinik
der Universitätsmedizin der Johannes Gutenberg-Universität Mainz

Einsatz künstlicher Intelligenz als Diagnostikum in der Hautkrebsversorgung aus
Sicht der Patienten und der behandelnden Ärzte

Inauguraldissertation
zur Erlangung des Doktorgrades der
Medizin
der Universitätsmedizin
der Johannes Gutenberg-Universität Mainz

Vorgelegt von

Damian Klaedtke
aus Lahnstein

Mainz, 2025

Wissenschaftlicher Vorstand: Univ.-Prof. Dr. med. Philipp Drees

1. Gutachter: Univ.-Prof. Dr. med. Stephan Grabbe

2. Gutachter: Prof. Dr. med. Carolin Brockmann

Tag der Promotion: 11.11.2025

Nachnutzungslizenz: CC-BY-ND-4.0

Inhaltsverzeichnis

Abkürzungsverzeichnis	I
Abbildungsverzeichnis.....	II
Tabellenverzeichnis	IV
1 Einleitung.....	1
1.1 Einführung in die Thematik	1
1.2 Ziel der Dissertation.....	1
2 Literaturdiskussion.....	2
2.1 Definition Künstliche Intelligenz.....	2
2.2 Entstehungsgeschichte der Künstlichen Intelligenz	2
2.3 Wichtige Begriffe zur Künstlichen Intelligenz.....	3
2.3.1 Maschinelles Lernen.....	3
2.3.2 Überwachtes Lernen (engl.: supervised learning).....	3
2.3.3 Halb-überwachtes Lernen (engl.: semi-supervised learning).....	3
2.3.4 Unüberwachtes Lernen (engl.: unsupervised learning)	3
2.3.5 Verstärktes Lernen (engl.: reinforcement learning)	4
2.3.6 Deep Learning.....	4
2.3.7 Künstliches neuronales Netzwerk.....	4
2.3.8 Convolutional Neural Network	4
2.4 Rechtliche Grundlagen zum Einsatz von KI	5
2.4.1 Schweigepflicht	5
2.4.2 Aufklärungspflicht	5
2.4.3 Digitale-Versorgung-Gesetz	5
2.4.4 Datenschutz	6
2.4.5 Haftungsrecht.....	6
2.4.6 KI-Gesetz: Verordnung des EU-Parlaments.....	7
2.4.7 Medizinprodukteverordnung.....	7
2.5 Ethische Grundlagen zum Einsatz von KI	7
2.5.1 Freiheit/Autonomie.....	7
2.5.2 Menschenwürde und Privatheit	8
2.5.3 Erklärbarkeit.....	8
2.5.4 Schadensverhütung und Wohltätigkeit.....	8
2.5.5 Gerechtigkeit/Fairness	9
2.5.6 Solidarität.....	9
2.5.7 Verantwortung.....	9
2.5.8 Souveränität und Datenschutz.....	10
2.6 Voraussetzungen im Versorgungssystem zum Einsatz von KI	10
2.6.1 Einsatz von KI im privaten Bereich	10
2.6.2 Einsatz von KI im klinischen Bereich	11
2.7 Potential von KI im Rahmen der Hautkrebsdiagnostik	11
2.7.1 Melanom Diagnostik	11

2.7.2	Multiklassifizierende Diagnostik.....	13
2.7.3	KI als Assistenzsystem zur Hautkrebsdiagnostik.....	16
2.7.4	Mobiler Einsatz von KI.....	17
2.7.5	Verbesserungsansätze von KI.....	18
2.7.6	Einsatz von KI.....	19
2.8	KI in der Hautkrebsdiagnostik aus Sicht der Patienten und der Ärzte.....	19
2.8.1	Sicht der Patienten.....	19
2.8.2	Sicht der Ärzte.....	21
3	Material und Methoden.....	21
3.1	Literaturrecherche.....	21
3.2	Studiendesign, Studiendurchführung und Befragungskollektiv.....	22
3.2.1	Studiendesign.....	22
3.2.2	Patientenbefragung.....	22
3.2.3	Ärztebefragung.....	22
3.3	Erstellung der Befragung.....	23
3.3.1	Aufklärung.....	23
3.3.2	Vergleichbarkeit der Befragung zwischen Patienten und Ärzten.....	23
3.3.3	Medium der Befragung.....	24
3.3.4	Validierung der Fragebögen.....	24
3.4	Aufbau der Fragebögen.....	25
3.4.1	Patientenfragebogen (PF).....	25
3.4.2	Arztfragebogen.....	26
3.5	Durchführung der Befragung.....	28
3.5.1	Fallzahlbestimmung.....	28
3.6	Statistische Auswertung.....	28
3.7	Datenschutz.....	29
4	Ergebnisse.....	30
4.1	Ergebnisse der Patientenbefragung.....	30
4.2	Ergebnisse der Ärztebefragung.....	38
4.3	Kombinierte Ergebnisse der Patienten- und Ärztebefragung.....	48
5	Diskussion.....	52
5.1	Einsatz von KI aus Sicht der Patienten.....	52
5.2	Einsatz von KI aus Sicht der Ärzte.....	55
5.3	Vergleich der Ansichten von Patienten und Ärzten.....	60
5.4	Limitationen.....	62
6	Zusammenfassung.....	64
7	Literaturverzeichnis.....	65
8	Anhang.....	70
8.1	Aushang für die Patientenbefragung.....	70
8.2	Aufklärungstext.....	71
8.3	Fragebogen: Patienten.....	73

8.4	Fragebogen: Ärzte.....	78
8.5	Ergebnisse.....	84
9	Danksagung.....	97
10	Tabellarischer Lebenslauf.....	98

Abkürzungsverzeichnis

AF	Arztfragebogen
AUC	Area under the curve
BCC	Basalzellkarzinom (engl.: basal cell carcinoma)
CNN	Convolutional Neural Network
DSGVO	Datenschutz-Grundverordnung
ISIC	International Skin Imaging Collaboration
KI	Künstliche Intelligenz
MBO-Ä	(Muster-) Berufsordnung für die in Deutschland tätigen Ärztinnen und Ärzte
MRT	Magnetresonanztomographie
PF	Patientenfragebogen
ROC	Receiver Operating Characteristic
SCC	Plattenepithelkarzinom (engl.: squamous cell carcinoma)
SD	Standard Deviation (engl. für Standardabweichung)
StGB	Strafgesetzbuch

Abbildungsverzeichnis

Abbildung 1, Seite 4: Beispielhafte Darstellung des grundlegenden Aufbaus eines künstlichen neuronalen Netzwerks (1)

Abbildung 2, Seite 30: Säulendiagramme zur geschlechterspezifischen Altersverteilung der Patienten

Abbildung 3, Seite 31: Säulendiagramme zur geschlechterspezifischen Einschätzung der Patienten in Bezug auf ihr Wissen um die Funktionsweise von KI

Abbildung 4, Seite 32: Säulendiagramm zur Häufigkeit der Mediennutzung der Patienten zur Recherche von Krankheitssymptomen

Abbildung 5, Seite 33: Säulendiagramme zu der Bereitschaft in den Subgruppen "Stimulus" und "Kontrollgruppe" KI einzusetzen

Abbildung 6, Seite 34: Säulendiagramm zu der Bereitschaft der Patienten, KI zur Selbstdiagnostik einzusetzen

Abbildung 7, Seite 35: Säulendiagramm zu der Verunsicherung der Patienten durch den unterstützenden Einsatz von KI beim Arzt

Abbildung 8, Seite 36: Säulendiagramm zu dem Vertrauen von Patienten in KI

Abbildung 9, Seite 36: Säulendiagramm zu dem Einsatz von KI in der Klinik aus Sicht der Patienten

Abbildung 10, Seite 37: Säulendiagramm zu dem Einsatz von KI in der Klinik aus Sicht der Patienten in den Subgruppen "Stimulus" und "Kontrollgruppe"

Abbildung 11, Seite 38: Säulendiagramm zu der Bereitschaft der Patienten, Daten zum Training von KI zur Verfügung zu stellen, in den Subgruppen "Stimulus" und "Kontrollgruppe"

Abbildung 12, Seite 39: Säulendiagramme zu der geschlechtsspezifischen Altersverteilung der Ärzte

Abbildung 13, Seite 41: Säulendiagramme zu der Einschätzung der Ärzte in Bezug auf ihr Wissen um die Funktionsweise von KI

Abbildung 14, Seite 41: Säulendiagramme zu dem Einsatz von KI in der Hautkrebsdiagnostik aus Sicht der Ärzte

Abbildung 15, Seite 42: Säulendiagramm zu der Mediennutzung der Ärzte bei der Recherche von Krankheitssymptomen

Abbildung 16, Seite 43: Säulendiagramm zu der Zustimmung der Ärzte zum selbstständigen Einsatz von KI durch Patienten

Abbildung 17, Seite 44: Säulendiagramm zu der geschlechterspezifischen Bereitschaft der Ärzte KI einzusetzen

Abbildung 18, Seite 44: Säulendiagramme zu der Bereitschaft der Ärzte KI einzusetzen, in Abhängigkeit des Vorwissens, in den Subgruppen "Stimulus" und "Kontrollgruppe"

Abbildung 19, Seite 45: Säulendiagramm zu dem Vertrauen der Ärzte in den stand-alone Einsatz von KI

Abbildung 20, Seite 47: Säulendiagramm zu der Bereitstellung von Daten zum Training von KI in Abhängigkeit von der Aufklärung

Abbildung 21, Seite 49: Säulendiagramme zu der Einschätzung des Wissens von Ärzten und Patienten um die Funktionsweise von KI

Abbildung 22, Seite 50: Säulendiagramme zu der Ansicht von Ärzten und Patienten, im Falle einer gutartigen Diagnose durch die KI den Arzt aufzusuchen

Abbildung 23, Seite 50: Säulendiagramme zu dem Vertrauen von Ärzten und Patienten in die Diagnose einer KI

Abbildung 24, Seite 51: Säulendiagramme zu der Bereitschaft der Ärzte und Patienten Daten zum Training von KI zur Verfügung zu stellen

Tabellenverzeichnis

Tabelle 1, Seite 33: Entdeckung von Hautkrebs bei den Patienten

Tabelle 2, Seite 40: Häufigkeit der Hautkrebsdiagnostik im Alltag der Ärzte

Tabelle 3, Seite 46: Ansicht der Ärzte mit hohem Vorwissen, ob das Wissen um die Funktionsweise als Voraussetzung für deren Einsatz zu betrachten ist

Tabelle 4, Seite 84: Angabe zum höchsten Bildungsabschluss der Patienten

Tabelle 5, Seite 84: Angabe der Patienten, ob sie schonmal etwas von KI gehört haben und ihre Einschätzung, ob sie diese im Alltag einsetzen

Tabelle 6, Seite 84: Einschätzung der Patienten zum eigenen Wissen um die Funktionsweise von KI

Tabelle 7, Seite 85: Anzahl der Smartphonebesitzer, Häufigkeit der Nutzung des Mediums und Häufigkeit des Einsatzes

Tabelle 8, Seite 85: Angabe der Patienten, worauf sie bei der Recherche der Symptome zurückgegriffen haben

Tabelle 9, Seite 86: Anzahl der Patienten mit und ohne Hautkrebs

Tabelle 10, Seite 86: Hautkrebsverteilung bei den Patienten

Tabelle 11, Seite 86: Präferenz der Patienten, auf welche Weise sie KI einsetzen würden

Tabelle 12, Seite 87: Bereitschaft, sich selbstständig mittels KI zu untersuchen

Tabelle 13, Seite 87: Bereitschaft, sich selbstständig mittels KI zu untersuchen, bei hohem Vorwissen

Tabelle 14, Seite 87: Verlangen, den Arzt aufzusuchen, in Abhängigkeit vom Ergebnis der Untersuchung mittels KI

Tabelle 15, Seite 88: Anteil der Patienten, die auf den Einsatz von KI beim Arzt bestehen

Tabelle 16, Seite 88: Vertrauen in die Diagnose einer KI, ohne vom Arzt untersucht worden zu sein

Tabelle 17, Seite 89: Vertrauen der Patienten in die KI oder den Arzt in den Subgruppen Stimulus und Kontrollgruppe

Tabelle 18, Seite 89: Zustimmung der Patienten für den klinischen Einsatz von KI bei hohem Vorwissen

Tabelle 19, Seite 90: Prioritäten der Patienten für den präklinischen Einsatz von KI

Tabelle 20, Seite 90: Bereitschaft KI Verwandten weiterzuempfehlen in den Subgruppen Stimulus und Kontrollgruppe

Tabelle 21, Seite 91: Bedenken der Patienten bezüglich der Datensicherheit beim Einsatz von KI

Tabelle 22, Seite 91: Fachrichtungen und Anstellungsverhältnis der Ärzte

Tabelle 23, Seite 91: Ärztliches Ausbildungsniveau

Tabelle 24, Seite 92: Bekanntheit von KI und Wissen um die Funktionsweise von KI

Tabelle 25, Seite 92: Einsatz von KI im Rahmen der Hautkrebsdiagnostik

Tabelle 26, Seite 92: Mediennutzung der Ärzte zur Information über Krankheitssymptome

Tabelle 27, Seite 92: Patienten den selbstständigen Einsatz von KI empfehlen

Tabelle 28, Seite 93: Empfehlung der Ärzte an die Patienten, einen Arzt aufzusuchen, in unterschiedlichen hypothetischen Situationen

Tabelle 29, Seite 93: Bereitschaft der Ärzte KI bei ihrer Untersuchung als Hilfsmittel einzusetzen

Tabelle 30, Seite 93: Bereitschaft der Ärzte bei hohem Vorwissen KI bei ihrer Untersuchung als Hilfsmittel einzusetzen

Tabelle 31, Seite 94: Vorgehen der Ärzte in verschiedenen hypothetischen Situationen

Tabelle 32, Seite 94: Bereitschaft der Ärzte dem Patientenwunsch nach dem Einsatz von KI nachzukommen

Tabelle 33, Seite 94: Ansicht der Ärzte, ob das Verständnis um die Funktionsweise von KI als Voraussetzung für deren Einsatz gelten sollte

Tabelle 34, Seite 95: Meinung der Ärzte, ob KI Teil ihrer Facharztausbildung sein sollte

Tabelle 35, Seite 95: Einschätzung der Ärzte in Bezug auf das Potential von KI in der Hautkrebsdiagnostik

Tabelle 36, Seite 95: Bereitschaft der Ärzte, die während der Diagnostik anfallenden Daten zum Training von KI bereitzustellen

Tabelle 37, Seite 95: Rating der Ärzte zu wichtigen Eigenschaften von KI

Tabelle 38, Seite 96: Angabe der Ärzte, was ihnen bei der Anwendung einer KI wichtig wäre

1 Einleitung

1.1 Einführung in die Thematik

„Besser als die Hausärzte? Eins zu Null für Dr. Computer“ titelte die Frankfurter Allgemeine Zeitung am 30.05.2018 (2). „KI [Künstliche Intelligenz] in der Medizin: Algorithmus erkennt Hautkrebs präziser als Dermatologen“ heißt es in einem Artikel vom 17.04.2019 in der Berliner-Zeitung (3). „KI bei der medizinischen Bildanalyse – Expertenblick weiter unverzichtbar“ lautet hingegen die Bildüberschrift in der Ausgabe des Ärzteblatts vom 18.02.2021 (4). Solche und ähnliche Artikel sind in den letzten Jahren immer wieder in Magazinen, Zeitungen und Fachzeitschriften veröffentlicht worden, mit auf den ersten Blick zum Teil stark divergierenden Aussagen.

Auch wenn KI keine neue Erfindung ist und sie uns auf vielfältige Weise im Alltag tangiert, so scheint es doch, dass sich die wenigsten Menschen dessen bewusst sind, dass sie KI regelmäßig in ihrem Alltag einsetzen (5, S. 7). Sie findet beispielsweise Einsatz in Form einer Autokorrektur beim Schreiben von Texten, bei Suchanfragen im Internet oder der personalisierten Werbung, welche uns in zahlreichen digitalen Medien begegnet.

KI ist noch kein Pflichtbestandteil der Hautkrebsdiagnostik und findet bislang keinen flächendeckenden Einsatz. Was sind die Gründe dafür? Immerhin erkrankten in Deutschland im Jahr 2018 knapp 23.000 Menschen am Malignem Melanom und etwa 200.000 Menschen an hellem Hautkrebs (6). Gleichzeitig wird seit Jahren auf einen drohenden Ärztemangel hingewiesen (7). Kann KI den daraus resultierenden, zunehmenden Bedarf an Hautkrebsdiagnostik kompensieren?

Folgende Voraussetzungen erscheinen aufgrund aktueller Diskussionen um das Thema notwendig, um KI im Bereich der Hautkrebsdiagnostik in der klinischen Anwendung etablieren zu können:

1. Es bedarf eines rechtlichen Rahmens, um den Einsatz von KI-Systemen zu regulieren.
2. Der Einsatz einer KI wirft ethische Fragen auf, welche diskutiert werden müssen.
3. Das Versorgungssystem muss verschiedene Bedingungen erfüllen, um den Einsatz einer KI realisierbar zu machen.
4. Die KI muss einen positiven Nutzen für die Hautkrebsversorgung haben.
5. Sowohl bei Patienten als auch bei Ärzten muss eine hinreichende Akzeptanz vorliegen, um eine KI einsetzen zu können.

1.2 Ziel der Dissertation

Das Ziel der Dissertation besteht darin, die Sichtweise von Patienten, sowie von Haut- und Hausärzten auf den Einsatz von KI im Bereich der Hautkrebsdiagnostik in Form einer schriftlichen Befragung zu erörtern. Aus dieser Erhebung wiederum soll ersichtlich werden, welche Hürden einem möglichen, flächendeckenden Einsatz von KI im Wege stehen. Dabei werden sowohl der aktuelle Stand der Forschung als auch die ethischen und rechtlichen Voraussetzungen für den Einsatz von KI in der Medizin berücksichtigt.

2 Literaturdiskussion

2.1 Definition Künstliche Intelligenz

Um zu verstehen, was künstliche Intelligenz bedeutet, bietet es sich an, den Turing-Test näher zu betrachten, welcher auch als „The Imitation Game“ bezeichnet wird (8). Dieser Test sieht vor, dass drei Personen schriftlich miteinander kommunizieren, wobei eine Person die Rolle des Interviewers einnimmt. Diese muss durch ihre Fragen herausfinden, welche der beiden anderen Personen (ein Mann und eine Frau) männlich und welche weiblich ist. Dabei versucht eine der beiden Personen den Interviewer in die Irre zu führen, während die andere Person versucht, dem Interviewer zu helfen. Schließlich stellt Turing die Frage, was passiert, wenn die Person, welche den Interviewer in die Irre leiten will, durch einen Computer ersetzt wird (8). Kann der Mensch zwischen dem Menschen und dem Computer unterscheiden?

Wenn man sagt, dass der Mensch intelligent ist und ein Computer vorgeben kann, ein Mensch zu sein, kann man diesen nach Turing als intelligent bezeichnen (8).

Es gibt jedoch eine Vielzahl von Ansätzen, die sich mit der Definition von Künstlicher Intelligenz beschäftigen (9, S. 1-2). Dies liegt nicht zuletzt daran, dass KI ein facettenreiches Anwendungsspektrum bietet, welches sowohl grundlegende Fähigkeiten wie die Lernfähigkeit, aber auch hochspezifische Aufgabengebiete wie das Fahren eines Autos einschließt (9, S. 1).

Die Grundlagen der Künstlichen Intelligenz sind interdisziplinär zu suchen (9, S. 5). Wichtige Teilbereiche sind dabei unter anderem die Philosophie, die Mathematik, die Neurowissenschaft und die Informatik (9, S. 5-15).

In Grundzügen kann zwischen einer schwachen und einer starken KI unterschieden werden (10, S. 20-21). Eine schwache KI kann als eine Art „Werkzeug“ angesehen werden (11). Sie besitzt keinen Verstand im eigentlichen Sinne und kann vom Menschen eingesetzt werden. Beispielsweise um schwierige oder langwierige Aufgaben auszuführen, welche durch den Einsatz von KI bewältigt werden können (12, S. 18). Die schwache KI besitzt die Fähigkeit mit großen Datenmengen effektiv und objektiv umzugehen (12, S. 18).

Starke KI hingegen kann als eine dem Menschen ebenbürtige Intelligenz angesehen werden (10, S. 21). Sie besitzt einen Verstand und kann verschiedene mentale Zustände annehmen (11). Folglich vereint diese die Fähigkeiten des eigenständigen Lernens, Planens und der Problemlösung.

Noch einen Schritt weiter geht die Idee, eine dem Menschen überlegene künstliche Superintelligenz zu schaffen. Dafür gibt es verschiedene Ansätze (13, S. 206-208). Einer dieser Ansätze sieht eine Form von Intelligenz vor, die der eines Menschen entspricht, jedoch weniger Zeit für die Denkprozesse benötigt, die sogenannte „schnelle Superintelligenz“ (13, S. 206-207). Um die dafür benötigte Rechenleistung bereitzustellen, eignet sich der Einsatz von Quantencomputern (13, S. 206-207). Ein weiterer möglicher Ansatz ist die „kollektive Superintelligenz“, welche sich aus einzelnen Komponenten zusammensetzt, die dem Menschen einzeln unterlegen, im Zusammenspiel aber überlegen sind (13, S. 207-208). Dieses Phänomen ist auch in der Natur zu finden und wird dort als Schwarmintelligenz bezeichnet (13, S. 207-208). Schließlich gibt es noch die Idee der „neuen Superintelligenz“ (13, S. 208). Darunter stellt man sich vor, eine KI zu entwickeln, welche vollkommen neuartige Fähigkeiten besitzt, die dem Menschen bislang nicht zugänglich sind (13, S. 208). Doch dies sind bisher alles theoretische Konstrukte (13, S. 211).

2.2 Entstehungsgeschichte der Künstlichen Intelligenz

Ausgehend davon, dass die Meinungen divergieren, wie die Begrifflichkeit KI exakt zu definieren ist, scheint es nicht verwunderlich, dass auch der Ursprung der KI alles andere als genau festzulegen ist. Wo auch immer der Grundstein in der Geschichte gelegt worden ist,

ohne die Entwicklung des ersten programmierbaren Computers in den 40er Jahren des 20. Jahrhunderts, wäre die Existenz einer KI zu der heutigen Zeit kaum denkbar (10, S. 17). 1956, wenige Jahre nach diesem Ereignis, wurde schließlich der Begriff „Artificial Intelligence“ (aus dem Englischen für „Künstliche Intelligenz“) im Rahmen der Dartmouth-Konferenz für ein neues Gebiet der Computertechnologie eingeführt (10, S. 17). Damit war der Grundstein für die KI endgültig gelegt. Jedoch war man zunächst ziemlich enttäuscht, als man vergeblich versuchte Programme zu entwerfen, welche zu universellen Fragestellungen eingesetzt werden sollten (10, S. 17-18). Stattdessen war die Einführung sogenannter wissensbasierter Expertensysteme ein wichtiger Schritt auf dem Weg zur KI (14, S. 4). Diese konnten konkrete Lösungen für spezifische Fragestellungen finden (14, S. 4). Parallel dazu war die Entwicklung künstlicher neuronaler Netze ein weiterer Meilenstein, um die KI voranzutreiben (10, S. 18). Mittlerweile findet KI ein breites Anwendungsgebiet und tangiert uns bei der Nutzung von Diensten der großen Technologiefirmen wie Amazon, Apple, Facebook und Google regelmäßig im Alltag (14, S. 7). Heutzutage spielen vor allem intelligente Techniken wie das Deep Learning eine entscheidende Rolle bei der Entwicklung der KI (10, S. 18).

2.3 Wichtige Begriffe zur Künstlichen Intelligenz

2.3.1 Maschinelles Lernen

Maschinelles Lernen und Künstliche Intelligenz sind eng miteinander verknüpft (15, S. XV). Es bietet die Möglichkeit, mittels Feedbackmechanismen informationsverarbeitende Anwendungen zu trainieren (15, S. 5). Dabei lernt der Computer aus den verarbeiteten Daten (14, S.9) . Das Ziel besteht darin, aus großen Datenmengen Zusammenhänge zu erschließen und diese auf unbekannte Daten anzuwenden(14, S. 12). Das Ergebnis dieses Lernprozesses ermöglicht beispielsweise autonomes Fahren oder auch personalisierte Suchergebnisse bei Google. Einige dieser Ansätze des maschinellen Lernens werden im Folgenden näher definiert:

2.3.2 Überwachtes Lernen (engl.: supervised learning)

Das überwachte Lernen zeichnet sich dadurch aus, dass das Ergebnis für einen präsentierten Stimulus dem Programm bereits vorher bekannt gegeben wird (14, S. 12-13). Über eine Vielzahl ähnlicher Stimuli lernt das Programm Entscheidungsregeln zu finden, welche bei entsprechenden Stimuli zu dem gewünschten Ergebnis führen (14, S. 12-13). Voraussetzung für diese Art des Lernens ist eine große Menge gelabelter Daten (16). Gelabelt bedeutet in diesem Zusammenhang, dass für jeden Stimulus im Trainingsdatensatz ein Ergebnis definiert ist. Labels erhöhen die Qualität von Daten und führen zu besseren Ergebnissen (17). Dadurch wird beispielsweise verhindert, dass der Algorithmus beim Training einem Stimulus ein falsches Ergebnis zuordnet.

2.3.3 Halb-überwachtes Lernen (engl.: semi-supervised learning)

Das halb-überwachte Lernen soll das Problem des überwachten Lernens lösen, dass große Mengen gelabelter Daten benötigt werden. Hierzu werden gelabelte Daten mit nicht gelabelten Daten gemeinsam zum Training verwendet (16).

2.3.4 Unüberwachtes Lernen (engl.: unsupervised learning)

Das unüberwachte Lernen unterscheidet sich insofern vom überwachten Lernen, als dass dem Programm bei dieser Form im Lernprozess kein Ergebnis vorgegeben wird, zu welchem es bei dem jeweiligen Stimulus gelangen soll (5, S. 26). Stattdessen besteht das Ziel darin, dass das Programm selbständig lernt, in den eingegebenen Daten eine Struktur zu erkennen (5, S. 26).

Ein Beispiel hierzu ist das sogenannte Clustering, bei dem es durch eine Klassifikation zu einer Bildung von Gruppen kommt (5, S. 26-27).

2.3.5 Verstärktes Lernen (engl.: reinforcement learning)

Das verstärkte Lernen ist für den Laien das vermutlich „natürlichste“ Beispiel für Maschinelles Lernen. „Verstärkt“ bedeutet in diesem Zusammenhang, dass das Programm für richtige Ergebnisse belohnt wird (5, S. 29). Es lernt also aus Erfahrung. Dabei ist es wichtig, dass das Programm den Anspruch hat, die größte Belohnung zu erhalten und somit das bestmögliche Ergebnis zu erreichen (5, S. 29). Es wird folglich nicht wie bei dem überwachten Lernen die Zielgröße direkt beeinflusst, sondern es findet eine Modulation über die Belohnung statt (5, S. 29).

2.3.6 Deep Learning

Deep Learning ist ein Teilbereich des maschinellen Lernens (14, S. 9). Eine zentrale Rolle in diesem Teilbereich spielen dabei sogenannte künstliche neuronale Netzwerke (14).

2.3.7 Künstliches neuronales Netzwerk

Mit künstlichen neuronalen Netzwerken sind Algorithmen im Bereich des Deep Learning gemeint, welche zwar nicht in ihrem spezifischen Aufbau, jedoch in ihren Eigenschaften dem biologischen Nervensystem nachempfunden sind (5, S. 31). Dazu gehört zum einen die Aufteilung in verschiedene Recheneinheiten, welche im Nervensystem als Neurone bezeichnet werden, und zum anderen die Art und Weise, wie diese einzelnen Recheneinheiten Informationen weitergeben (5, S. 31-32). Dabei spielt sowohl das Prinzip der Gewichtung, also der unterschiedliche informationsbasierte Wert, als auch das Prinzip des Schwellenwerts für die Informationsweitergabe eine zentrale Rolle (5, S. 31-32). Ein solches Netzwerk besteht aus mehreren Schichten (layer), in denen unterschiedliche Anteile der Informationsverarbeitung stattfinden: Über eine oberflächliche Schicht wird der Input, also die zu verarbeitende Information wie etwa ein Bild, in das Netz gespeist wird (5, S. 31-32). Es folgt wenigstens eine Schicht, welche als Hidden-Layer bezeichnet wird (5, S. 31-32). Den Schluss bildet eine weitere oberflächliche Schicht, welche als Output-Layer bezeichnet wird (5, S. 31-32). Diese gibt das Ergebnis der Informationsverarbeitung an (5, S. 31-32) (siehe Abbildung 1).

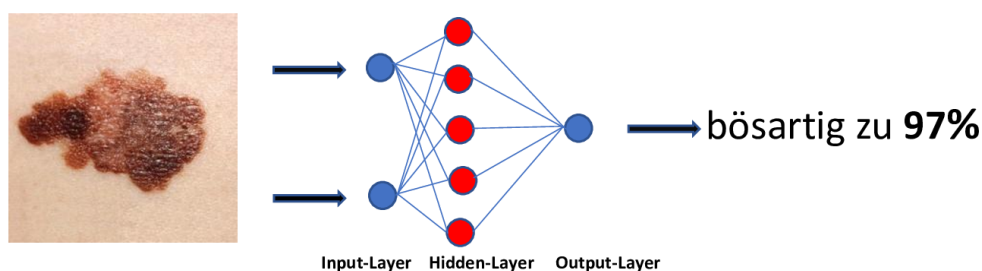


Abbildung 1: Beispielhafte Darstellung des grundlegenden Aufbaus eines künstlichen neuronalen Netzwerks (1)

2.3.8 Convolutional Neural Network

Convolutional Neural Network (CNN) bedeutet so viel wie „faltendes Neuronales Netz“ und beschreibt eine spezielle Art von künstlichen Neuronalen Netzen (5, S. 32-33). Faltungen stehen in diesem Fall für Funktionen, welchen in unterschiedlichen Schichten zur Anwendung kommen können und dem Netz die Fähigkeit geben, Bildinformationen zu abstrahieren (5, S.

32-33). Dies ermöglicht es, dass zum Beispiel Formen oder Grenzen auf Bildern erkannt werden können (5, S. 32-33).

2.4 Rechtliche Grundlagen zum Einsatz von KI

Im folgenden Abschnitt werden mehrere Aspekte aus dem Buch „Digitalisierung im Gesundheitswesen“ von Alexandra Jorzig und Frank Sarangi aufgeführt und auf den Einsatz von KI insbesondere im Bereich der Hautkrebsdiagnostik übertragen.

Bislang gab es keinen eigens für KI geschaffenen Rechtsrahmen, welcher zweifelsfrei und eindeutig den Einsatz von KI auch im Rahmen der Hautkrebsdiagnostik regelte (18, S. 8). Durch die neue Verordnung des EU-Parlaments zum Einsatz Künstlicher Intelligenz, welche am 02.08.2026 endgültig in Kraft treten soll, wird sich dies ändern. Zunächst werden im Folgenden Aspekte in Bezug auf Recht und KI bedacht und es wird erläutert, wie diese beim Einsatz von KI berücksichtigt werden können. Anschließend wird auf die neue Verordnung des EU-Parlaments eingegangen.

2.4.1 Schweigepflicht

Die Schweigepflicht ist in der Berufsordnung für Ärzte verankert (19, S. 4). Eine Entbindung von dieser Pflicht kann in der Regel nur durch den Patienten selbst vollzogen werden (19, S. 4). Dies scheint zunächst auch eine Weitergabe von Daten an Dienstleister wie beispielsweise externe IT-Unternehmen auszuschließen, sofern dieser nicht explizit zugestimmt worden ist. Jedoch trat am 08. November 2017 eine Neuregelung des § 203 StGB in Kraft, welche die Einbindung externer Dienstleister, die ebenfalls zur Geheimhaltung der Patientendaten verpflichtet sind, ermöglicht (20, 21, S. 16).

Da im Rahmen der KI-Anwendung sensible personenbezogene Daten gesammelt werden, ist es wichtig, dass diese vor fremdem Zugriff und Missbrauch geschützt werden. Eine Art, für diesen Schutz zu sorgen, ist die Verpflichtung zum Schweigen. Dies gilt auch im Hinblick auf das Training von KI, zu dessen Zwecke große Datenmengen benötigt werden und in dessen Prozess verschiedene Parteien eingeschlossen sind.

2.4.2 Aufklärungspflicht

Wie im Falle der Schweigepflicht liegt die Aufklärungspflicht für Ärzte in ihrer Berufsordnung begründet (19, S. 4, 21, S. 28). Außerdem ist die Aufklärungspflicht rechtlich im Bürgerlichen Gesetzbuch verankert (21, S.125, 22). Wird dieser nicht im angemessenen Maße nachgekommen, kann dies nach §223 StGB strafrechtlich geahndet werden (21, S. 19, 23).

Der Umfang der Aufklärung hängt unter anderem von der medizinischen Notwendigkeit sowie der Tragweite der Maßnahme ab (19, S. 4). Zwar hat die Anwendung von KI im Rahmen der Hautkrebsdiagnostik keinen direkten Einfluss auf die Unversehrtheit des Patienten, jedoch kann eine von einer KI mitentschiedene Diagnose weitreichende Folgen für den Patienten haben. Aus diesem Grund scheint es naheliegend, den Patienten beim Einsatz einer KI aufzuklären.

2.4.3 Digitale-Versorgung-Gesetz

Das Gesetz zur digitalen Versorgung sieht Krankenkassenleistungen in Form digitaler und telemedizinischer Anwendungen vor (21, S. 41, 24). Hierbei sind in erster Linie Apps zu nennen, welche der Erkennung und Linderung von Krankheiten dienen sollen (25). Eine KI basierte Anwendung zur Früherkennung von Hautkrebs ist in dieser Aufstellung bislang nicht enthalten (26). Das bedeutet, dass bislang keine vom Bundesinstitut für Arzneimittel und

Medizinprodukte gelistete und zertifizierte KI-Anwendung auf Kosten der Krankenkassen vom Arzt verschrieben werden kann.

2.4.4 Datenschutz

Da das Ansammeln von Patientendaten ein wichtiger Aspekt des Trainings, der Anwendung und der fortlaufenden Optimierung von KI ist (5, S. 37), stellt sich die rechtliche Betrachtung des Datenschutzes als äußerst wichtig dar. Nach der DSGVO (Datenschutz-Grundverordnung) gilt es, die Rechtmäßigkeit und Zweckgebundenheit der Datenverarbeitung, die Datenminimierung im Sinne des Meidens der Ansammlung nicht zwingend notwendiger Daten, die Richtigkeit der erhobenen Daten, die Speicherbegrenzung im zeitlichen Sinne und die Integrität und Vertraulichkeit bei der Speicherung der Daten zu berücksichtigen (21, S. 73, 27).

Außerdem stehen den Personen, deren Daten erhoben und verarbeitet worden sind, eine Vielzahl an Rechten wie das Widerspruchsrecht zu der Verarbeitung ihrer Daten und das Recht auf Löschung ihrer erhobenen Daten zu (21, S.138, 28, 29). Es ist jedoch vorstellbar, dass die Umsetzung dieser Rechte zum Teil mit einem außerordentlichen Aufwand verbunden oder gar unmöglich ist. Um diese Problematik im Falle der Datenerhebung zum Training von KI zur Hautkrebsdiagnostik zu umgehen, wäre es denkbar, sich auf eine entsprechende Öffnungsklausel zu berufen, welche der Datenerhebung und -verarbeitung für das öffentliche Interesse oder auch wissenschaftlichen Forschungszwecken ein Sonderrecht eingesteht (21, S. 139, 30). Das Gesetz zur verbesserten Nutzung von Gesundheitsdaten, welches am 22.03.2024 in Kraft getreten ist, soll den Umgang mit Daten insbesondere im Bereich der Forschung regulieren und vereinfachen (31).

Der Koalitionsvertrag der aktuellen Bundesregierung sieht ein Forschungsdatengesetz vor, welches es ermöglichen soll, das Potential erhobener Daten maximal auszuschöpfen. Zugleich sollen wichtige Aspekte wie Sicherheit, Aufbewahrung und Weitergabe von Daten geregelt werden (32).

2.4.5 Haftungsrecht

Bei der Frage des Haftungsrechts geht es um die Klärung, wer die Verantwortung für die Folgen des Einsatzes von KI trägt, vor allem vor dem Hintergrund, dass es dadurch zu einem Schadenseintritt kommen kann. Die aktuelle Gesetzeslage gemäß des Produkthaftungsgesetzes sieht eine Haftungsmöglichkeit eines KI-Systems selbst nicht vor, jedoch ist der Hersteller als sogenanntes Haftungssubjekt in der Verantwortung zu sehen (21, S.127, 33). Der Hersteller ist allerdings unter bestimmten Bedingungen von der Ersatzpflicht ausgenommen. Darunter zählt zum Beispiel der Fall, dass der Produktfehler erst nach dem Vertrieb des Produktes entstanden ist (21, S.127, 34). Versucht man diesen Fall auf KI zu übertragen, wird schnell klar, dass das Gesetz Grenzen in Bezug auf KI hat. Schließlich besitzt KI die Möglichkeit sich weiterzuentwickeln, auch während der Anwendung nach dem Kauf eines KI-Produktes. Es ist anzunehmen, dass es kaum möglich ist, festzustellen, ob ein Fehler im Algorithmus bereits vor oder erst nach dem Kauf des Produktes aufgetreten ist.

Auf EU-Ebene gibt es bereits einen Gesetzesvorschlag zur Haftungsregelung beim Einsatz künstlicher Intelligenz. Dieser soll einerseits mittels spezifische Vorschläge für die zivilrechtliche Haftung bei dem Betrieb einer KI Geschädigten ermöglichen, den ihnen zustehenden Schadensersatz zu erhalten und andererseits wirtschaftliche Anreize setzen, Schadensersatzzahlungen und somit auch Personen- und Sachschäden zu vermeiden (35).

Eine weitere Betrachtungsweise bietet schließlich noch das Bürgerliche Gesetzbuch nach § 831 Absatz 1(21, S. 127). Da die KI der Weisungsbindung des Arztes unterliegt, könnte man mögliche Schadensfälle, auch wenn deren Kausalität kritisch zu hinterfragen ist, auf den Arzt zurückführen (36).

2.4.6 KI-Gesetz: Verordnung des EU-Parlaments

Seit dem 13.06.2024 gilt in der gesamten EU eine Verordnung zum Einsatz von Künstlicher Intelligenz. Die entsprechenden Maßnahmen müssen bis zum 02.08.2026 umgesetzt werden, Verbote sowie allgemeine Bestimmungen gelten jedoch schon ab dem 02.02.2025 (37). Die Verordnung sieht vor, KI in Risikostufen einzuordnen. Entsprechend der Risikostufen unterliegt die KI dann verschiedenen Regulierungen (38). Das EU-Parlament habe den Anspruch, dass die KI sicher, transparent, nichtdiskriminierend und umweltfreundlich sein soll. (39). Es ist davon auszugehen, dass die KI im Rahmen der Hautkrebsdiagnostik unter die Kategorie „Hochrisiko-KI-Systeme“ fallen würde. Ein Begriff, der immer wieder mit den „Hochrisiko-KI-Systemen“ in der Verordnung genannt wird, ist die „Konformitätsbewertung“ (40). Diese dient der Überprüfung, ob ein Produkt genau die Anforderungen erfüllt, welche durch die jeweiligen Rechtsvorschriften vorgegeben sind (41). Zu diesen Anforderungen gehört unter anderem die Einrichtung eines Risikomanagementsystems, wozu regelmäßige, systematische Aktualisierungen des KI-Systems zählen (42). Eine weitere Anforderung ist unter „Daten und Daten-Governance“ zusammengefasst. Diese sieht vor, dass Trainingsdaten und Testdatensätze möglichst fehlerfrei und vollständig sein müssen (43).

2.4.7 Medizinprodukteverordnung

Die Medizinprodukteverordnung der EU sieht eine Einordnung von Medizinprodukten in verschiedene Risikogruppen vor, worunter auch Medizinische Apps fallen (21, S.43, 44). Software, welche auf diagnostische und therapeutische Entscheidungen Einfluss nimmt, gehört mindestens zur Risikogruppe IIa. Kann sie schwerwiegende Verschlechterungen des Gesundheitszustandes einer Person herbeiführen oder chirurgische Eingriffe zur Folge haben, ist sie in die Klasse IIb einzuordnen. Die Klasse III wird über den Tod oder irreversible Verschlechterungen des Gesundheitszustandes einer Person definiert (44).

Für den Einsatz von KI durch Patienten wäre die Einordnung in die Klasse IIa naheliegend, wenn man davon ausgeht, dass erst durch die Konsultation eines Arztes eine endgültige diagnostische und therapeutische Entscheidung getroffen wird. Im Falle eines unterstützenden Einsatzes durch KI während einer ärztlichen Konsultation, bleibt zu diskutieren, ob auch eine Einordnung in die Klassen IIb oder III vorzunehmen wären.

2.5 Ethische Grundlagen zum Einsatz von KI

Der Deutsche Ethikrat veröffentlichte 2018 eine Stellungnahme unter dem Titel „Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung“, in dem wichtige ethische Fragen in Bezug auf den Umgang mit Daten in der zunehmend technisierten Welt der Medizin diskutiert werden (45). Ebenfalls 2018 setzte die europäische Kommission eine Expertengruppe zusammen, um zu entwerfen (46). Einige der in diesen Berichten debattierten ethischen Gesichtspunkte werden im Folgenden näher untersucht und im Hinblick auf den Einsatz Künstlicher Intelligenz im Bereich der Hautkrebsdiagnostik betrachtet.

2.5.1 Freiheit/Autonomie

Freiheit in Bezug auf Daten hat ihre rechtliche Verankerung im Persönlichkeitsrecht unter dem Aspekt der informationellen Selbstbestimmung, in den Zielen der DSGVO und in dem Recht auf Achtung der Privatheit der Europäischen Menschenrechtskonvention (47-49). Daran wird deutlich, welchen Stellenwert die Freiheit in der Gesellschaft einnimmt.

In Bezug auf die praktische Umsetzung des Freiheitskonzeptes vor dem Hintergrund von Big Data rückt das Kaskadenmodell, auch Meta-Einwilligung genannt, in den Mittelpunkt (45, S.184-185). Es bildet einen Kompromiss aus der Blanko-Einwilligung, bei der man all seine

Rechte auf die gespeicherten Daten abgibt, und der dynamischen Einwilligung, bei der man jeweils in Teilprojekte einwilligen kann, wie es schon bei Biobanken der Fall ist (50). Dieses sieht vor, dass die Teilnehmer für die einzelnen Teilprojekte sowohl zu Beginn zustimmen können als auch zwischenzeitlich für spezielle Projekte um ihre Zustimmung gefragt werden können (50). Es ergibt sich folglich ein individuelles Konzept für die Studienteilnehmer.

Freiheit ist jedoch auch in Bezug auf die Barrierefreiheit zu berücksichtigen (46, S. 23). Demzufolge sollte danach gestrebt werden, die KI möglichst allen Mitgliedern der Gesellschaft zugänglich zu machen (46, S. 23). Diese Überlegung in Bezug auf die Einwilligung könnte auch auf den Einsatz von KI im Rahmen der Hautkrebsdiagnostik übertragen werden, um die Entscheidungsfreiheit in Bezug auf die persönlichen Daten der Patienten im höchstmöglichen Maß zu wahren. Zugleich sollte KI möglichst einfach für alle Menschen zugänglich gemacht werden.

2.5.2 Menschenwürde und Privatheit

Die Bewahrung der Privatsphäre stellt im digitalen Zeitalter eine zunehmende Herausforderung dar. Hier sei als Beispiel die Möglichkeit genannt, aufgrund von ursprünglich getrennten, öffentlich verfügbare Daten auf Merkmale oder Eigenschaften Dritter zurückzuschließen (45, S. 199).

Als weiterer Punkt sei hier die Löschung von Daten im Internet zu nennen. Wer erstmal Informationen über sich in der digitalen Welt preisgegeben hat, hat kaum noch die Chance, diese wieder zurückzunehmen (21, S. 138).

Auch ist die Möglichkeit der Deanonymisierung von Daten vor dem Aspekt der Privatheit zu bedenken (45, S. 87), ebenso wie die Tatsache, dass Zusammenhangsanalysen von Massendaten dazu führen können, dass Menschen, zumeist unabsichtlich, in Gruppen gefasst werden, was schnell in einer Stigmatisierung enden kann (45, S. 198).

Beim Einsatz von KI darf zudem die Würde des Menschen nicht außer Acht gelassen werden. Die Würde des Menschen darf nicht unter dem Einsatz von KI leiden (46, S. 13).

Da das Prinzip der KI auf der Verarbeitung großer Datenmengen basiert, nimmt die Privatheit an dieser Stelle eine wichtige Rolle ein. Zur Sicherung der Privatheit und somit auch der Würde des Patienten trägt unter anderem die Neuregelung des § 203 StGB bei (20). Privatheit ist an dieser Stelle als Recht auf die informationelle Selbstbestimmung zu sehen.

2.5.3 Erklärbarkeit

KI sollte bis zu einem gewissen, situationsabhängigen Maß erklärbar sein (46, S. 16). Erklärbarkeit bildet nach den Ethikleitlinien der EU die Basis für das Vertrauen in die Anwendung einer KI (46, S. 16). Da Transparenz im Sinne der Erklärbarkeit nicht immer vollumfänglich vorhanden ist, dienen Maßnahmen wie z. B. die Rückverfolgbarkeit oder die Nachprüfbarkeit einer Entscheidung, um der Erklärbarkeit gerecht zu werden (46, S. 16).

Prozesse, bei denen es keine genaue Erklärung gibt, wie diese zu ihrem Ergebnis gelangt sind, werden in diesem Zusammenhang als „Blackbox“-Algorithmus bezeichnet (46, S. 16).

In Bezug auf den Einsatz von KI im Rahmen der Hautkrebsdiagnostik ist folglich zu erwarten, dass die Akzeptanz für KI wächst, wenn die anwendende Person versteht, wie die KI zu ihrer Diagnose gelangt. Diese Überlegung betrifft sowohl Patienten als auch Ärzte.

2.5.4 Schadensverhütung und Wohltätigkeit

Schadensverhütung und Wohltätigkeit sind ebenfalls ein wichtiger Bestandteil bei der ethischen Betrachtung von KI (46, S. 15). In diesem Kontext rückt zum einen der Aspekt des Wissens- und Erkenntniszuwachses und zum anderen der therapeutische Mehrwert in den

Mittelpunkt (45, S. 215). Beiden, der auf der Gewinnung und Verarbeitung gesundheitsbezogener Daten beruhenden Aspekten, wohne ein Mehrwert inne, welcher sowohl für das Individuum als auch für die Gesellschaft im Ganzen von Bedeutung ist (45, S. 215). Voraussetzung dafür ist, dass die KI in ihrer Anwendung robust ist (46, S. 8-9).

In Bezug auf die Hautkrebsdiagnostik bedeutet Wohltätigkeit beispielsweise eine Zunahme der Lebensqualität oder der Überlebenszeit. Dabei soll die Leistung einer KI möglichst unabhängig von der jeweiligen Umgebung sein, um Fehler zu vermeiden und Schaden zu verhüten (46, S. 8-9).

Doch auch die Telemedizin schafft neue Möglichkeiten. So werden Arztkonsultationen auch über räumliche Distanz ermöglicht (21, S. 170). Im Rahmen der Telemedizin kann KI zur Hautkrebsdiagnostik zum Einsatz kommen und einen möglichen Weg zum Arzt ersparen oder im Sinne eines Triage-Systems zu einer schnelleren Vorstellung führen, wenn das KI-System eine Gefahr erkennt.

2.5.5 Gerechtigkeit/Fairness

Gerechtigkeit in Bezug auf KI lässt sich auf mehreren Ebenen betrachten. Zunächst einmal sind Daten von außerordentlicher Wichtigkeit für die Entwicklung einer KI (21, S. 111). Dabei gilt es ein Gleichgewicht zwischen der Datensicherheit in Form von Zugangsbeschränkungen zu Datenbanken und der freien Datennutzung für Entwickler, welchen durch den Zugriff auf Datenbanken eine unnötige Datenerhebung erspart bleibt, zu finden (45, S. 103, 223). Eine Lösung für diese Problematik könnte z. B. eine Publikationspflicht sein (45, S. 223). Wenn die zu Forschungszwecken erhobenen Daten publiziert werden müssen, jedoch in ihrem Zugang bestimmten Beschränkungen unterliegen, könnte die Datensicherheit bewahrt werden und zugleich die unnötige Erhebung bereits gesammelter Daten verhindert werden.

Gerechtigkeit bedeutet auch, einen freien Zugang für alle Menschen zu KI zu schaffen (46, S. 13). Die barrierefreie Verfügbarkeit von KI soll dafür sorgen, dass alle Menschen die gleichen Chancen haben, KI einzusetzen (46, S. 23).

2.5.6 Solidarität

Das Sammeln immer spezifischerer Daten ermöglicht eine zunehmende Risikostratifizierung einzelner Individuen (45, S.230). Dass diese Risikostratifizierung nicht zum Nachteil einzelner ausgenutzt werden sollte, ist aus solidarischer Sicht selbsterklärend. Im Gegenzug besteht sogar die Möglichkeit, durch die intensivierten Datenerhebungen Informationen von und über seltene Erkrankungen zusammenzuführen und somit die Forschung in dieser Richtung voranzutreiben (45, S. 237-238).

Auch der Einsatz von KI im Rahmen der Hautkrebsdiagnostik ist mit einer vermehrten Datenansammlung verknüpft. Im Sinne der Solidarität ist folglich ein verantwortungsvoller Umgang mit der Technik gefordert. Im Umkehrschluss kann der Einsatz von KI helfen, seltene Formen von Hautkrebs früh und vermehrt zu erkennen und somit langfristig die Versorgung in diesem Bereich zu optimieren.

2.5.7 Verantwortung

In Bezug auf Big Data und somit auch indirekt auf KI, gilt es eine Multiakteurverantwortung zu berücksichtigen (45, S. 249). Einer dieser Akteure stellt das Individuum dar. Übertragen auf den Einsatz von KI lässt sich die Verantwortung des Individuums im Umgang mit seinen eigenen Daten nach der Vorstellung des Ethikrats wie folgt verstehen: Die KI sammelt die Daten des Patienten, speichert diese und legt diese möglicherweise zum Training weiterer KI in Datenbanken ab. Diesem Prozess muss der Patient jeweils zustimmen (45, S. 249). Das

Individuum entscheidet also selbst über die Verwendung der persönlichen Daten und trägt die Verantwortung für seine Entscheidung.

Doch auch die Hersteller von KI stehen in der Verantwortung (46, S. 18-19). So ist die Transparenz und Rückverfolgbarkeit von KI-Prozessen wichtig, um mögliche Fehlentscheidungen der KI nachvollziehen zu können (46, S. 22). Eine mögliche Maßnahme in diesem Zusammenhang wäre, dass Unternehmen Verfahren ihrer KI-Systeme prüfbar machen (46, S. 24).

Im Rahmen der Hautkrebsdiagnostik kann Transparenz zudem zur Qualitätssicherung beitragen, wenn zum Beispiel offensichtlich fehlerhafte Prozesse in der Entscheidungsfindung einer KI erkannt werden.

Auch eine Aufsichtsfunktion von Staaten für Gütesiegel ist eine denkbare Maßnahme, um so einen regulatorischen Einfluss auszuüben (45, S. 248).

Überträgt man diese Überlegungen auf den Einsatz von KI in der Hautkrebsdiagnostik, so scheint aus Sicht der privaten Nutzung einer KI mittels einer App die Einführung eines Gütesiegels eine sinnvolle Maßnahme. Dadurch würde es der Privatperson vereinfacht werden, vertrauenswürdige und qualitätsgesicherte Anwendungen zu finden und einzusetzen.

2.5.8 Souveränität und Datenschutz

In Bezug auf die Digitalisierung spielt vor allem der Begriff Datensouveränität im Sinne einer informationellen Freiheitsgestaltung eine entscheidende Rolle (45, S. 251). Ein mögliches Mittel zur Umsetzung dieser Souveränität bietet das bereits erwähnte Kaskadenmodell der Zustimmung. Diese Überlegung ist vor allem in Bezug auf das Training von KI wichtig.

Der Datenschutz selbst wird unter anderem durch die DSGVO gefordert (49). Diese sieht zunächst ein grundlegendes Verbot der Bearbeitung personenbezogener Daten vor, welches unter anderem durch die Zustimmung der Person aufgehoben werden kann (51). Es gilt die dem Anwender einzugestehende, weitestmögliche Entscheidungsfreiheit in Bezug auf die Verwendung seiner eigenen Daten zu gewährleisten.

2.6 Voraussetzungen im Versorgungssystem zum Einsatz von KI

2.6.1 Einsatz von KI im privaten Bereich

Dass KI zur Hautkrebsdiagnostik von häufigen Tumorentitäten wie dem Malignen Melanom, dem Basalzellkarzinom (BCC, aus dem engl. für basal cell carcinoma) und dem Plattenepithelkarzinom (SCC, aus dem engl. für squamous cell carcinoma) beispielsweise in Form von Apps eingesetzt werden kann, hat man bereits mehrfach zeigen können (52-54).

Die notwendige Ausgangsbedingung für einen privaten und mobilen Einsatz von KI liegt vor. Knapp die Hälfte der Bevölkerung weltweit besaß im Jahr 2021 ein Smartphone und die Prognose bis zum Jahr 2024 verspricht weiterhin eine deutlichen Zunahme (55). In Deutschland betrug der Anteil der Smartphone-Nutzer im Jahr 2021 rund 89% (56). Für den kleinen Anteil der Bevölkerung, welcher nicht in Besitz solch einen mobilen Endgeräts ist oder dieses zu diesem Zweck nicht einsetzten wollte, wäre auch die Entwicklung eines eigenständigen Endgeräts, welches allein für die Hautkrebsdiagnostik einsetzbar wäre, denkbar.

Wie sich der Versorgungsprozess nach der selbständigen Diagnostik durch den Patienten gestaltet, bietet Raum für Diskussionen.

Es liegt auf jeden Fall nahe, den Patienten gut an das Hautkrebsversorgungssystem anzugliedern, beispielsweise mittels einer automatischen Befundübermittlung an den eigenen Hautarzt oder an eine nahegelegene Klinik. Auch eine beschleunigte Terminvermittlung bei

suspekten Befunden wäre denkbar. Klar ist, dass der Patient mit seiner Diagnose nicht allein gelassen werden sollte. Eine anschließende telemedizinische Befundbesprechung wäre ebenfalls in Betracht zu ziehen und im Rahmen des Beschlusses der Bundesärztekammer aus dem Jahr 2018 mit der MBO-Ä ((Muster-) Berufsordnung für die in Deutschland tätigen Ärztinnen und Ärzte) vereinbar (19, S. 3).

2.6.2 Einsatz von KI im klinischen Bereich

Auch im klinischen Setting ist ein mobiler Einsatz von KI vorstellbar (53, 54). Vor allem im Hinblick auf den flexiblen Einsatz wäre die mobile KI einer festen Installation, welche räumlich fixiert wäre, vorzuziehen. Dies würden auch einen unkomplizierten Einsatz in Ambulanzen ermöglichen. Dann könnte der diensthabende Arzt die KI immer mit sich führen und hätte sie jederzeit zur Verfügung. Auch der konsiliarische Einsatz durch die Mitnahme der KI auf fachfremde Stationen wäre auf diese Weise problemlos möglich.

In Bezug auf die Befunddokumentation wären mehrere Szenarien in Betracht zu ziehen. Zum einen könnte das Ergebnis der KI händisch in der digitalen und/oder der analogen Patientenakte dokumentiert werden. Andererseits wäre auch die Kommunikation mit dem klinikinternen Datenverwaltungssystem denkbar. Auf diese Weise könnte man beispielsweise Befunde retrospektiv analysieren und die Diagnostik dadurch auch objektiver gestalten. Zudem wäre eine Verlaufsanalyse möglich, wenn identische Hautläsionen zu verschiedenen Zeitpunkten analysiert würden.

Diese theoretischen Überlegungen werfen die Frage auf, warum KI auf diese Weise bislang keine Verwendung findet.

2.7 Potential von KI im Rahmen der Hautkrebsdiagnostik

Zum Zeitpunkt der Erstellung der vorliegenden Arbeit lassen sich die programmiertechnischen Ansätze in die groben Kategorien „Melanom Diagnostik“ und „Multiklassifizierende Diagnostik“ unterteilen.

2.7.1 Melanom Diagnostik

In den vergangenen Jahren sind viele verschiedene auf KI basierende Ansätze getestet worden, welche sich mit der Melanomdiagnostik im Bereich der Hautkrebsuntersuchung auseinandersetzen. So ist zum Beispiel in einer Studie von Haenssle et al. das CNN Inception v4 von Google mit 58 Dermatologen aus 17 verschiedenen Ländern verglichen worden (57). Nachdem das CNN u.a. mit Bildmaterial von der International Skin Imaging Collaboration (ISIC) trainiert worden ist, ist dieses mit einem Testset, bestehend aus dermatoskopischen Aufnahmen von Melanomen und Nävi unterschiedlicher Hautregionen und Subtypen gegen die Dermatologen, von denen 30 Personen eine Erfahrung von über 5 Jahren im Bereich der Dermatoskopie besaßen, angetreten. Die primären Endpunkte wurden durch die Sensitivität, Spezifität und die AUC (area under the curve) der Grenzwertoptimierungskurve (ROC für receiver operating characteristic) abgebildet. Es zeigte sich, dass die Spezifität des CNN signifikant ($p < 0,01$) größer war als die mittlere Spezifität der Dermatologen (82,5% vs. 71,5%) bei gleicher Sensitivität von 86,6%). In einem zweiten Schritt haben die Dermatologen zusätzlich klinische Informationen zu den Patienten sowie weitere Nahaufnahmen der Läsionen bekommen. In diesem Fall stieg die mittlere Sensitivität der Dermatologen auf 88,9% und die mittlere Spezifität auf 79,7%. Dennoch zeigte sich diese Leistung der Performance der CNN signifikant ($p < 0,01$) unterlegen. Insgesamt lag der Wert für die ROC AUC der CNN bei 0,86 und war somit signifikant ($p < 0,01$) höher als der der Dermatologen (0,79). Die Autoren folgern daraus, dass der Einsatz von KI im Bereich der Melanomdiagnostik hilfreich ist, unabhängig von der Erfahrung des Dermatologen. Sie merken jedoch an, dass die Leistung von 22,4% der Dermatologen die des CNN übertraf. Zudem gibt es weitere

Differentialdiagnosen, welche nicht in das Training der KI eingeschlossen worden sind, wie beispielsweise die Seborrhoische Keratose und pigmentierte BCC. Diese können von der KI folglich nicht erkannt werden.

Eine andere Studie hat sich mit der Optimierung von KI durch Fusion von Algorithmen beschäftigt (58). Auch hier wurde auf dermatoskopische Aufnahmen aus dem Datenpool der ISIC zurückgegriffen. Insgesamt wurden 25 verschiedenen Deep Learning Algorithmen und verschiedenen Fusionen aus diesen Algorithmen (die beste Fusion schloss 16 Algorithmen ein) getestet und mit der Leistung von 8 Dermatologen verglichen, welche eine durchschnittliche Erfahrung von 13,5 Jahren im Bereich der Dermatoskopie angaben und sich auf Hautkrebs spezialisiert haben. Die Aufgabe bestand darin, 100 Läsionen zu klassifizieren (benigne vs. maligne) und ein entsprechendes Management zu empfehlen (Biopsie vs. beobachten/beruhigen). Für die binäre Klassifizierung haben die Dermatologen bei einer Sensitivität von 82% im Mittel eine Spezifität von 59% erreicht, welche mit der Spezifität des besten Algorithmus vergleichbar war (62%, $p=0,68$). Die beste Fusion hingegen erreichte eine Spezifität von 76%, welche signifikant höher ($p<0,02$) war als die mittlere Spezifität der Dermatologen. Auch die AUC der ROC der besten Fusion unterschied sich signifikant von der AUC der ROC der Dermatologen (0,86 vs. 0,71, $p<0,001$). Ebenso zeigte sich, dass die Spezifität der besten Fusion bei einer Sensitivität von 89% signifikant höher war als die mittlere Spezifität der Dermatologen (64% vs. 47%, $p<0,02$). Es ist jedoch anzumerken, dass ein paar Dermatologen in der Diagnostik und/oder im Management bessere Ergebnisse als die beste Fusion erzielten. Zudem gilt es zu berücksichtigen, dass die Stichprobengröße von 8 Dermatologen gering ist. Außerdem hat die Studie von Haenssle et al. gezeigt, dass zusätzliche klinische Informationen die Diagnosegenauigkeit von Dermatologen verbessern können (57).

Eine weitere Arbeitsgruppe hat sich mit dem Einsatz von KI zur Fotoanalyse von Hautläsionen beschäftigt (59). In diesem Fall ist die KI mit dermatoskopisch aufgenommen Bildern trainiert und mit 100 Bildern von Läsionen getestet worden. Hierzu wurde auf Bilder der ISIC sowie des HAM10000 Datensatzes zurückgegriffen, in welchem pigmentierte Läsionen aus verschiedenen Populationen hinterlegt sind. Bei einer Sensitivität von 89,4% zeigten die Dermatologen im Durchschnitt eine Spezifität von 64,4% (Range: 22.5%–92.5%, welche mit der Spezifität der trainierten KI vergleichbar war (68,2%, Range: 47.5%–86.25%). Den Autoren zufolge deutet die geringere Varianz der Ergebnisse durch die KI auf eine höhere Robustheit eben jener hin. Zudem merken sie an, dass das Trainingsmaterial überwiegend von helleren Hauttypen stammte, sodass davon auszugehen ist, dass die KI bei dunkleren Hauttypen schlechter abschneidet. Außerdem standen auch in diesem Fall den Dermatologen keine klinischen Informationen zur Verfügung. Die Autoren merken an, dass 13,1% der Dermatologen eine höhere Sensitivität erzielten als die KI. Dennoch hat sich gezeigt, dass es möglich ist, eine KI für Fotos von Hautläsionen einzusetzen, obwohl diese nur mit dermatoskopischen Aufnahmen trainiert worden ist.

Eine weitere Studie zeigt, welche Leistung eine KI, die mit dermatoskopisch aufgenommen Open Source Bilddateien trainiert worden ist, im Vergleich zu 157 Dermatologen, von denen die meisten an deutschen Universitätskliniken arbeiteten, erzielen konnte (60). In diesem Fall wurde ebenso auf Bildmaterial der ISIC zurückgegriffen. Es wurden 100 Läsionen, bestehend aus Melanomen und atypischen Nävi, zum Testen herangezogen. Bei einer Sensitivität von 74,1% erreichten die Dermatologen im Mittel eine Spezifität von 60%. Die Spezifität der KI lag in diesem Fall bei 86,5%. Auch wenn dieses Ergebnis keine statistische Signifikanz aufwies ($p=0,31$), stuften die Autoren dieses als relevant ein. Bei einer Spezifität von 60% erreichte die KI eine Sensitivität von 87,5%. Insgesamt war die KI in Bezug auf die Sensitivität und die Spezifität 136 der 157 Dermatologen überlegen. Jedoch ist anzumerken, dass auch in dieser Studie den Dermatologen keine klinischen Informationen zur Verfügung gestellt worden sind.

Schließlich wurden in einer anderen Studie, bei welcher eine KI ebenfalls mit Bildmaterial der ISIC trainiert worden und gegen 144 Dermatologen angetreten ist, nur Biopsie gesicherte Läsionen im Testset verwendet (61). Dieses bestand aus 804 Läsionen, von denen die Hälfte

Melanome und die andere Hälfte Nävi waren. Diese wurden dann 1:1 auf 6 Fragebögen mit jeweils 134 Läsionen verteilt. Insgesamt sind 11,1% der Läsionen ausgeschlossen worden, da diese von den Befragten in ihrer Bildqualität als unzureichend eingestuft worden sind. Bei den übrigen Läsionen galt es zu unterscheiden, ob diese therapiebedürftig seien bzw. biopsiert werden sollten oder ob man den Patienten beruhigen konnte, da kein Handlungsbedarf bestehe. Die Dermatologen erzielten im Durchschnitt eine Sensitivität von 67,2% und eine Spezifität von 62,2%. Die KI erreichte eine Sensitivität von 82,3% und eine Spezifität von 77,9%. Im McNemar's Test zeigte sich eine statistisch signifikante Überlegenheit der KI gegenüber den getesteten Dermatologen mit $p < 0,001$. Wie in den zuvor diskutierten Studien, sind auch hier keine zusätzlichen klinischen Informationen in die Diagnostik eingeflossen. Zudem muss darauf hingewiesen werden, dass alle Nävi des Testsets biopsiert worden sind, da sie klinisch als suspekt eingestuft wurden. Sämtliche Läsionen waren daher schwer zu unterscheiden. Damit erklären sich die Autoren das im Vergleich zu vergangenen Studien schlechtere Abschneiden der Dermatologen.

Zusammenfassend lässt sich festhalten, dass die verschiedenen KI in den dargestellten Studien dem zum Vergleich herangezogenen Ärztekollektiv in Bezug auf die Sensitivität und Spezifität der Diagnose meist überlegen war, häufig auch signifikant. Dies war in einer Studie auch dann der Fall, als den Dermatologen zusätzliche klinische Informationen zur Verfügung standen. In einer Studie konnte zudem gezeigt werden, dass die Leistung der KI durch Fusion mehrerer KI gesteigert werden konnte. All diesen Studien ist gleich, dass diese in einem künstlichen Umfeld durchgeführt worden sind, welches nicht dem klinischen Alltag und einem realistischen Einsatz der getesteten KI entspricht. Zudem ist die Klassifikation von Melanomen aufgrund fehlender Integration zahlreicher Differentialdiagnosen für den ärztlichen Alltag nur bedingt geeignet (61).

2.7.2 Multiklassifizierende Diagnostik

Da die Inzidenz des nicht melanozytären Hautkrebs die des melanozytären Hautkrebs in Deutschland übersteigt, scheint es naheliegend, die KI gestützte Hautkrebsdiagnostik nicht nur auf die Erkennung von Melanomen zu beschränken. Laut des Zentrums für Krebsregisterdaten erkrankten im Jahr 2019 etwa 22.000 Menschen in Deutschland an einem malignen Melanom, 2018 lag die Inzidenz für nicht melanozytären Hautkrebs in Deutschland bei 200.000 (62, 63). In verschiedenen Studien sind nicht melanozytäre Hautkrebsformen in die Testung von KI eingeschlossen worden, welche im Folgenden vorgestellt werden. Der Begriff Multiklassifikation bezieht sich in diesem Zusammenhang auf die Unterscheidung von mehr als 2 Differentialdiagnosen.

Tschandl et al. haben in einer Studie eine kombinierte KI für dermatoskopische Aufnahmen und Nahaufnahmen von Läsionen getestet (64). Dabei sind folgende Läsionen eingeschlossen worden: Aktinische Keratose, Morbus Bowen, BCC in verschiedenen Subtypen, Lentigo solaris, Seborrhoische Keratose, Lichen planus, Dermatofibrome, Melanome, SCC, Keratoakanthome, Nävus sebaceus und verwandte gutartige Läsionen, benigne Haarfollikeltumore. Das Datenmaterial stammte aus unterschiedlichen Quellen, u. a. aus einer Hautkrebsklinik in Queensland und der Universitätsklinik in Wien. Die KI wurde in Bezug auf die Unterscheidung benigne vs. maligne, der spezifischen Diagnose und dem Management der Läsionen mit 95 Dermatologen unterschiedlicher Berufserfahrung verglichen. Es mussten jeweils 50 Läsionen bewertet werden. Die AUC der KI war mit 0,74 signifikant höher als die der Dermatologen mit 0,70 ($p < 0,001$). Bei einer Spezifität von 51,3% lag die Sensitivität der KI bei 80,5%. Die durchschnittliche Sensitivität der Dermatologen betrug 77,6%. Die KI benannte in 37,6% der Fälle die spezifische Diagnose richtig, womit sie signifikant besser abschnitt als die Dermatologen (33,5%, $p = 0,001$). Unter den Dermatologen mit einer Berufserfahrung von mehr als 10 Jahren lag der Anteil der richtigen Diagnosen hingegen bei vergleichbaren 37,0%. In Bezug auf das Erkennen amelanotischer Melanome zeigte sich die KI mit einer Sensitivität von 52,3% den Dermatologen mit einer durchschnittlichen Sensitivität von 69,3% unterlegen. Als Ursache führen die Autoren in diesem Zusammenhang die geringe Anzahl amelanotischer

Melanome im Trainingsdatenset an. Die Subgruppenanalysen zeigen, dass die KI bei der spezifischen Erkennung maligner Läsionen den getesteten Dermatologen überlegen war (55,5% vs. 42,2%, $p < 0,001$). Für benigne Läsionen war der Anteil korrekt klassifizierter Läsionen bei den Dermatologen signifikant höher als bei der KI (18,1% vs. 23,4%, $p = 0,001$). Es hat sich gezeigt, dass die KI, welche die Nahaufnahmen analysiert hat, besser bei benignen Läsionen abschnitt, während die KI, welche dermatoskopische Aufnahmen analysiert hat, die malignen Läsionen besser erkannt hat. In diesem Fall standen den Dermatologen keine klinischen Informationen zur Verfügung.

In einer anderen Studie ist eine KI mit Bildern von Läsionen sowie zusätzlichen Informationen zu den Patienten aus 17 verschiedenen teledermatologischen Diensten aus den USA trainiert worden (65). Diese kann in erster Linie 26 verschiedene Hautläsionen unterscheiden, darunter verschiedenen Dermatitisen, Dermatosen, Psoriasis, Melanome und Alopezien, welche laut der Autoren dieser Studie 80% der Fälle in der Primärversorgung ausmachten. Zudem können zusätzlich Vorhersagen unter Berücksichtigung von insgesamt 419 verschiedenen Hautläsionen getroffen werden. Das Besondere ist, dass von der KI außer Bilddaten auch weitere 45 Variablen berücksichtigt werden, u.a. die Krankengeschichte sowie die demographische Situation. Außerdem kann die KI eine beliebige Anzahl an Bildern zu einem Fall bearbeiten. Die Top-1-Genauigkeit (Wahrscheinlichkeit, die richtige Diagnose getroffen zu haben) betrug im Testset B für die KI 66%. Dagegen wurden 6 Dermatologen getestet, welche im Durchschnitt auf 63% kamen und weitere 6 Hausärzte, welche auf 44% kamen. Zudem wurden auch 6 „nurse practitioners“ getestet. Diese kamen auf 40%. Dieser Berufszweig lässt sich von der medizinischen Ausbildung her zwischen dem Medizinstudium und der Ausbildung zur Krankenpflege verordnen. Zudem wurde auch die Top-3 Genauigkeit getestet. Diese gibt die Wahrscheinlichkeit an, dass die gesuchte Hautläsion in einer Auswahl von 3 Differentialdiagnosen war, welche von der KI und den Testpersonen angegeben worden sind. Die Top-3-Genauigkeit im Testset B lag für die KI bei 90%. Die Dermatologen kamen auf 75%, die Hausärzte auf 60% und die nurse practitioners auf 55%. Ebenso hat sich gezeigt, dass die Top-1-Genauigkeit der KI um fast 10% Punkte höher war, wenn wenigstens 3 verschiedene Bilder zur Analyse zur Verfügung standen, als wenn nur ein Bild einer Läsion als Input diente. Ab 5 Bildern stagnierte die Zunahme der der Top-1-Genauigkeit. Diese Studie legt nahe, dass die KI nicht nur zur spezifischen Diagnostik eingesetzt werden kann, sondern auch zur Differentialdiagnostik geeignet ist.

Eine ähnliche Studie hat sich auf die Klassifikation pigmentierter Hautläsionen fokussiert, deren Farbe von braun zu schwarz reichte (66). In diesem Fall sind 6 Differentialdiagnosen (Melanome, BCC, Nävi, Altersflecken, Seborrhoische Keratose und Hämatome/Hämangiome) eingeschlossen worden. Das Besondere ist, dass sowohl für das Training als auch für den Test der KI ausschließlich von einer Digitalkamera aufgenommen Bilder verwendet worden sind. Alle Bilder stammten aus dem National Cancer Center Hospital in Tokyo. Zum Test sind 10 erfahrene Ärzte sowie 10 weniger erfahrene Ärzte der Dermatologie hinzugezogen worden, welche Testsets mit jeweils 200 Läsionen beurteilen sollten. Zum einen ging es um die Einordnung der Läsion in eine der oben genannten Kategorien und zum anderen ging es um die Unterscheidung zwischen malignen und benignen Läsionen. Die Zielgrößen waren u.a. die spezifische Diagnosegenauigkeit, die Sensitivität und die Spezifität. Die Genauigkeit der Diagnose lag bei der KI bei 86,2%. Die erfahrenen Ärzte erreichten im Mittel eine Genauigkeit von 79,5% ($p = 0,0081$) und die weniger erfahrenen Ärzte kamen auf 75,1% ($p < 0,00001$). Die Genauigkeit für die dichotome Unterscheidung benigne vs. maligne lag bei der KI bei 91,5% (Sensitivität=83,3%, Spezifität=94,5%). Die Genauigkeit der erfahrenen Ärzte betrug 86,6% (Sensitivität=86,3%, Spezifität=86,6%) und die der weniger erfahrenen Ärzte betrug 85,3% (Sensitivität=83,5%, Spezifität=85,9%). Ein Dermatologe habe eine höhere spezifische Diagnosegenauigkeit erreicht als die KI. Es sollte berücksichtigt werden, dass den Ärzten keine weiteren Informationen zu den Fällen zur Verfügung standen. Zudem ist zu berücksichtigen, dass diese KI nur bei braun bis schwarz pigmentierten Läsionen einsetzbar ist. Ein möglicher Grund, warum die Diagnosegenauigkeit der KI und der Ärzte höher ausfiel, als in der zuvor aufgeführten Studie von Liu et al. ist, dass in dieser Studie weniger Differentialdiagnosen eingeschlossen worden sind.

Maron et al. haben ebenfalls eine multiklassifizierende KI getestet, wobei der primäre Endpunkt jedoch durch die Unterscheidung benigne vs. maligne definiert war (67). Der sekundäre Endpunkt wurde durch die Einordnung der Läsion in eine der folgenden Klassen abgebildet: (1): aktinische Keratose, M. Bowen, SCC, (2): BCC, (3): Seborrhoische Keratose, Lentigo solaris, Lichen ruber planus, (4): Nävus, (5): Malignes Melanom. Die KI wurde mit dermatoskopischen Aufnahmen der ISIC trainiert und im Vergleich zu 112 Dermatologen aus deutschen Universitätskliniken getestet, welche eine durchschnittliche Berufserfahrung von vier Jahren hatten. Das Testset bestand jeweils aus 50 Bildern. Für die dichotome Unterscheidung benigne vs. maligne lag die durchschnittliche Sensitivität der Ärzte bei 74,4%, die Spezifität betrug 59,8%. Bei gleicher Sensitivität war die Spezifität der KI (91,3%) signifikant größer ($p < 0,001$). Ein ähnliches Bild hat sich in Bezug auf die Einordnung der Läsionen in eine der 5 zuvor genannten Klassen gezeigt. Bei einer durchschnittlichen Sensitivität der Ärzte von 56,5% lag die Spezifität bei 89,2%. Die Spezifität der KI war bei gleicher Sensitivität mit 98,8% signifikant größer ($p < 0,001$). Davon ausgenommen ist die Klasse der BCC. In diesem Fall war die Spezifität der Ärzte vergleichbar mit der der KI. Auffällig war, dass sowohl die Ärzte als auch die KI in 21% der Fälle Melanome als Nävi fehldiagnostiziert haben. Umgekehrt hat die KI in 8% der Fälle Nävi als Melanome gedeutet, während der Anteil unter den Ärzten 30% betrug. An dieser Stelle muss berücksichtigt werden, dass alle Läsionen histologisch gesichert waren, sodass davon auszugehen ist, dass die benignen Läsionen schwer zu diagnostizieren waren. Zudem standen den Ärzten keine weiteren Informationen zu den Fällen zur Verfügung. Dennoch konnte gezeigt werden, dass eine multiklassifizierende KI in einem künstlichen Testumfeld einem Dermatologenkollektiv mit einer durchschnittlichen Berufserfahrung von vier Jahren in Bezug auf die Sensitivität und Spezifität der dichotomen Klassifikation und der Multiklassifikation von Hautläsionen überlegen sein kann.

Sharma et al. haben in einem Review 52 Studien zu dem Themen „Maschinelles Lernen“ und „Nicht-melanozytärer Hautkrebs“ analysiert und zusammengefasst (68). Die Literatursuche wurde in den Datenbanken PubMed, MEDLINE und arXiv durchgeführt. Dabei hat sich für die Diagnostik von BCC durch KI eine mittlere Sensitivität von 89% und eine mittlere Spezifität von 85% gezeigt. Bei der Diagnostik von SCC kam die KI im Durchschnitt auf eine Sensitivität von 83% und eine Spezifität von 98%. Diese Ergebnisse wurden mit zwei Studien zu den Themen Dermatoskopie von SCC und BCC verglichen (69, 70). In diesen zeigten Dermatologen mit Hilfe der Dermatoskopie für BCC eine Sensitivität von 89% und eine Spezifität von 86%. Für SCC wurde eine Sensitivität von 79% und eine Spezifität von 87% erreicht. Im gepaarten t-Test zeigte sich kein signifikanter Unterschied zu der Leistung der KI ($p < 0,05$). Die Autoren folgerten daraus, dass es nicht nachgewiesen ist, dass KI Dermatologen bei der Diagnostik von BCC und SCC überlegen ist. Zudem wurde angemerkt, dass die eingeschlossenen Studien in künstlichen Testumfeldern durchgeführt worden sind und ein klinischer Stresstest fehle.

In den einzelnen Studien waren die KI den Dermatologen in ihrer Diagnosegenauigkeit zumeist überlegen, teilweise signifikant. Dabei sind jedoch verschiedene Diagnosen in den Studien berücksichtigt worden, sodass diese nur bedingt vergleichbar sind. Zudem gibt es Unterschiede der Diagnosegenauigkeiten für verschiedene Läsionen. In einer Studie war die KI den Dermatologen in der Genauigkeit für maligne Läsionen signifikant überlegen, während diese für benigne Läsionen signifikant unterlegen war. Jedoch waren die Dermatologen in diesem Fall deutlich besser bei der Erkennung amelanotischer Melanome. Im Sinne der Multiklassifikation kann die KI zudem Differentialdiagnosen beispielsweise in Form einer Top-3-Genauigkeit angeben. In einem Review zu der Erkennung von SCC und BCC hat sich gezeigt, dass die getesteten KI den zum Vergleich herangezogenen Dermatologen nicht überlegen waren.

2.7.3 KI als Assistenzsystem zur Hautkrebsdiagnostik

Einige Studien haben sich mit dem unterstützenden Einsatz von KI beschäftigt, um eine Diagnose oder Therapieentscheidung zu treffen. Diese werden nachfolgend vorgestellt.

Die Studie von Hekler et al. beschäftigt sich mit der Kombination der unabhängigen Diagnosen von KI und Dermatologen, ohne dass diese explizit zusammengearbeitet haben (71). Zu diesem Zweck wurde ein Gradient Boosting verwendet, welches Teil des maschinellen Lernens ist und auf Entscheidungsbäumen basiert. In dieser Studie sind dermatoskopische Aufnahmen der ISIC zum Training und Testen der KI eingesetzt worden. Zugleich sind 112 Dermatologen deutscher Universitätskliniken getestet worden, in Testsets von 50 Bildern Läsionen in 5 verschiedene Klassen einzusortieren (primärer Endpunkt) und zwischen benigne und maligne (sekundärer Endpunkt) zu unterscheiden. Die Klassen sahen wie folgt aus: (1): aktinische Keratose, M. Bowen, SCC, (2): BCC, (3): Seborrhoische Keratose, Lentigo solaris, Lichen ruber planus, (4): Nävus, (5): Malignes Melanom. Für die Klassifikation hat sich gezeigt, dass die Kombination aus Arzt und KI eine Genauigkeit von 82,95% erzielen konnte, womit sie geringfügig besser war (1,36%) als die KI, welche wiederum besser als der Durchschnitt der Dermatologen war. Des Weiteren zeigte sich, dass die Sensitivität der Kombination (89%) für die Unterscheidung benigne vs. maligne signifikant ($p < 0,05$) höher war als die der KI (84,1%), welche auch in diesem Fall den Dermatologen überlegen war. Die Spezifität der Kombination nahm von 89,2% (KI allein) auf 84% ab.

In einer Studie von Tschandl et al. ist eine KI mit pigmentierten Läsionen aus öffentlichen Datenbanken getestet worden (72). Dabei sind Melanome, BCC, SCC, Aktinische Keratose, melanozytäre Nävi, benigne keratinozytäre Läsionen, Dermatofibrome sowie vaskuläre Läsionen eingeschlossen worden. Diese KI wurde anschließend 302 Ärzten aus 41 Ländern, es handelte sich überwiegend um Dermatologen, lediglich 12,6% der Ärzte waren Hausärzte, zur unterstützenden Hautkrebsdiagnostik zur Verfügung gestellt. Es hat sich gezeigt, dass die Genauigkeit der Diagnose der Ärzte von 63,6% auf 77% gestiegen ist, wenn diese die KI unterstützend einsetzten. Auffällig war, dass diejenigen Ärzte mit der geringsten Erfahrung deutlich häufiger (26%) ihre Diagnose aufgrund der KI verändert haben als die Ärzte mit viel Erfahrung (14,7%). Zudem hat sich gezeigt, dass wenn sich erfahrene Ärzte in ihrer Diagnose sicher waren, diese auch zumeist richtig war. Des Weiteren sind in einem Abschnitt der Studie bewusst falsche Diagnosen durch die KI eingestreut worden, um zu sehen, wie sich dies auf die Diagnosen durch die Ärzte ausgewirkt. Dabei konnte gezeigt werden, dass die Diagnosegenauigkeit der Ärzte nicht gebessert, in den meisten Fällen sogar deutlich verschlechtert wurde. Auffällig war, dass die Unterstützung durch die KI vor allem für Aktinische Keratosen und SCC von Vorteil war. In diesem Fall konnte die Diagnosegenauigkeit von 31,5 % (Arzt) auf 40,1% (Kombination aus Arzt und KI) gesteigert werden.

In einer weiteren Studie sind von Han et al. Bilder von 174 verschiedenen Hauterkrankungen aus 6 verschiedenen Datenbanken ausgewählt worden, um eine KI zu trainieren (73). Anschließend wurde diese mit Aufnahmen aus der Edinburgh Datenbank (10 verschiedene Läsionstypen) und der SNU Datenbank (134 verschiedene Läsionstypen) getestet. Es konnte gezeigt werden, dass durch den unterstützenden Einsatz der KI die durchschnittliche Sensitivität der Malignitätsvorhersage von 47 getesteten Dermatologen um 12,1% anstieg ($p < 0,0001$) im Vergleich zu der alleinigen Diagnose durch den Dermatologen. Die Spezifität stieg um 1,1% an ($p < 0,0001$). Ebenso sind 23 Nicht-Mediziner getestet worden. Deren Sensitivität für Malignität steigerte sich um 83,8% ($p < 0,0001$). Ebenso konnte die Top-1- und die Top-3-Genauigkeit von 4 getesteten Ärzten deutlich gesteigert werden (um 7%, $p = 0,045$ und um 10,1%, $p = 0,002$). Es zeigt sich, dass die KI dazu geeignet ist, durch den unterstützenden Einsatz bei einer großen Anzahl von Differentialdiagnosen die Diagnosegenauigkeit sowie die Sensitivität und Spezifität für die Malignitätsvorhersage von Dermatologen zu steigern.

Sowohl in der binären als auch in der Multiklassifikation konnten Ärzte von dem unterstützenden Einsatz durch KI in den dargelegten Studien profitieren. Dabei nahmen sowohl die Diagnosegenauigkeit in der Multiklassifikation als auch die Sensitivität und

Spezifität im Vergleich zu der durchschnittlichen Leistung des getesteten Dermatologenkollektivs bei der Malignitätsbewertung zu. In einer Studie wurde jedoch auch gezeigt, dass die Diagnosegenauigkeit abnahm, als durch die KI gezielt falsche Diagnosen eingestreut worden sind, was die Forderung nach einer robusten KI unterstreicht.

2.7.4 Mobiler Einsatz von KI

Udrea et al. haben in ihrer Studie die App SkinVision getestet, welche auf Basis einer KI funktioniert (52). Die Anwender dieser App können Fotos von ihren Hautläsionen mit dem Smartphone machen und diese nach ihrem Risiko für Malignität bewerten lassen (hohes Risiko vs. niedriges Risiko). Zum Training und Testen der KI sind Daten der Nutzer selbst sowie von Studien der Universitätsklinik München und Eindhoven verwendet worden. Dabei zeigte die KI eine Sensitivität von 95% sowie eine Spezifität von 78,3%. Zum Training der Malignitätserkennung sind Bilder von Läsionen wie Melanome, SCC, BCC, Aktinische Keratose, M. Bowen sowie Vorstufen dieser verwendet worden. Jedoch merken die Autoren an, dass die App kein Gerät zur Hautkrebsdiagnostik sei und nicht den Arztbesuch ersetze. Vielmehr könne diese eingesetzt werden, um Patienten zu priorisieren, beispielsweise in einem überlasteten Gesundheitssystem. Deeks et al. haben jedoch eine Kritik zu dieser Studie veröffentlicht, in der sie davon ausgehen, dass sowohl die Sensitivität als auch die Spezifität dieser App überschätzt wird (74). Als einen Grund führen sie das verwendete Studiendesign der Fall-Kontroll-Studie an. Es habe sich gezeigt, dass dieses in vielen klinischen Anwendungen zur Überschätzung der Sensitivität und Spezifität geführt habe. Zudem sind für die Evaluation der App Bilder aus Studien verwendet worden. Diese App richte sich jedoch an den Laien, weshalb nur Bildmaterial aus dem Kontext der App-Anwendung hätte verwendet werden sollen. Nicht zuletzt sei es sehr wahrscheinlich, dass die Datenselektion für die Evaluation der App dazu geführt habe, dass schwer zu diagnostizierende Läsionen aussortiert worden sind.

In einer anderen Studie ist eine KI trainiert und getestet worden, welche über eine Cloudplattform und mobile Geräte verwendet werden kann (54). Zu diesem Zweck ist auf Datenmaterial des Kaohsiung Chang Gung Memorial Hospital (KCGMH) und des Datenpools HAM10000 zurückgegriffen worden. Der Datensatz des KCGMH enthielt BCC, SCC, Melanome, melanozytäre Nävi und Seborrhoische Keratosen. In den HAM10000 Daten waren (1) BCC, (2) Aktinische Keratosen, M. Bowen, (3) Lentigo solaris, Seborrhoische Keratosen, Lichen planus-like Läsionen, (4) Dermatofibrome, (5) Melanome, (6) Melanozytäre Nävi und (7) Angiome, Angiokeratome, Pyogene Granulome, Hämatome. Für die Unterscheidung zwischen benigne und maligne im KCGMH Datensatz erreichte die KI eine Genauigkeit von 89,5%. Die Sensitivität betrug 90,8% und die Spezifität betrug 88,1%. Für die Multiklassifikation im KCGMH Datensatz erreichte die KI eine Genauigkeit von 72,1%. Im HAM10000 Datensatz lag die Genauigkeit bei 85,8%.

In einem Review zu dem Thema Apps zur Melanomerkennung sind nach einer Suche in den Datenbanken Cochrane Central Register of Controlled Trials, MEDLINE, Embase, CINAHL, CPCI, Zetoc, Science Citation Index, US National Institutes of Health Ongoing Trials Register, NIHR Clinical Research Network Portfolio Database, and the World Health Organization International Clinical Trials Registry Platform 2 Studien aufgenommen worden, in denen insgesamt vier KI basierte Apps getestet worden sind (75). Diese wiesen eine Sensitivität für die Erkennung von Melanomen bzw. Malignität von 7% bis 78% auf. Die Spezifität lag zwischen 37% und 94%. 27% bis 93% aller Melanome oder atypischen melanozytären Läsionen sind von der KI nicht erkannt worden. Zu den Studien ist angemerkt worden, dass diese nicht im realen Setting getestet worden sind, so sind beispielsweise die Aufnahmen von den Läsionen von Ärzten gemacht worden und nicht den Anwendern der Apps. Ebenso sind 2% bis 18% der Bilder ausgeschlossen worden, da die KI diese nicht einordnen konnte. Insgesamt kamen die Autoren zu dem Schluss, dass die diskutierten Studien ein hohes Risiko für einen Bias hatten. Die Aussagekraft des Reviews ist aufgrund der geringen Studienanzahl eingeschränkt.

Die dargestellten Studien zeigen, dass es bislang keine überzeugende App zur Hautkrebsdiagnostik gibt, da die Studien alle ein hohes Risiko für Bias besitzen.

2.7.5 Verbesserungsansätze von KI

In einer Studie von Han et al. ist eine KI mit Datensätzen verschiedener Ethnien trainiert und validiert worden (76). und im Hinblick auf die AUC, Sensitivität und Spezifität für bis zu 12 verschiedene Läsionen getestet worden (BCC, SCC, Melanome, melanozytäre Nävi, intraepitheliale Karzinome, Aktinische Keratose, Pyogene Granulome, Hämangiome, Dermatofibrome, Lentigo, Warzen, Seborrhoische Keratose). Zunächst ist dabei lediglich der Asan Datensatz verwendet worden, welcher mittels Aufnahmen von Personen asiatischer Herkunft erstellt worden ist. Anschließend wurde die KI mit Aufnahmen aus dem Edinburgh Datensatz validiert, welcher aus Aufnahmen von Personen kaukasischen Hauttyps besteht. Dabei war auffällig, dass die KI mit einer AUC von 0,78 für BCC relativ schlecht abschnitt. Als zusätzlich atlas site images, ein Datensatz, welcher aus einer Vielzahl an im Internet verfügbaren Dermatologie Atlanten besteht, als Trainingsmaterial hinzugezogen worden ist, erhöhte sich die AUC für BCC im Edinburgh Datensatz auf 0,90. Als potenzielle Erklärung für diese Beobachtung gaben die Autoren an, dass dies auf die unterschiedliche Prävalenz verschiedener BCC-Subtypen in verschiedenen Ethnien zurückführbar wäre. Ebenso haben die 16 zum Vergleich herangezogenen koreanischen Dermatologen im Edinburgh Datensatz für BCC sowie für melanozytäre Nävi eine Sensitivität von <40% erzielt. Dies wurde auf die Tatsache zurückgeführt, dass die Ärzte es gewohnt waren, Läsionen auf der Haut von Personen asiatischer Abstammung zu sehen und weniger bei Kaukasiern.

In einem Review zu dem Thema KI in der Hautkrebsdiagnostik ist die Bedeutung von Bildern von Hautläsionen in direktem Zusammenhang mit Patientendaten analysiert worden (77). Dazu fand eine Literaturrecherche in Google Scholar, PubMed, MEDLINE und ScienceDirect statt. Schließlich sind 11 Studien in das Review aufgenommen worden. 5 Studien beschäftigten sich mit binärer Klassifikation und 6 Studien mit Multiklassifikation. In den meisten Studien sind das Alter, das Geschlecht und die Lokalisation der Läsion berücksichtigt worden. Dabei konnte zwar gezeigt werden, dass diese zusätzlichen Informationen die Sensitivität der KI erhöht haben, jedoch ist die Spezifität in diesem Zusammenhang zum Teil nicht angegeben worden oder auch gesunken. Andere Faktoren wie beispielsweise die Erhabenheit der Läsion oder Symptomen wie Blutung oder Schmerz sorgten für eine Zunahme der Sensitivität und der Spezifität bei Läsionen wie Melanomen, Nävi, BCC oder Aktinischer Keratose. Die Autoren sprechen insgesamt von einer Verbesserung der Diagnosegenauigkeit der KI durch den Einfluss zusätzlicher Patientendaten.

Eine weitere Studie hat sich ebenfalls mit dem Einfluss von Patientendaten auf die Leistung einer KI beschäftigt (78). In diesem Fall sind 8 verschiedene KI mit Bildern von der ISIC trainiert worden, wobei im Training sowohl niedrigaufgelöste als auch hochaufgelöste Aufnahmen eingesetzt worden sind. Zudem wurden die KI einmal mit und einmal ohne zusätzliche Patientendaten (Alter, Geschlecht, Lokalisation der Läsion) trainiert. Die Diagnosekategorien waren (1) Melanozytäre Nävi, (2) vaskuläre Läsionen, (3) benigne Keratosen, (4) Dermatofibrome, (5) intraepitheliale Karzinome, (6) BCC und (7) Melanome. Zudem sind die Leistung der KI in Bezug auf die Diagnosegenauigkeit mit 22 Hausärzten verglichen worden. Für niedrig aufgelöste Bilder konnte gezeigt werden, dass die Diagnosegenauigkeit der 8 KI im Durchschnitt signifikant gesteigert wurde ($p=0,004$), wenn zusätzlich Patientendaten durch die KI berücksichtigt worden sind. Der stärkste Zuwachs an der Diagnosegenauigkeit zeigte sich bei der KI VGG19. Dieser betrug 5,22%. Die stärkste Zunahme in Bezug auf die einzelnen Läsionen zeigte sich bei der KI MobileNet. In diesem Fall stieg die Genauigkeit der Erkennung von Dermatofibromen um 46%. Auch bei den KI, welche mit hochaufgelösten Bildern trainiert und getestet worden sind, stieg die Diagnosegenauigkeit unter Einfluss der Patientendaten signifikant an ($p=0,02$). In diesem Fall profitierte vor allem die KI ResNet50, deren Diagnosegenauigkeit um 4,75% gesteigert wurde. Bei den spezifischen Läsionen steigerte sich u.a. die Diagnosegenauigkeit von ResNet101 für vaskuläre Läsionen um 37% und die von

VGG19 für Melanome um 32%. Schließlich lag die Top 1 Genauigkeit der schlechtesten KI (EfficientNetB5) bei 78,4% und somit deutlich höher als die der getesteten Hausärzte (27,74%).

In einer Studie von Young et al. sind mehrere KI trainiert und einem Stresstest ausgesetzt worden, um zu untersuchen, wie sich diese in verschiedenen künstlichen Testumfeldern verhalten (79). Die KI beschränkten sich in diesem Fall auf die binäre Unterscheidung zwischen Melanomen und Nävi. In diesem Stresstest wurden verschiedene Datensets verwendet, welche u. a. aus anderen Quellen als die Trainingsdatensets stammten, um zum Beispiel Selektionsbias zu verhindern. Außerdem wurden die Bilder teilweise rotiert oder vergrößert. Zum Teil wurden auch Aufnahmen von Läsionen gezeigt, welche nicht Teil des Trainings der KI waren. Model A erreichte in verschiedenen Testsets eine AUROC von 0,83 bis 0,89 für die Rate richtig positiver Vorhersagen. Veränderungen der Bilder in den Testsets durch veränderte Aufnahmewinkel oder Vergrößerungen haben jedoch dazu geführt, dass zwischen 15,8% und 36,7% der Klassenzuordnung Melanom vs. Nävus anders ausfielen. Die Autoren weisen darauf hin, dass die getesteten KI anfällig für Veränderungen der Aufnahmebedingungen der Läsionen und somit in diesem Aspekt nicht robust sind.

Zusammenfassend lässt sich sagen, dass KI zur Hautkrebsdiagnostik mehrere Ansatzpunkte für Verbesserungspotential besitzt. Zum einen müssen verschiedenen Ethnien berücksichtigt werden, zum Beispiel im Hinblick auf die Prävalenz bestimmter Läsionen. Sind diese in dem Trainingsdatensatz unterrepräsentiert, wirkt sich das negativ auf die Diagnoseleistung der KI aus. Auch scheinen zusätzlichen Patientendaten wie Alter, Geschlecht, Lokalisation der Läsion, Symptome wie Blutung, Schmerzen oder die Erhabenheit der Läsion die Diagnosegenauigkeit, Sensitivität und Spezifität der KI steigern zu können. Jedoch muss berücksichtigt werden, dass die Verbesserung eines Parameters wie der Sensitivität für eine bestimmte Läsion die Spezifität für diese Läsion senken kann. Außerdem hat der Test mehrerer KIs gezeigt, dass diese anfällig für Vergrößerungen oder Veränderungen des Aufnahmewinkels der Läsionen waren.

2.7.6 Einsatz von KI

KI zur Hautkrebsdiagnostik gibt es auch bereits zu kaufen. Beispielsweise bietet Foto Finder Systems GmbH aus Bayern KI basierte Lösungen als App aber auch für Ganzkörper-scans an (80). Die KI gibt eine Risikobewertung von melanozytären und nicht-melanozytären Hautläsion in Prozent an und nutzt eine sogenannte Heatmap, aus welcher der Anwender erfährt, auf welche Strukturen die KI ihren Fokus bei der Analyse gelegt hat. Auf diese Weise können mögliche Störfaktoren wie beispielsweise Tattoos identifiziert werden.

Auch kann KI nicht nur zur Hautkrebsdiagnostik eingesetzt werden, sondern auch für ein Ganzkörper-Assessment, wie es etwa von Neko angeboten wird (81). Dieses Health Start-up nutzt über 70 verschiedene Sensoren, um insgesamt etwa 50 Millionen Datenpunkte pro Patienten zu erheben. Dabei findet u.a. ein Ganzkörper-scans, eine nicht-invasive Blutdruckmessung mit 8 Sensoren, eine Bestimmung der Herzfrequenz und eine Messung der Steifigkeit von Arterien statt. Diese Daten werden gesammelt und mit Daten gesunder Menschen verglichen, um mögliche Krankheiten frühzeitig oder auch vorzeitig zu erkennen. Anschließend werden diese Daten mit Ärzten besprochen. KI findet in diesem Zusammenhang Einsatz, um den Körperscan auszuwerten und mögliche Abweichungen der erhobenen Parameter von denen gesunder Menschen zu erkennen

2.8 KI in der Hautkrebsdiagnostik aus Sicht der Patienten und der Ärzte

2.8.1 Sicht der Patienten

In einer auf Deutsch durchgeführten, anonymen Onlinebefragung gaben 88% der Befragten an, den Begriff KI schonmal gehört zu haben und ein gewisses Verständnis für diese Thematik zu besitzen (82). 94% haben dabei gesagt, den Einsatz von KI im Bereich der Medizin

grundsätzlich zu befürworten. Über 90% der Befragten stimmten zu, KI als Assistenzsysteme im Bereich der Hautkrebsdiagnostik einzusetzen und 56% konnten sich vorstellen, KI im Zuge der Früherkennung selbstständig zuhause einzusetzen. Insgesamt waren 88% der Befragten bereit, ihre Daten zur Verfügung zu stellen, um KI weiterzuentwickeln. Im Falle eines Konflikts, bei dem sich die Aussagen der KI und des Arztes widersprechen, gab die Mehrheit an, eine Biopsie nehmen lassen zu wollen. Häufig genannte Bedenken waren Datensicherheit, die unpersönliche Beziehung zur KI als Diagnostikum und die Fehleranfälligkeit der KI. Dahingegen erhofften sich die Befragten einen beschleunigten diagnostischen Prozess und insgesamt eine frühere Erkennung von Hautkrebs. Außerdem gaben sie an, dass die Diagnostik durch den Einsatz von KI verlässlicher, objektiver und transparenter werden kann. Zuletzt wurde auch die Möglichkeit benannt, dass Ärzte durch den Einsatz von KI ihre eigenen Fähigkeiten verbessern können. Das Alter und die Bildung der Befragten hatten in diesem Zusammenhang keinen signifikanten Einfluss auf die Ergebnisse der Befragung. Dahingegen sprachen sich signifikant mehr Patienten mit einem Melanom in der Vorgeschichte für den Einsatz von KI in der Medizin und im häuslichen Umfeld zur Früherkennung von Hautkrebs aus.

Ein anderer Ansatz ist in Form eines halbstrukturierten Interviews in einem Krankenhaus in Boston gewählt worden (83). In diesem Fall ist die eine Hälfte der Patienten zu KI in Form eines Assistenzsystems und die andere Hälfte zu KI in Form eines stand-alone Systems befragt worden. 90% der Befragten gaben an, ein elektronisches Endgerät zu besitzen. Ebenso viele Patienten sagten, dass sie digitale Gesundheitsdienste verwenden würden, wobei in über 80% der Fälle auf Google zurückgegriffen worden sei. 21% der Befragten gaben in dem Interview an, keine Vorstellung von KI zu haben. Als Vorteile durch den Einsatz von KI in der Hautkrebsdiagnostik sind ein beschleunigter diagnostischer Vorgang sowie der verbesserte Zugang zur Gesundheitsversorgung genannt worden, ebenso wie eine Verringerung von Gesundheitskosten. Auch könnte die Angst vor einer Diagnose reduziert werden, da die Wartezeit bis zur Diagnose durch die KI wegfallen würde. Zugleich war Angst ein häufig benanntes Risiko, welches der Einsatz von KI vor allem als stand-alone System mit sich bringe. Es ist die Angst mit einer Krebsdiagnose allein gelassen zu werden. Aber auch die Depersonalisierung der gesundheitlichen Betreuung ist als Nachteil benannt worden. Einen Vorteil sahen die Patienten vor allem in der Möglichkeit, genauere und objektivere Diagnosen durch den Einsatz von KI zu ermöglichen. Zugleich ist die fehlende Genauigkeit von vielen Patienten als Kritikpunkt benannt worden. Außerdem wurde gesagt, dass der Einsatz von stand-alone Systemen Patienten die Möglichkeit gebe, sich niederschwellig und somit möglicherweise früher um ihre Gesundheit zu kümmern. Als wichtigste Möglichkeit, für den Fall, dass die KI und der Arzt verschiedene Aussagen treffen würden, benannten 67% der Befragten die Entnahme einer Biopsie. Etwa die Hälfte (52%) der Patienten war der Meinung, dass der Hersteller für die Genauigkeit der KI verantwortlich sei. 75% (71% in der KI stand-alone Gruppe) der Patienten gaben an, dass sie KI ihren Verwandten weiterempfehlen würden. In beiden Gruppen war die symbiotische Beziehung zwischen Menschen und KI das dominierende Thema.

In einer mittels Fragebögen in zwei Krankenhäusern des Vereinigten Königreichs durchgeführten Befragung gaben 39% der Patienten an, dass sie den Einsatz von KI durch ihren Hausarzt unterstützen würden, wenn diese dazu eingesetzt würde, den Hausarzt bei der Entscheidung zu unterstützen, den Patienten möglicherweise an einen Hautspezialisten weiterzuleiten (84). 81% der Befragten gaben an, eine Beurteilung und Bestätigung einer von einer KI gemachten Diagnose durch einen Dermatologen als wichtig zu erachten und lediglich 12% der Patienten gaben an, dass sie auf die alleinige Diagnose einer KI vertrauen würden. Dahingegen äußerten 87% der Befragten Verständnis für die Tatsache, dass ihre Daten dabei helfen könnten, KI weiterzuentwickeln. Diese Einstellungen stellte sich als vom Alter der Befragten unabhängig dar. Eine positive Hautkrebsanamnese ließ die Einstellung gegenüber dem unterstützenden Einsatz von KI unbeeinflusst. Schließlich war den Patienten wichtig, trotz des Einsatzes von KI, nicht die Interaktion mit dem Arzt zu verlieren.

In einer in den USA durchgeführten Onlineumfrage, ist die Compliance in Bezug auf Therapieempfehlungen im Zusammenhang mit dem Einsatz von KI in der Hautkrebsdiagnostik hypothetisch erfragt worden (85). Die Compliance war geringer, wenn die Diagnose allein durch eine KI gefällt worden ist, im Vergleich zu der Compliance bei der Diagnose durch einen Arzt mit oder ohne Assistenzsystem. Ob der Arzt ein Assistenzsystem einsetzte oder nicht, beeinflusste die Compliance hingegen nicht.

2.8.2 Sicht der Ärzte

In einer von einer südkoreanischen Universität durchgeführten Befragung von Ärzten und Medizinstudenten gaben lediglich 6% der Befragten an, gut mit KI vertraut zu sein (86). Dennoch waren rund 73% der Meinung, dass KI einen Nutzen für die Medizin habe. 44% der Befragten schätzten KI als dem Arzt überlegen ein und knapp 80% sagten, dass sie im Falle eines Konflikts eher auf den Arzt vertrauen würden. Etwa 50% sagten, dass sie im Falle eines Problems im Rahmen des Einsatzes von KI den Arzt verantwortlich machen würden, während ca. 20% den Hersteller in der Verantwortung sahen.

In einer an Dermatologen gerichteten, internationalen Onlinebefragung gaben knapp 24% der Ärzte an, eine sehr genaue Vorstellung zu KI im Bereich der Dermatologie zu haben (87). Knapp 6% der Befragten waren der Meinung, dass KI in absehbarer Zeit den Dermatologen ersetzen wird und etwas weniger als 20% gaben Angst in Bezug auf den zunehmenden Einsatz von KI im Bereich der Dermatologie an. Rund 77% der Ärzte waren der Meinung, dass KI für die Dermatologie von Nutzen ist und knapp 80% der Befragten stimmten zu, dass KI ein Teil der medizinischen Ausbildung sein solle.

In einer Befragung von Hausärzten in Frankreich in Bezug auf den Einsatz von KI zur Diagnostik von nicht melanozytärem Hautkrebs hat sich gezeigt, dass die Akzeptanz für den Einsatz von KI mit 45% (starke Zustimmung) beziehungsweise 41% (Zustimmung) sehr hoch ist. 86% der Befragten sagten, dass sie KI zur Hautkrebsdiagnostik in ihrer Praxis als nützlich einschätzen. 98% der Hausärzte gab in diesem Zusammenhang an, Schwierigkeiten mit der Diagnostik von nicht melanozytärem Hautkrebs zu haben. Jedoch sagten 68%, dass sie nicht für die KI zahlen wollten. Dem Vorschlag, diese KI dem Patienten frei zur Verfügung zu stellen, standen 20% der Befragten sehr kritisch und 37% kritisch gegenüber (88).

3 Material und Methoden

3.1 Literaturrecherche

Die für die Erarbeitung der Thematik, für die Erstellung der Befragung und für die Diskussion des aktuellen Forschungsstands zugrunde gelegte Literatur entstammt überwiegend einer strukturierten Recherche in den Datenbanken PubMed, Cochrane Library und Web of Science,

welche vom 01.12.2020 bis zum 15.04.2022 erfolgt ist. Andere Quellen entstammen Querverweisen aus den recherchierten Quellen sowie der Funktion „Similar articles“ auf PubMed. Die zitierten Bücher stammen alle aus dem Zugang zur Universitätsbibliothek Mainz.

In PubMed sind unter der Funktion „Advanced Search Builder“ die Termini “(patient compliance[MeSH Terms]) AND (ai artificial intelligence[MeSH Terms])” und “((skin neoplasms[MeSH Terms]) AND (data collection[MeSH Terms])) AND (ai artificial intelligence[MeSH Terms])” verwendet worden. Ebenso erfolgte eine Freitextsuche mit den Begriffen “artificial intelligence patient perspective skin cancer“ und „physician confidence artificial intelligence“.

Bei der Recherche in der Cochrane Library kam die „Advanced Search“ Funktion zum Einsatz. Dabei wurden jeweils unter der Kategorie „All Text“ die Begrifflichkeiten „skin neoplasms“ und artificial intelligence“ und „patient perspective“ mit dem logischen Konnektor „AND“ verbunden.

Für die Literatursuche im Web of Science ist auf die allgemeine Suchfunktion zurückgegriffen worden. Unter der Kategorie „All fields“ sind die Begriffe „skin neoplasms“, „artificial intelligence“ und „patient perspective“ mit dem Konnektor „AND“ verwendet worden.

3.2 Studiendesign, Studiendurchführung und Befragungskollektiv

3.2.1 Studiendesign

Bei dieser Studie handelt es sich um eine zweiarmige Querschnittstudie. Die anonyme Befragung der Patienten ist monozentrisch durchgeführt worden. Sie erfolgte in der Hautklinik der Universitätsmedizin Mainz. Die Ärzte hingegen wurden im Sinne eines multizentrischen Designs deutschlandweit befragt. Bei jeweils etwa der Hälfte der beiden Gruppen war eine zusätzliche Aufklärung in dem Fragebogen beinhaltet.

3.2.2 Patientenbefragung

Die Patientenbefragung wurde ausschließlich in der Hautklinik der Universitätsklinik Mainz durchgeführt. In diesem Rahmen wurden 3 verschiedene Ansätze verfolgt. Der erste Ansatz zielte auf die Befragung der stationären Patienten ab. Die Patienten stammten von der Station 4b der Hautklinik. Die Befragung der stationären Patienten erfolgte mittels Fragebögen. Diese wurden den Patienten bei ihrer Aufnahme auf die Station vom Pflegepersonal angeboten. Die Befragung erfolgte von 24.11.2021 bis zum 13.01.2022. Parallel dazu gab es einen Aushang im Patientenwartebereich vor dem OP der Hautklinik sowie in den Umkleiden des Haut-OP. Dieser bot einen kurzen Informationstext für die Patienten sowie einen QR-Code, welcher direkt zu der Befragung führte (siehe **Anhang 8.1**). Die Onlinebefragung war zwischen dem 16.08.2021 und dem 16.12.2021 freigeschaltet. Der dritte Ansatz bestand aus der händischen Verteilung von Fragebögen im Wartebereich der Ambulanz im Zeitraum vom 15.02.2022 bis zum 25.02.2022. Für die Patienten gab es keine Ausschlusskriterien. Es sind gezielt Patienten aus dem Fachbereich Dermatologie befragt worden.

3.2.3 Ärztebefragung

Die Befragung der Ärzte zielte darauf ab, sowohl den stationären als auch den ambulanten Berufszweig zu erreichen. In diesem Rahmen wurde ein kurzer Informationstext mit einem Link zu der Onlinebefragung von einer offiziellen Studentenmailadresse der Universität Mainz verschickt. Die Mails an die Unikliniken in Deutschland wurden im Zeitraum vom 20.09.2021 bis zum 28.09.2021 verschickt. In diesem Zusammenhang sind alle Hautkliniken deutscher Universitätsklinken angeschrieben worden. Sofern eine Mailadresse vorhanden war, ging die Mail direkt an die leitende ärztliche Stelle. War dies nicht der Fall, so ist das Chefsekretariat

angeschrieben worden. In der Hautklinik der Mainzer Universitätsklinik selbst wurde eine interne Benachrichtigung versandt. Die E-Mails an die niedergelassenen Dermatologen gingen zwischen dem 04.10.2021 und dem 23.10.2021 raus. Dabei erfolgte die Auswahl der Praxen über Google Maps. Voraussetzung für die Kontaktaufnahme war, dass die eingezeichnete Praxis Kontaktdaten in Form einer Mailadresse im Internet abgelegt hatte, was fast ausschließlich dann der Fall war, wenn auch eine eigene Homepage vorhanden gewesen ist. In Mainz und im näheren Umkreis beginnend sind zunächst Praxen in Rheinlandpfalz, Hessen und Nordrhein-Westfalen angeschrieben worden. Schließlich ist die Befragung auf Praxen in Norddeutschland, Süddeutschland und zum Schluss auch auf Praxen in Ostdeutschland erweitert worden. Dabei wurden die niedergelassenen Dermatologen sowohl in Großstädten als auch im ländlichen Bereich erfasst. Die Hausärzte wurden im Zeitraum vom 26.10.2021 bis zum 11.12.2021 kontaktiert. Dabei wurde nach dem gleichen Prinzip wie bei der Befragung der niedergelassenen Dermatologen vorgegangen. Die Befragung selbst war für den Zeitraum vom 15.09.2021 bis zum 10.01.2022 freigeschaltet und konnte in dieser Zeit bearbeitet werden. Das Einschlusskriterium für die Ärzte war die Facharzttrichtung. In diesem Zusammenhang sind Hausärzte/Internisten und Dermatologen eingeschlossen worden.

3.3 Erstellung der Befragung

3.3.1 Aufklärung

In den zuvor diskutierten Studien über Befragungen von Patienten wurde zwar festgestellt, dass der überwiegende Anteil der Befragten schonmal etwas von KI gehört hat, jedoch bleibt unklar, wie tief das Verständnis der jeweiligen Person für die Funktionsweise einer KI, vor allem in Bezug auf die Hautkrebsdiagnostik, ist (82, 83). Auch bei den zuvor aufgeführten Studien, bei denen Ärzte und Medizinstudenten befragt worden sind, hat sich gezeigt, dass nur ein geringer Anteil sich stärker mit KI vertraut fühlte (86). Ein größerer Anteil der befragten Dermatologen gab an, Ängste in Bezug auf den Einsatz von KI zu hegen (87). In der Annahme, dass die „Angst vor dem Unbekannten“, also dem Einsatz unvertrauter Technik in Form von KI, die Akzeptanz in Bezug auf den Einsatz von KI mindert, stellt die randomisierte Aufklärung der Befragten einen zentralen Bestandteil dieser Befragung dar (siehe **Anhang 8.2**). Damit sollte die Hypothese untersucht werden, ob die Akzeptanz für den Einsatz von KI durch das Wissen über die Funktionsweise und die Anwendungsmöglichkeit von KI in Bezug auf die Hautkrebsdiagnostik beeinflusst werden kann. Diese Aufklärung fand ihre Umsetzung in einer Randomisierung der Befragten, wobei die Hälfte der Teilnehmer einen zweiseitigen Aufklärungstext im Laufe der Befragung erhalten hat. Um den Einfluss durch das Vorwissen der Befragten zu vermindern, mussten alle Befragten zuvor angeben, ob sie schonmal etwas von KI gehört haben und wie sie ihr Verständnis von KI einschätzten. Diese Einschätzung wiederum gab Auskunft über den Zusammenhang zwischen dem individuellen Wissen der Person in Bezug auf KI und der damit verbundenen Einstellung gegenüber dem Einsatz eben jener. Zudem ist die Reihenfolge der Fragen streng vorgegeben worden, sodass diese Frage wiederum nicht durch die mögliche Aufklärung verfälscht werden konnte. Der Aufklärungstext ist in Laiensprache verfasst worden und war für Ärzte und Patienten identisch. Zunächst ist der Begriff KI näher definiert und eingeordnet worden. Im Anschluss wurde der primitivste Aufbau einer KI dargelegt, um deren Funktionsweise zu erklären. Es folgte eine Erläuterung zu dem Training und der Arbeitsweise einer KI. Diese Aspekte wurden dann auf die Hautkrebsdiagnostik übertragen und in einer Grafik veranschaulicht. Zum Schluss sind das Potential sowie das Risiko des Einsatzes von KI dargelegt worden und es wurde zusätzlich auf die Wichtigkeit geeigneter Daten eingegangen, welche notwendig sind, um KI zu trainieren.

3.3.2 Vergleichbarkeit der Befragung zwischen Patienten und Ärzten

Da KI das Potential besitzt, die Arzt-Patienten-Interaktion und somit auch die Arzt-Patienten-Beziehung zu verändern, ist es von wissenschaftlichem Interesse, die Sichtweisen beider

Parteien im Hinblick auf den Einsatz von KI zu eruieren. Um möglichst aussagekräftige und vergleichbare Daten zu erhalten, sind einige Fragen sowohl in der Ärzefragebogen als auch im Patientenfragebogen zu finden. Im Sinne der bereits thematisierten randomisierten Aufklärung der Befragten sollten beide Parteien ihr Wissen in Bezug auf KI einschätzen. Einen wichtigen Aspekt neben dem Alter und dem Geschlecht stellte zudem die Nutzung von Medien, auch in Bezug auf das Gesundheitsverhalten, dar. Des Weiteren wurde erfragt, wie Patient und Arzt zu dem eigenständigen Gebrauch der KI durch den Patienten stehen. Ebenso ist die Einstellung in Bezug auf den Einsatz von KI im Rahmen der Hautkrebsdiagnostik erfragt worden. Zudem wurde auch das Vertrauen beider Parteien in die KI erörtert. Schließlich wurde noch das Thema der Datensammlung in Bezug auf das Training von KI thematisiert. Ziel dieser Art der Befragung war es, herauszufinden, inwieweit sich die Vorstellung und die Einstellung von Arzt und Patient in Bezug auf den Einsatz von KI im Rahmen der Hautkrebsdiagnostik überschneiden. Umgekehrt soll daraus geschlossen werden können, an welchen Stellen Konfliktpunkte in der KI basierten Hautkrebsdiagnostik entstehen könnten.

3.3.3 Medium der Befragung

Um möglichst viele Ärzte unkompliziert zu erreichen, ist die Befragung der Ärzteschaft mittels einer Onlinebefragung umgesetzt worden. Zu diesem Zweck ist die kostenlose Software SoSci Survey verwendet worden (89). Diese bietet die Möglichkeit, mittels einer einfachen PHP-Codierung und der Erstellung eines Zufallsgenerators, verschiedene Bausteine einer Befragung zu randomisieren. Dieser Ansatz hat Verwendung gefunden, um den Aufklärungstext in Form einer Bilddatei an der entsprechenden Stelle bei der zufallsverteilten Hälfte der Befragten einzufügen. In diesem Zusammenhang wurde die Option „Uneingeschränkte Zufallsauswahl“ gewählt. Dies bedeutet, dass die gezogene Aufklärung nach jeder Befragung (Ziehung) zurück in den Zufallstopf gelegt wurde, sodass jede befragte Person mit einer Chance von 50% die Aufklärung über KI erhalten hat. Um das unbewusste Auslassen von einzelnen Fragen zu verhindern, wurde für alle Fragen die Option „bei fehlender Antwort nachhaken“ gewählt. Dadurch sollten möglichst vollständige Datensätze garantiert werden. Dennoch bestand die Option, eine Frage bewusst nicht zu beantworten, um keine Antworten zu erzwingen und somit das Ergebnis zu beeinflussen. Zudem bietet SoSci Survey die Option, die erhobenen Daten über die SPSS-Syntax in SPSS zu importieren, um eine statistische Auswertung durchzuführen.

Anders als bei den Ärzten, sind die Patientendaten mehrgleisig erhoben worden. Grund für die händische Umfrage war die Tatsache, dass die Patienten sowohl im stationären als auch im ambulanten Bereich der Hautklinik der Universität Mainz auf diese Weise unkompliziert befragt werden konnten. Außerdem sollten so auch diejenigen Patienten angesprochen werden, welche keine internetfähigen, mobilen Endgeräte besitzen, um die Befragung online auszufüllen. Zusätzlich sind in der Ambulanz und vor dem OP-Bereich Aushänge mit QR-Codes gemacht worden, um weitere Patientengruppen auch außerhalb der Zeiten der analogen Befragung zu erreichen. Die Randomisierung bei der Onlinebefragung erfolgte für die Patientenbefragung nach dem gleichen Prinzip wie bei der Ärztebefragung. Die Randomisierung der händischen Patientenbefragung wurde mittels eines gestaffelten Zufallsprinzips vollzogen. Dazu wurden die Fragebögen, welche zur Hälfte eine Aufklärung beinhalteten, mittels eines Münzwurfes in eine zufällige Reihenfolge gebracht. Anschließend wurden diese in der Reihenfolge an die Patienten auf der Station und in der Ambulanz verteilt, in der die Patienten dort eintrafen.

3.3.4 Validierung der Fragebögen

Um die Fragebögen zu validieren, ist vor dem Start der offiziellen Befragung eine kleine Gruppe testweise befragt worden. Dieser Gruppe gehörten Mitglieder unterschiedlichen Alters und Geschlechts an. Um die Diversität zusätzlich zu steigern, entstammte ein Teil der Befragten dem medizinischen Gesundheitssektor. Zudem war ein Teil der Testpersonen mit

dem Einsatz von KI im Bereich der Hautkrebsdiagnostik vertraut. Dem anderen Teil war das Thema KI bislang unbekannt. Auf diese Weise sollte das weite Spektrum des Patientenlientels simuliert werden. Im Hinblick auf die Befragung der Haut- und Hausärzte sind zudem Testpersonen aus dem medizinischen Bereich befragt worden, darunter Dermatologen und Studenten der Humanmedizin. In diesem Zusammenhang ist eine fehlerhafte Programmierung des Zufallsgenerators in der Arztbefragung aufgefallen. Diese hatte dafür gesorgt, dass die Befragung in der Hälfte der Fälle zwischendrin abbrach. Dieses Problem konnte daraufhin behoben werden. Grund dieses Fehlers war eine nachträgliche Anpassung der Fragereihenfolge in der Arztbefragung gewesen, welche jedoch nicht in dem zuvor programmierten Zufallsgenerator Berücksichtigung fand.

3.4 Aufbau der Fragebögen

Der Aufbau der händischen Befragung stimmte exakt mit dem der digitalen Befragung überein, weshalb bei der Patientenbefragung auf die Onlinebefragung im Folgenden nicht gesondert eingegangen wird. Für die Arztbefragung wird die Onlinebefragung erläutert. Durch die eingeklammerten Abkürzungen wird auf die Originalfragen der Fragebögen im Anhang verwiesen. PF steht dabei für Patientenfragebogen und AF für Arztfragebogen. Die nachfolgende Ziffer gibt die Nummerierung der Frage an und wird zum Teil durch einen kleingeschriebenen Buchstaben unterklassifiziert.

3.4.1 Patientenfragebogen (PF)

Der Fragebogen wird durch einen kurzen Informationstext eingeleitet, welcher das Thema der Befragung und die geschätzte Bearbeitungszeit angibt. Zudem wird definiert, in welchem Rahmen und Umfang die Befragung durchgeführt wird. Des Weiteren wird darauf verwiesen, dass die Erhebung der Daten anonym stattfindet, sodass keine Rückschlüsse auf einzelne Individuen möglich sind. Aus formalen Gründen wird zusätzlich begründet, warum die Befragung nicht gegendert worden ist. Schließlich wird noch darauf verwiesen, die Reihenfolge der Fragen und der Blätter bei dem Ausfüllen zu berücksichtigen, um das Ergebnis der Befragung nicht zu verfälschen.

Der erste Abschnitt der Befragung befasst sich unter anderem mit den epidemiologischen Daten der Patienten, welche an dieser Stelle ihr Geschlecht (PF1) sowie ihr Alter (PF2) angeben sollen. Es folgt die Frage nach dem höchsten Bildungsstand der Patienten (PF3). Anschließend wird das Thema KI eingeleitet. Zunächst sollen die Patienten angeben, ob sie schonmal etwas von KI gehört haben (PF4). An dieser Stelle wird die Abkürzung KI für „Künstliche Intelligenz“ eingeführt. Daraufhin werden die Patienten dazu aufgefordert, ihr Wissen in Bezug auf KI einzuschätzen (PF5). Schließlich wird noch erfragt, ob die Patienten KI in ihrem Alltag einsetzen (PF6).

An dieser Stelle erhält die Hälfte der Patienten eine zweiseitige Aufklärung zum Thema KI. Die andere Hälfte der Patienten springt direkt zum nächsten Abschnitt.

Der zweite Abschnitt beschäftigt sich mit dem Einsatz digitaler Medien durch die Patienten. Dabei wird auch auf das Krankheitsverhalten der Patienten eingegangen. Zunächst werden die Patienten gefragt, ob diese ein Smartphone oder ein anderes Internetfähiges Endgerät besitzen (PF 7). Für den Fall, dass die Patienten diese Frage verneinen, werden sie gebeten, die anschließenden, vertiefenden Fragen zu überspringen und mit PF11 weiterzumachen. Wer die Frage PF7 bejaht, soll als nächstes angeben, welches Gerät er am häufigsten in seiner Freizeit verwendet (PF8). Dafür stehen die Auswahloptionen Smartphone, Tablet, Computer und Sonstiges zur Verfügung. Daraufhin wird erfragt, wie oft auf das ausgewählte Medium zurückgegriffen wird, um sich über Krankheitssymptome zu informieren (PF9). Wird diese Anfrage mit „nie“ beantwortet, soll der Patient ebenfalls zu Frage PF11 springen. In allen

anderen Fällen soll der Patient angeben, worauf er bei der Informationssuche nach Krankheitssymptomen zurückgegriffen hat (PF10).

Im dritten Abschnitt der Befragung rückt das Thema Hautkrebs in den Mittelpunkt. Dazu werden zunächst einige grundlegende Daten erhoben. Als erstes wird der Patient gebeten, anzugeben, ob er Hautkrebs hat oder in der Vergangenheit hatte (PF11). Wird diese Frage mit „nein“ angekreuzt, so soll mit Frage PF14 fortgefahren werden. Als zweites wird erfragt, wie der Patient auf den Hautkrebs aufmerksam geworden ist (PF12). Schließlich soll im folgenden Schritt in Form einer kombinierten Mehrfachauswahl mit integriertem Freitext angegeben werden, um welche Form von Hautkrebs es sich handelt (PF13).

Im vierten Abschnitt werden die Themen KI und Hautkrebs zusammengeführt. In diesem Zusammenhang wird der Patient gebeten, sich in verschiedene hypothetische Situationen hineinzusetzen. Zuerst wird der Patient aufgefordert, seine Präferenz in Bezug auf die Form der Anwendung von KI beim eigenständigen Gebrauch anzugeben (PF14). An dieser Stelle besteht die Möglichkeit, auszuwählen, dass man unter keinen Umständen eine KI verwenden würde. In diesem Fall soll der Patient zu Frage PF16 springen. Wird der Gebrauch einer KI nicht verneint, folgt das erste hypothetische Fallbeispiel. Die Ausgangssituation ist, dass der Patient eine auffällige Hautveränderung an seinem Körper entdeckt (PF15). Zuerst soll angegeben werden, ob er sich zunächst eigenständig mittels KI untersuchen würde (PF15a). Wenn an dieser Stelle „auf keinen Fall“ ausgewählt wird, soll der Patient mit Frage PF16 weitermachen. Alle anderen Patienten fahren mit der Frage PF15 fort. Nachfolgend wird die Hypothese konkretisiert und die KI diagnostiziert die Hautveränderung einmal als Krebs (PF15b) und einmal als gutartig (PF15c). In beiden Fällen soll angegeben werden, ob der Patient in der jeweiligen Situation einen Arzt aufsuchen würde. Die zweite Hypothese bezieht sich auf einen Besuch beim Arzt. Der Patient hat erneut eine auffällige Hautläsion entdeckt, wobei er sich diesmal jedoch nicht mit einer KI untersucht hat (PF16). Es schließen sich vier konkretisierende Fragen an. Als erstes wird erfragt, ob es den Patienten verunsichern würde, wenn der Arzt zusätzlich eine KI einsetzen würde (PF16a). Im Anschluss soll angegeben werden, ob der Patient darauf bestehen würde, dass der Arzt eine KI einsetzt (PF16b). Bei der dritten Frage soll der Patient angeben, ob er auf die Diagnose der KI vertrauen würde, ohne vorher von einem Arzt untersucht worden zu sein (PF16c). Zuletzt muss sich der Patient entscheiden, wem er eher vertrauen würde, wenn Arzt und KI bei ihrer Diagnose zu unterschiedlichen Ergebnissen kämen (PF16d). Abgeschlossen wird dieser Abschnitt mit der Frage, ob der Patient den Einsatz von KI im Rahmen der Hautkrebsdiagnostik befürworten würde (PF17).

Im letzten Abschnitt wird unter anderem auf verschiedene Teilaspekte eingegangen, welche es beim Einsatz von KI zu berücksichtigen gilt. Zunächst wird nochmal das Thema „KI im Hausgebrauch“ aufgegriffen. An dieser Stelle sollen sich die Patienten zwischen vier verschiedenen Aspekten entscheiden und priorisieren, was ihnen davon in Bezug auf den eigenständigen Gebrauch von KI wichtiger wäre (PF18). Im Anschluss sollen die Patienten angeben, ob sie den eigenständigen Gebrauch von KI an Verwandte oder Bekannte weiterempfehlen würden (PF19). Zum Schluss der Befragung wird der Umgang mit Daten thematisiert. Der Patient wird aufgefordert, anzugeben, ob er Bedenken um die Sicherheit seiner Daten habe, welche im Rahmen der Diagnostik angesammelt würden (PF20) und ob er seine Daten für das Training von KI zur Verfügung stellen würde (PF21). Schließlich endet die Befragung mit einem Freitextfeld, in dem der Patient Anmerkungen hinterlassen kann (PF22) (siehe Anlage 8.3).

3.4.2 Arztfragebogen

Der Arztfragebogen beginnt ebenfalls mit einer Einleitung, in welcher das Thema der Befragung und die geschätzte Bearbeitungszeit vorgestellt werden. Im Anschluss wird die Rahmenbedingung sowie der Umfang der Befragung erläutert, auf die Anonymität der Daten verwiesen sowie das Unterlassen des Genderns begründet. Die Bitte um die Einhaltung der

Fragereihenfolge fällt an dieser Stelle weg, da im Rahmen der Onlinebefragung das Aufrufen bereits bearbeiteter Fragen nicht möglich ist.

Der erste Abschnitt beginnt wie bei der Patientenbefragung mit der Erhebung epidemiologischer Daten. In diesem Rahmen sollen Geschlecht (AF1) und Alter (AF2) angegeben werden. Anschließend werden die Ärzte aufgefordert, ihre Fachrichtung anzugeben (AF3). Zudem sollen sie zusätzlich ihre Berufserfahrung in ihrer derzeitig ausgeübten Fachrichtung darlegen (AF4) und angeben, ob sie in ihrer derzeitig ausgeübten Fachrichtung einen Facharzt besitzen (AF5). Es folgt die Frage nach der Häufigkeit, mit der die Ärzte in ihrem Alltag mit der Hautkrebsdiagnostik konfrontiert sind (AF6). Analog zu der Patientenbefragung soll anschließend angegeben werden, ob die Ärzte schonmal etwas von KI gehört haben (AF7), und wie sie ihr Wissen in Bezug auf KI einschätzen (AF8). Schließlich soll angegeben werden, ob die Ärzte aktuell KI im Rahmen der Hautkrebsdiagnostik einsetzen (AF9). Darüber hinaus wird erfragt, ob sie den Einsatz von KI im Rahmen der Hautkrebsdiagnostik befürworten (AF10).

Es folgt die mittels eines Zufallsgenerators in SoSci Survey eingesetzte Aufklärung der einen Hälfte der Ärzte.

Im zweiten Abschnitt wird die Mediennutzung der Ärzte erfragt, wobei ein wichtiges Augenmerk auf dem Bezug zum individuellen Krankheitsverhalten liegt. Zuerst sollen die Ärzte angeben, ob sie ein Smartphone oder ein anderes internetfähiges Endgerät besitzen (AF11). Wenn diese Frage mit „nein“ beantwortet wird, wird man direkt zu Frage AF15 weitergeleitet. Alle anderen Ärzte werden vertiefend gefragt, welches internetfähige Gerät sie am häufigsten in ihrer Freizeit einsetzen (AF12). Im Anschluss sollten sie angeben, wie oft sie dieses Medium nutzen, um sich über Krankheitssymptome zu informieren (AF13). Wird diese Frage mit „nie“ beantwortet, werden die Ärzte direkt zu Frage AF15 weitergeleitet. Der Rest der Ärzte wird nachfolgend befragt, welche Programme zu der Informationssuche genutzt worden sind (AF14).

Im dritten Abschnitt wird der Arzt dazu aufgefordert, sich in verschiedene hypothetische Situationen hineinzusetzen, welche im Rahmen der Anwendung von KI durch den Patienten in Bezug auf die Hautkrebsdiagnostik denkbar wären. Die Ausgangssituation der ersten Frage ist, dass der Patient eine auffällige Hautveränderung an sich selbst entdeckt (AF15). Zunächst sollen die Ärzte ankreuzen, ob sie dem Patienten empfehlen würden, einen Arzt aufzusuchen (AF15a). Wenn diese Frage mit „auf keinen Fall“ beantwortet wird, wird man zu Frage AF16 weitergeleitet. Alle anderen sollen mit der Frage AF15 fortfahren. Anschließend untersucht sich der Patient selbstständig mittels einer KI. In einem Fall stellt die KI eine bösartige (AF15b) und in einem anderen Fall eine gutartige (AF15c) Läsion fest. Für beide Fälle sollen die Ärzte angeben, ob sie dem Patienten jeweils empfehlen würden, einen Arzt aufzusuchen.

Anschließend wird der Einsatz von KI durch den Arzt thematisiert. Als erstes wird gefragt, ob die Ärzte KI als diagnostisches Hilfsmittel für ihre Untersuchungen einsetzen würden (AF16). Wer diese Frage mit „auf keinen Fall“ beantwortet, wird zu Frage AF20 weitergeleitet. Alle anderen werden mit weiteren Hypothesen konfrontiert. Zum einen wird nach dem Vorgehen der Ärzte gefragt, wenn diese eine therapeutisch relevante Diagnose fällen, die KI dies allerdings verneint (AF17). Zum anderen wird der umgekehrte Fall präsentiert, bei dem die KI eine therapeutisch relevante Diagnose trifft, die die Ärzte jedoch verneinen (AF18). Nachfolgend sollen die Ärzte angeben, ob sie auf die Diagnose einer KI vertrauen würden, ohne den Patienten selbst untersucht zu haben (AF19). Danach werden die Ärzte mit der Situation konfrontiert, dass der Patient darauf besteht, dass der Arzt eine KI einsetze (AF20). Die Ärzte werden gefragt, inwieweit sie dieser Aufforderung nachkommen würden.

Im letzten Abschnitt werden verschiedene Aspekte in Bezug auf den Einsatz von KI im Rahmen der Hautkrebsdiagnostik thematisiert. Zum einen wird gefragt, ob die Ärzte es als Voraussetzung betrachten würden, dass ein Arzt die Funktionsweise einer KI kennt, bevor er diese einsetzen darf (AF21). Zum anderen wird die Meinung der Ärzte erfragt, ob KI zum

Bestandteil ihrer Facharztausbildung werden sollte (AF22). Nachfolgend soll das Potential von KI in Bezug auf die Hautkrebsdiagnostik eingeschätzt werden (AF23). Daraufhin wird die Möglichkeit dargeboten, Patientendaten nach der Zustimmung durch die Patienten zum Training von KI zur Verfügung zu stellen (AF24). Es folgt die Frage, welche Eigenschaft einer KI den Ärzten am wichtigsten wäre (AF25), in die neben einer Auswahl auch eine Freitextantwortmöglichkeit integriert ist. Anschließend werden vier verschiedene Aspekte einer KI in Bezug auf ihre Anwendung dargeboten, welche nach ihrer Wichtigkeit eingestuft werden sollen (AF26). Danach werden zwei Freitextfelder angeboten, in die jeweils die größte Stärke (AF27) beziehungsweise Schwäche der KI (AF28) eingetragen werden soll. Zum Abschluss besteht die Möglichkeit, Anmerkungen zu hinterlassen (AF29) (siehe Anlage 8.4).

3.5 Durchführung der Befragung

3.5.1 Fallzahlbestimmung

Die Fallzahlbestimmung erfolgte in Kooperation mit dem Institut für Medizinische Biometrie, Epidemiologie und Informatik der Universitätsmedizin Mainz.

In Bezug auf die Befragung der Ärzte wurden folgende Fragen und erwartete Antwortkonstellationen zugrunde gelegt: Zum einen wurde geschätzt, dass nicht aufgeklärte Ärzte die Frage, ob sie KI als diagnostisches Hilfsmittel einsetzen würden (AF16), im Schnitt mit einer 2 ankreuzen würden. Für aufgeklärte Ärzte ist eine 4 geschätzt worden. Die Zahlen beziehen sich in diesem Zusammenhang und im weiteren Verlauf auf die 5 abgestuften Antwortmöglichkeiten im Sinne einer Likert-Skala, mit 1 als niedrigstem und 5 als höchstem Wert. Ebenso ist die Frage AF21 berücksichtigt worden, welche erschließt, ob die Ärzte das Verständnis in Bezug auf die Funktionsweise von KI als Voraussetzung erachten, um diese einzusetzen. Hier wurde die gleiche Antwortkonstellation wie bei der vorherigen Frage als Annahme zugrunde gelegt. Schließlich wurde auch die Frage, wie die Ärzte das Potential von KI in Bezug auf die Hautkrebsdiagnostik einschätzen (AF23) mit in die Berechnung aufgenommen. In diesem Fall lagen drei abgestufte Antwortmöglichkeiten vor. Aus diesen Angaben errechnete sich schließlich eine Fallzahl von 16 Ärzten, welche nötig ist, um bei einem Signifikanzniveau von $\alpha=0.05$ eine Power von etwa 0.837 zu erreichen.

Für die Befragung der Patienten sind zwei Fragen in die Fallzahlberechnung eingeflossen. Dies war zum einen die Frage, ob sich die Patienten, sofern diese eine auffällige Hautveränderung an sich entdeckten, eigenständig mittels eines KI gestützten Programms untersuchen würden (PF15a). Von Patienten, welche eine Aufklärung erhalten hatten, wurde erwartet, dass sie diese Frage mit einer 4 beantworten würden. Für die nicht aufgeklärte Patientengruppe ist eine 2 bis 3 geschätzt worden. Zum anderen fand die Frage nach der Befürwortung des Einsatzes von KI zur Hautkrebsdiagnostik in der Klinik Einfluss in die Fallzahlbestimmung. Dabei wurde angenommen, dass nicht aufgeklärte Patienten diese Frage mit einer 3 beantworten würden, während aufgeklärte Patienten im Schnitt eine 5 angeben würden. Diese Überlegungen führten schließlich zu einer Fallzahl von 26 Patienten, um bei einem Signifikanzniveau von $\alpha=0.05$ eine Power von etwa 0.816 zu erreichen.

3.6 Statistische Auswertung

Die statistische Auswertung erfolgte mittels SPSS. Der überwiegende Anteil der Fragen wurde deskriptiv ausgewertet. In diesem Fall beziehen sich die prozentualen Anteile (Zahlen in Klammern) auf die Personen, welche die Frage beantwortet haben. Für den Fall, dass Angaben zum Mittelwert gemacht worden sind, ist jeweils die Standardabweichung (SD für standard deviation) nachfolgend in Klammern aufgeführt worden. Fragen nach dem Prinzip einer Likertskala weisen, wenn nicht anders benannt, Werte von 1 bis 5 auf, mit 3 als Mittelpunkt, welcher genau zwischen den Extrema liegt. Fragen besonderer Relevanz so wie

jene, die zur Fallzahlbestimmung herangezogen worden sind, wurden auf ihre statistische Signifikanz hin überprüft. Da die Aufklärung eine unabhängige Variable darstellt und die abhängigen Variablen ordinalskaliert sind, wurde zu diesem Zweck der Mann-Whitney-U-Test herangezogen. Zuvor wurde mittels des Kolmogorov-Smirnov-Tests für zwei Stichproben getestet, ob sich die beiden Verteilungen in ihrer Form unterscheiden. Die Bestimmung der Signifikanz unter den Fragen mit dichotomen Variablen erfolgte mittels eines Chi-Quadrat-Tests.

3.7 Datenschutz

Die Befragung der Patienten und der Ärzte erfolgte anonym. Es sind keine personenidentifizierenden Daten im Sinne der Definition der DSGVO Art. 4 erhoben worden. Die digitalen Daten sind in einem passwortgeschützten Rechner gesichert worden, die analogen Daten in einem nicht öffentlich zugänglichen und schlossgesicherten Raum aufbewahrt worden.

4 Ergebnisse

4.1 Ergebnisse der Patientenbefragung

Insgesamt haben 109 Patienten an der Befragung teilgenommen, darunter 54 (49,5%) Männer und 54 (49,5%) Frauen. Eine (0,9%) befragte Person gab an, divers zu sein. Aufgrund des geringen Anteils findet die Kategorie „divers“ in den anschließenden, geschlechtsspezifischen Auswertungen keine statistische Berücksichtigung. Das Durchschnittsalter der befragten Patienten lag bei 55,57 Jahren ($SD=18,56$) mit einer Altersspanne von 19 bis 88 Jahren. Betrachtet man die geschlechtsspezifische Altersverteilung, so waren die Frauen mit 52 Jahren ($SD=18,19$) im Schnitt ca. 7 Jahre jünger als die Männer, deren Durchschnittsalter 59 Jahre ($SD=18,36$) beträgt (siehe **Abbildung 2**).

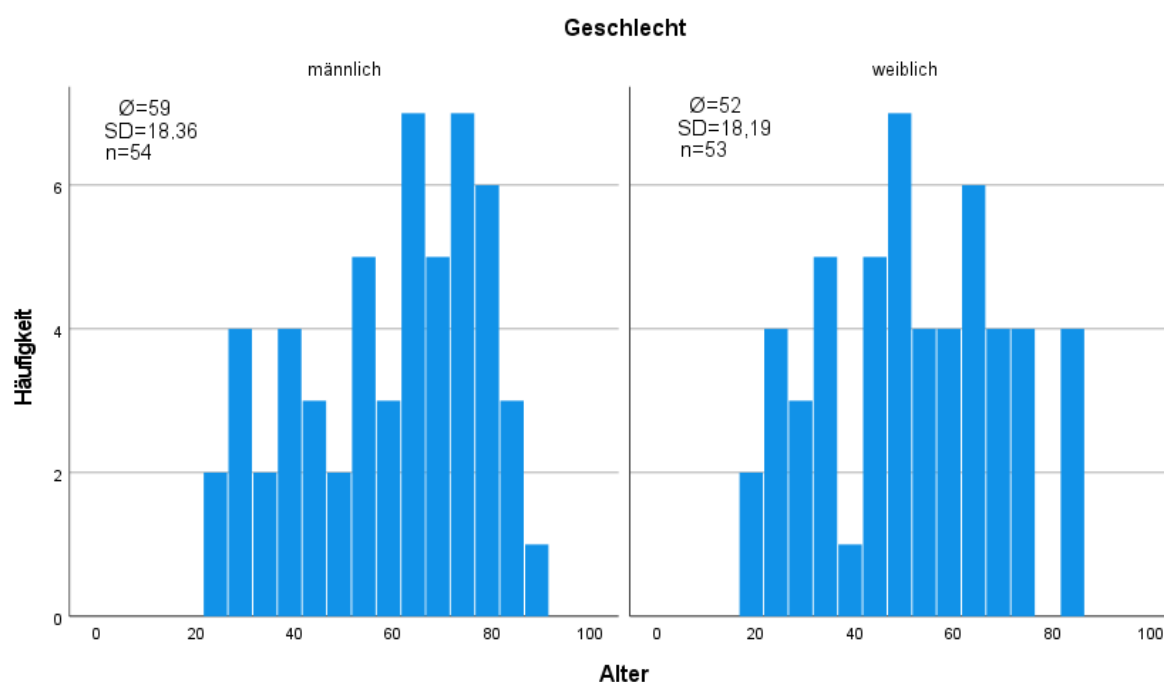


Abbildung 2: Säulendiagramme zur geschlechterspezifischen Altersverteilung der Patienten

Bei den Angaben zu dem höchsten Bildungsabschluss der Patienten zeigt sich ein gemischtes Bild. Jeweils 30 (27,8%) Patienten haben angegeben, ein abgeschlossenes Studium oder eine Ausbildung zu haben. 14 (13,0%) Patienten hatten einen Hauptschulabschluss, 13 (12,0%) ein (Fach-)Abitur und 12 (11,1%) einen Realschulabschluss. 8 (7,4%) Patienten wählten die Kategorie sonstiges und eine Person (0,9%) gab an, noch Schüler zu sein (siehe **Tabelle 4**).

88 (82,2%) Patienten gaben an, schonmal etwas von KI gehört zu haben. Des Weiteren sagten 32 (30,8%) Patienten, dass sie KI in ihrem Alltag nutzen würden. 38 (36,5%) Patienten verneinten den Gebrauch von KI im Alltag und 34 (32,7%) Patienten wussten nicht, ob sie KI in ihrem Alltag verwendeten (siehe **Tabelle 5**).

5 (4,8%) Patienten gaben an, dass sie ein sehr gutes Wissen und 16 (15,4%) Patienten sagten, dass sie ein gutes Wissen in Bezug auf die Funktionsweise von KI hätten. Dahingegen gaben 30 (28,8%) Patienten an, keine Kenntnisse zu der Funktionsweise von KI zu haben und 29 (27,9%) sagten, dass ihr Wissen zu der Funktionsweise gering ausgeprägt sei (siehe **Tabelle 6**). Im Durchschnitt schätzen die Patienten ihr Wissen auf eine 2,39 ($SD=1,19$). An dieser Stelle zeigt sich ein geschlechterabhängiger Unterschied. Frauen gaben im Schnitt eine 2,16 ($SD=1,07$) und Männer eine 2,65 ($SD=1,27$) an, womit die Männer ihr Wissen im Mittel höher einschätzten. (siehe **Abbildung 3**).

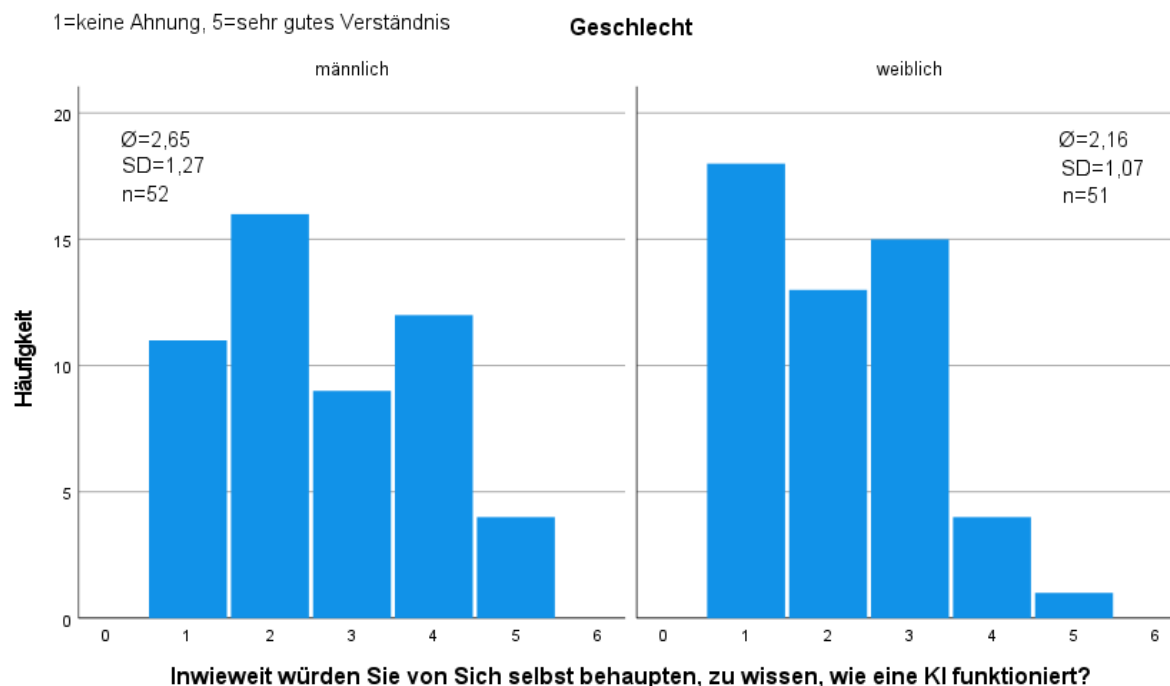


Abbildung 3: Säulendiagramme zur geschlechterspezifischen Einschätzung der Patienten in Bezug auf ihr Wissen um die Funktionsweise von KI

55 (50,9%) Patienten erhielten im Laufe der Befragung einen Aufklärungstext zu KI, während die restlichen 53 (49,1%) Patienten keinen Aufklärungstext bekamen. Im Mittel schätzen die Patienten ihr Vorwissen in Bezug auf KI in der Stimulusgruppe auf eine 2,48 (SD=1,23) und in der Kontrollgruppe auf eine 2,31 (SD=1,16), womit es keinen relevanten Unterschied zwischen diesen Gruppen gab.

100 (93,5%) Patienten gaben an, ein Smartphone oder ein anderes internetfähiges Endgerät zu besitzen. 58 (74,4%) Patienten, welche solch ein Endgerät besitzen, sagten im weiteren Verlauf, dass sie in ihrer Freizeit am häufigsten das Smartphone nutzen würden, gefolgt von dem Computer mit 13 (16,7%) Stimmen. 31 (28,4%) Patienten ließen diese Frage unbeantwortet beziehungsweise beantworteten diese uneindeutig. Insgesamt gaben 68 (87,2%) Patienten an, dass zuvor ausgewählte Medium zu nutzen (Skalenwerte von 2 bis 5), um sich über Krankheitssymptome zu informieren (siehe **Tabelle 7**). Im Mittel gaben die Patienten einen Wert von 3,19 (SD=1,32) an (siehe **Abbildung 4**).

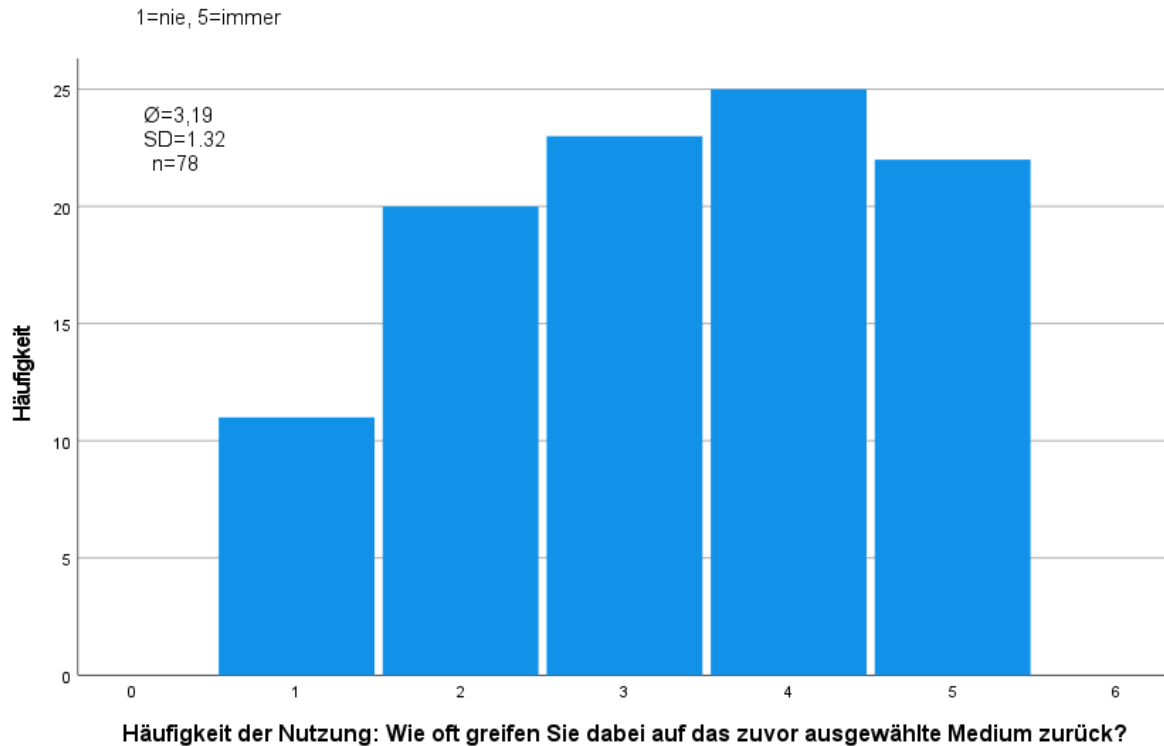


Abbildung 4: Säulendiagramm zur Häufigkeit der Mediennutzung der Patienten zur Recherche von Krankheitssymptomen

Mit 87 (87,9%) Patienten hat die Mehrheit der befragten, mobilen Endgerätbesitzer angegeben, auf Internetsuchmaschinen wie Google zurückzugreifen, um sich über Krankheitssymptome zu informieren. 12 Patienten (12,1%) nutzen Gesundheitsapps und 3 (3%) gaben an, spezielle Diagnoseprogramme für den Computer einzusetzen (siehe **Tabelle 8**).

58 (54,2%) Patienten haben bis zu dem Zeitpunkt der Befragung wenigstens eine Form von Hautkrebs gehabt (siehe **Tabelle 9**). In 16 Fällen (32,0%) ist der Hautkrebs durch den Hautarzt entdeckt worden. In 2 (4,0%) Fällen haben Bekannte oder Verwandte den Hautkrebs entdeckt und in 12 (24,0%) Fällen sind die Patienten selbstständig auf den Hautkrebs aufmerksam geworden. Durch den Hausarzt sind 11 (22,0%) und durch andere Fachärzte 9 (18,0%) Fälle entdeckt worden (siehe **Tabelle 1**). 8 Patienten haben keine oder mehr als eine Option angekreuzt.

Wie sind Sie auf den Hautkrebs aufmerksam geworden?

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	selbstständig	12	11,0	24,0	24,0
	durch Bekannte/Verwandte	2	1,8	4,0	28,0
	durch meinen Hausarzt	11	10,1	22,0	50,0
	durch meinen Hautarzt	16	14,7	32,0	82,0
	durch einen anderen Facharzt	9	8,3	18,0	100,0
	Gesamt	50	45,9	100,0	
Fehlend	System	59	54,1		
Gesamt		109	100,0		

Tabelle 1: Entdeckung von Hautkrebs bei den Patienten

Mit 28 (48,3%) Fällen war das Maligne Melanom die am häufigsten vertretene singuläre Entität. An zweiter Stelle kam das Basaliom mit 14 (24,1%), gefolgt von dem Spinaliom mit 3 (5,2%) Fällen. 4 (6,9%) Patienten haben zwei verschiedenen Arten von Hautkrebs gehabt (siehe **Tabelle 10**).

31 (32,0%) Patienten gaben an, dass sie KI bevorzugt in Form einer App für das Smartphone oder Tablet nutzen würden. 29 (29,9%) Patienten entschieden sich für ein selbstständiges Endgerät und 14 (14,4%) würden einen Einsatz am Computer bevorzugen (siehe **Tabelle 11**). 6 (12,0%) Patienten aus der Stimulusgruppe gaben an, dass sie unter keinen Umständen eine KI einsetzen würden, während in der Kontrollgruppe 17 (36,2%) Patienten den Einsatz von KI ausschlossen (siehe **Abbildung 5**). An dieser Stelle zeigte sich im Chi-Quadrat-Test, keine erwarteten Zelhäufigkeiten waren kleiner als 5, ein statistisch signifikanter Zusammenhang zwischen der Aufklärung und der Einstellung, KI einzusetzen ($\chi^2(1) = 7,82, p < 0,006$ $\phi = 0,28$).

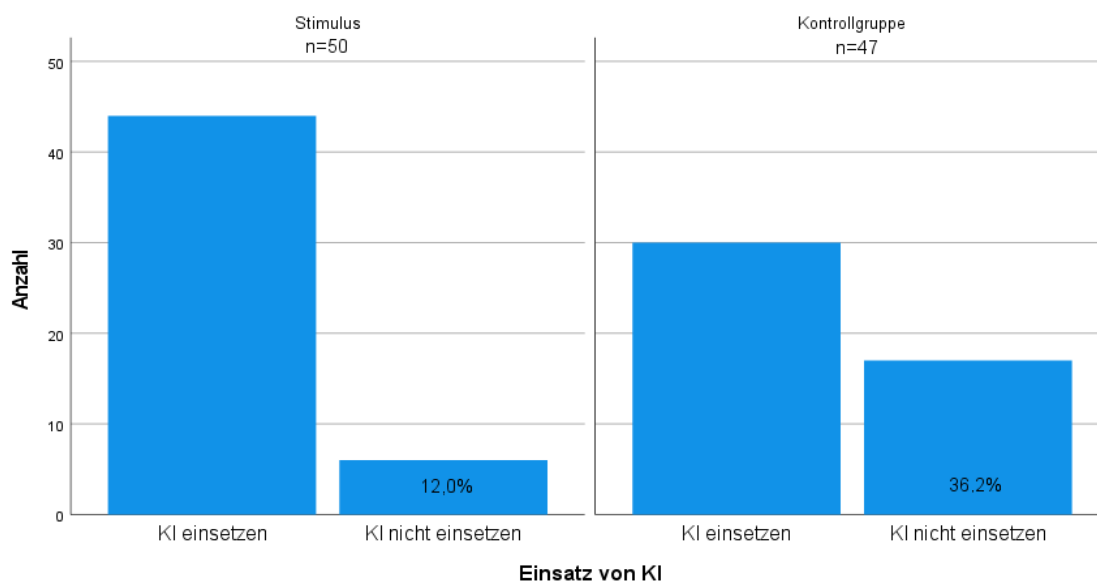
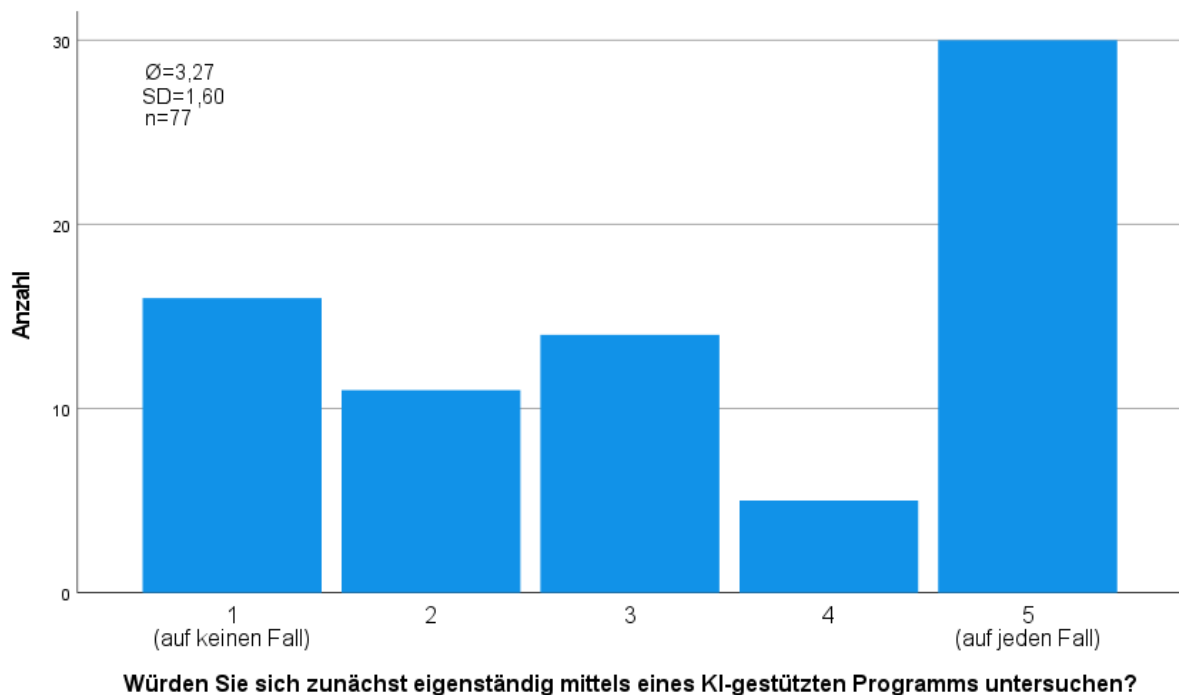


Abbildung 5: Säulendiagramme zu der Bereitschaft in den Subgruppen "Stimulus" und "Kontrollgruppe" KI einzusetzen

Die folgende Frage wurde nur von denjenigen Patienten ($n = 77$) beantwortet, die bei der vorherigen Frage den eigenständigen Einsatz von KI nicht grundsätzlich ausschlossen. Bei der expliziten Frage nach der Bereitschaft, sich eigenständig mittels KI zu untersuchen, gaben die Patienten im Mittel eine 3,27 ($SD=1,60$) an (siehe **Abbildung 6**). 16 (20,8%) Patienten sagten, dass sie auf keinen Fall eine KI zur Eigendiagnostik einsetzen würden. 12 (15,6%) Patienten würden sehr wahrscheinlich keine KI einsetzen und 14 (18,2%) waren weder für noch gegen den Einsatz von KI zur Selbstdiagnostik. 5 (6,5%) Patienten waren eher dazu geneigt, eine KI einzusetzen und 30 (39,0%) würden sich auf jeden Fall selbständig mittels einer KI untersuchen wollen (siehe **Tabelle 12**). Bei höherem Vorwissen (Werte von 4 und 5) zeigt sich mit einer Zustimmung von 58,8% eine höhere Bereitschaft für den selbständigen Einsatz von KI (siehe **Tabelle 13**).



Würden Sie sich zunächst eigenständig mittels eines KI-gestützten Programms untersuchen?

Abbildung 6: Säulendiagramm zu der Bereitschaft der Patienten, KI zur Selbstdiagnostik einzusetzen

Die Verteilungen in der Stimulus- und der Kontrollgruppe unterschieden sich voneinander, Kolmogorov-Smirnov $p < 0,05$. Der Mann-Whitney-U-Test ergab keinen statistisch signifikanten Unterschied zwischen der Bereitschaft in der Stimulus- und der Kontrollgruppe ($U = 697,00$, $Z = -0,077$, $p > 0,05$). Allgemein zeigten sich in Abhängigkeit vom Geschlecht und der Hautkrebsanamnese nur geringe Unterschiede in Bezug auf die Bereitschaft, sich selbstständig mittels KI zu untersuchen.

Für den Fall, dass die KI im eigenständigen Gebrauch eine Hautveränderung als bösartig einstuft, würden 58 (96,7%) Patienten auf jeden Fall ihren Arzt aufsuchen gehen. Wenn die Läsion hingegen als gutartig eingestuft würde, würden 44 (72,1%) Patienten auf jeden Fall zum Arzt gehen (siehe **Tabelle 14**). Diese Frage wurde nur von denjenigen Patienten beantwortet, welche den Einsatz von KI bei der vorherigen Frage nicht grundsätzlich ausschlossen.

10 (20,0%) Patienten aus der Kontrollgruppe gaben an, dass es sie auf jeden Fall verunsichern würde, wenn der Arzt zusätzlich eine KI zur Untersuchung einsetzen würde. In der Stimulusgruppe waren es 3 (5,7%). Aus der Kontrollgruppe würden 24 (48,0%) Patienten und aus der Stimulusgruppe 35 (66,0%) Patienten keine Verunsicherung bei einem unterstützenden Einsatz von KI durch den Arzt empfinden (siehe **Abbildung 7**).

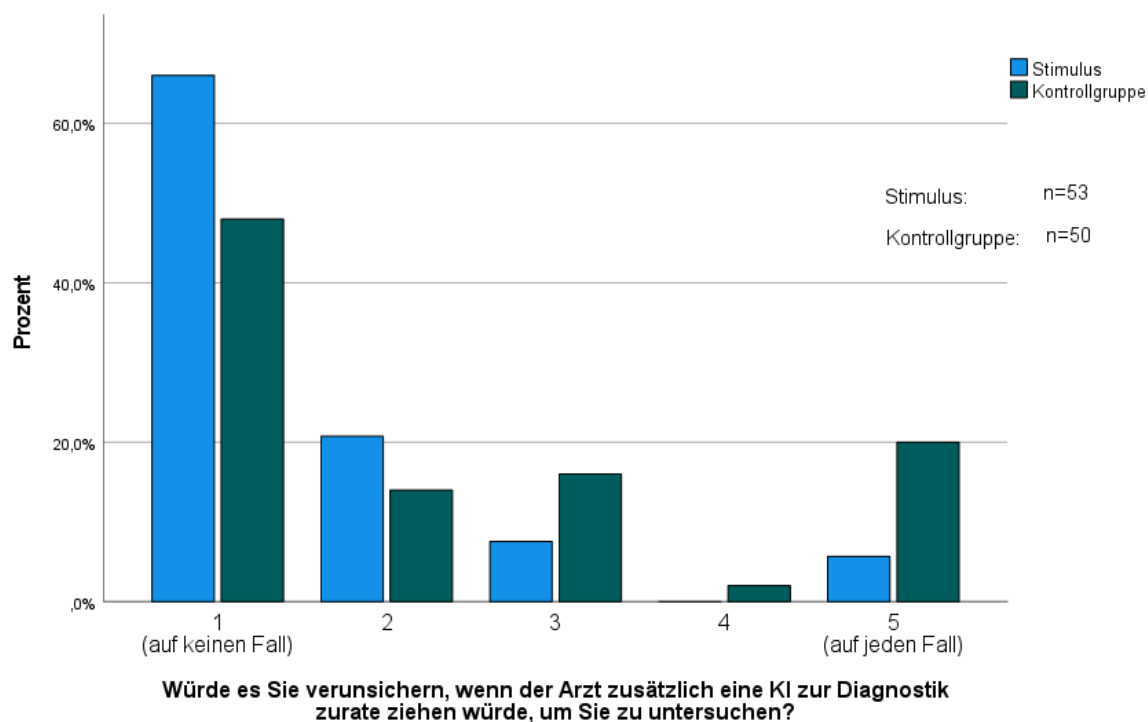


Abbildung 7: Säulendiagramm zu der Verunsicherung der Patienten durch den unterstützenden Einsatz von KI beim Arzt

Die Verteilungen in der Stimulus- und der Kontrollgruppe unterschieden sich voneinander, Kolmogorov-Smirnov $p < 0,05$. Der Mann-Whitney-U-Test ergab einen statistisch signifikanten Unterschied zwischen der Verunsicherung in der Stimulus- und der Kontrollgruppe ($U = 997,50$, $Z = -2,411303$, $p < 0,015$).

30 (29,1%) Patienten sagten, dass sie auf jeden Fall oder eher auf den Einsatz von KI durch den Arzt bestehen würden. 15 (14,6%) Patienten sagten hingegen, dass sie auf keinen Fall auf den Einsatz bestünden (siehe **Tabelle 15**). Im Mittel ist eine 2,99 (SD=1,26) angegeben worden. Der Mittelwert betrug in der Stimulusgruppe 3,04 (SD=1,23) und in der Kontrollgruppe 2,94 (SD=1,31).

51 (49,5%) Patienten gaben an, auf keinen Fall allein auf die Diagnose einer KI vertrauen zu würden und 19 (18,4%) Patienten sagten, dass sie eher nicht auf die alleinige Diagnose einer KI vertrauen würden. 8 (7,8%) Patienten sagten, dass sie allein oder eher allein auf die Diagnose einer KI vertrauen würden. Im Mittel gaben die Patienten eine 1,94 an (SD=1,12) (siehe **Abbildung 8**). In der Kontrollgruppe waren es 29 (56,9%) Patienten, die nicht allein auf die Diagnose einer KI vertrauen würden. Im Vergleich dazu waren es in der Stimulusgruppe 22 (42,3%) (siehe **Tabelle 16**).

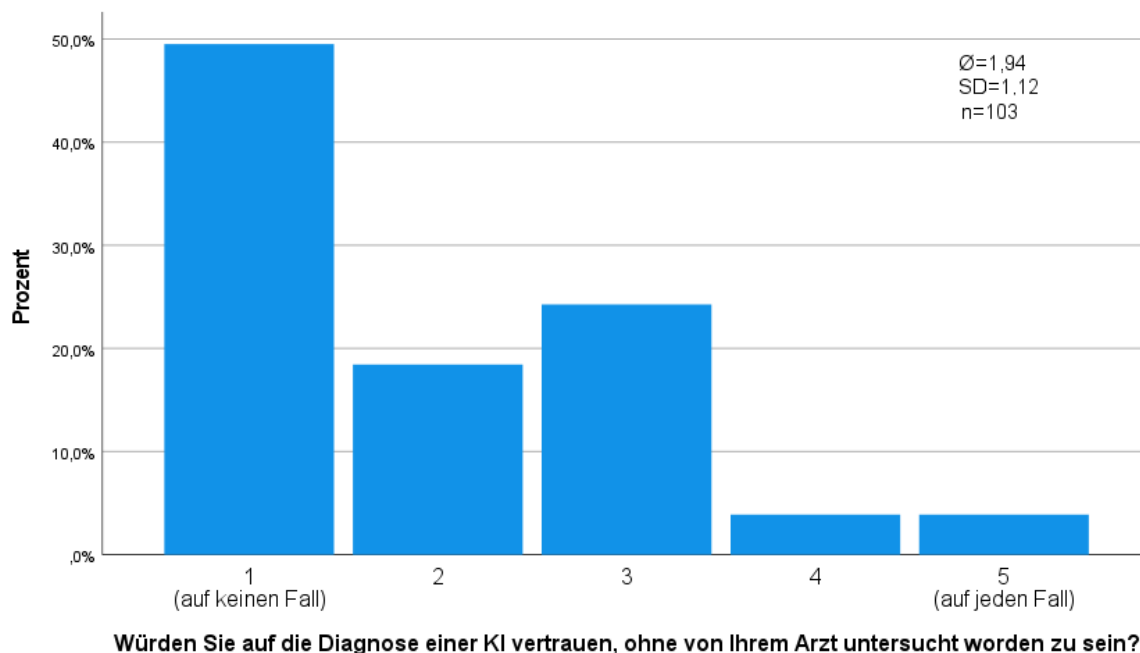


Abbildung 8: Säulendiagramm zu dem Vertrauen von Patienten in KI

Für den Fall, dass die KI und der Arzt zu unterschiedlichen Untersuchungsergebnissen kämen, würden 22 (45,8%) Patienten aus der Stimulusgruppe auf jeden Fall auf den Arzt vertrauen. In der Kontrollgruppe gaben 28 (59,6%) Patienten an, auf jeden Fall auf den Arzt zu vertrauen (siehe **Tabelle 17**). Im Mittel gaben die Patienten in der Kontrollgruppe eine 1,62 ($SD = 0,82$) und in der Stimulusgruppe eine 1,88 ($SD = 0,94$) an (1=Arzt vertrauen, 5=KI vertrauen). Kein Patient sagte, auf jeden Fall auf die KI vertrauen zu wollen.

Insgesamt sagten 38 (36,9%) Patienten, dass sie den Einsatz von KI in der Klinik auf jeden Fall befürworten würden, 25 (24,3%) Patienten würden dazu tendieren. 3 (2,9%) Patienten sagten hingegen, dass sie diesen auf keinen Fall befürworten würden und 6 (5,8%) Patienten waren eher dazwischen (siehe **Abbildung 9**).

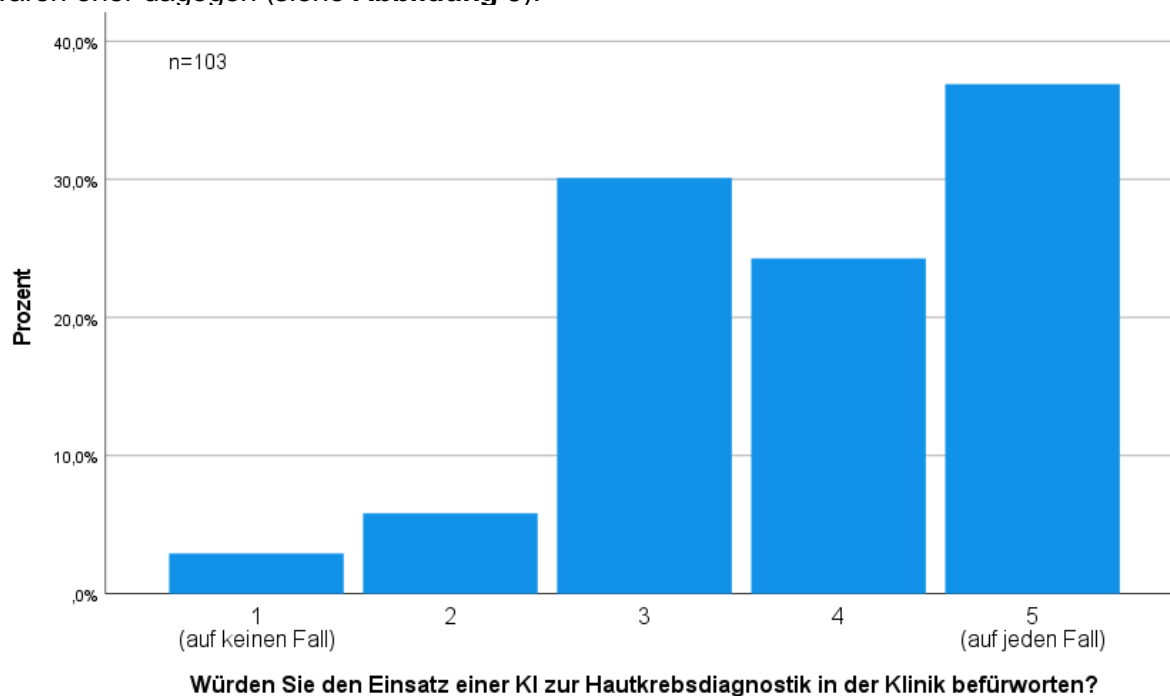


Abbildung 9: Säulendiagramm zu dem Einsatz von KI in der Klinik aus Sicht der Patienten

In der Stimulusgruppe lag der Mittelwert bei 3,94 (SD=1,06) und in der Kontrollgruppe bei 3,78 (SD=1,10). Der Mittelwert betrug unter den Frauen 3,84 (SD=1,07) und unter den Männern 3,86 (SD=1,10). Die Mittelwerte der Patienten in den Subgruppen älter oder jünger gleich 50 Jahre und in den Subgruppen Hautkrebs oder kein Hautkrebs zeigten keinen relevanten Unterschied. In der Stimulusgruppe sagten 19 (36,5%) Patienten, dass sie den Einsatz von KI in der Klinik auf jeden Fall befürworten würden, 17 (32,7%) weitere Patienten tendierten ebenfalls dazu. Dahingegen waren es in der Kontrollgruppe 19 (37,3%) Patienten, welche den Einsatz auf jeden Fall befürworteten und 8 (15,7%), die dazu neigten, KI in der Klinik einzusetzen (siehe **Abbildung 10**). Insgesamt zeigte sich kein signifikanter Unterschied zwischen diesen beiden Gruppen. Die Verteilungen in der Stimulus- und der Kontrollgruppe unterschieden sich voneinander, Kolmogorov-Smirnov $p < 0,05$. Der Mann-Whitney-U-Test ergab keinen statistisch signifikanten Unterschied zwischen der Befürwortung in der Stimulus- und der Kontrollgruppe ($U = 1158,50$, $Z = -0,999$, $p > 0,05$). Bei hohem Vorwissen (Werte von 4 und 5) lag die Zustimmung bei 65,0% (siehe **Tabelle 18**).

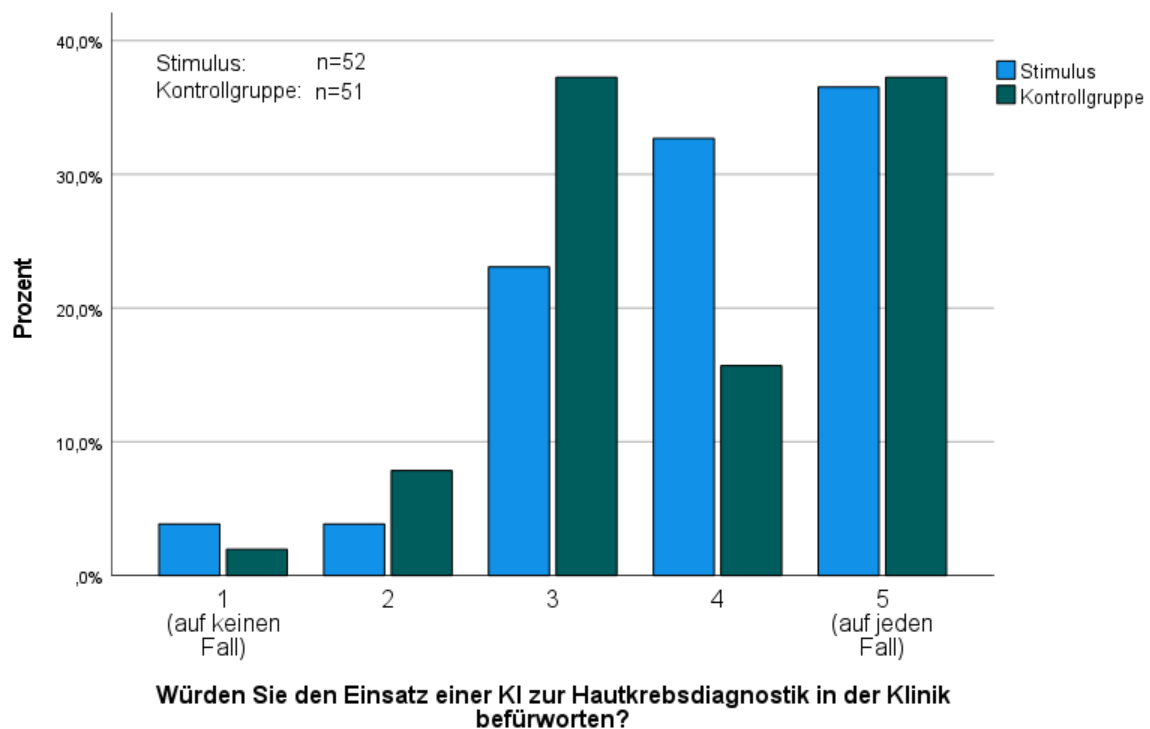


Abbildung 10: Säulendiagramm zu dem Einsatz von KI in der Klinik aus Sicht der Patienten in den Subgruppen "Stimulus" und "Kontrollgruppe"

Die Patienten gaben an, dass ihnen die Diagnosesicherheit für den häuslichen Einsatz einer KI am wichtigsten ist. Diese wurde 16-mal (42,1%) auf Rang 1 gewählt. Auf Rang zwei folgt die „anschließend schnelle Terminvermittlung zum Arzt“, welche 19-mal (47,5%) für diese Position ausgewählt worden ist. „Jederzeit Verfügbarkeit“ landete 15-mal (39,5%) auf Rang 3 und „Erstattung durch die Krankenkasse“ wurde in 22 Fällen (57,9%) der Fälle auf den hintersten Rang gewählt. 71 Patienten haben diese Frage nicht, unvollständig oder falsch ausgefüllt (siehe **Tabelle 19**).

11 (20,8%) Patienten aus der Stimulusgruppe würden den eigenständigen Gebrauch von KI auf jeden Fall weiterempfehlen, 18 (34,0%) weitere Patienten tendierten ebenfalls dazu. In der Kontrollgruppe ergaben sich Vergleichswerte von 10 (19,6%) und 6 (11,8%) (siehe **Tabelle 20**). Im Mittel gaben die Patienten aus der Stimulusgruppe einen Wert von 3,51 (SD=1,15) an, bei den Patienten aus der Kontrollgruppe ergab sich ein Mittelwert von 3,16 (SD=1,22).

16 (15,2%) Patienten sagten, Bedenken wegen der Datensicherheit im Zusammenhang mit dem Einsatz von KI zu haben und 15 (14,3%) Patienten gaben, dass sie eher Bedenken wegen

der Datensicherheit hätten. Dahingegen gaben 30 (28,6%) Patienten an, auf keinen Fall Bedenken wegen der Datensicherheit zu haben und 19 (18,1%) sagten, eher keine Bedenken zu haben (siehe **Tabelle 21**).

In der Stimulusgruppe lag die Bereitschaft, die eigenen Daten für das Training von KI zur Verfügung zu stellen im Schnitt bei 4,21 (SD=1,21) und in der Kontrollgruppe bei 3,85 (SD=1,43). 32 (60,4%) Patienten aus der Stimulusgruppe sagten, dass sie auf jeden Fall ihre Daten zum Training zur Verfügung stellen würden. In der Kontrollgruppe waren es 27 (50,9%) (siehe **Abbildung 11**).

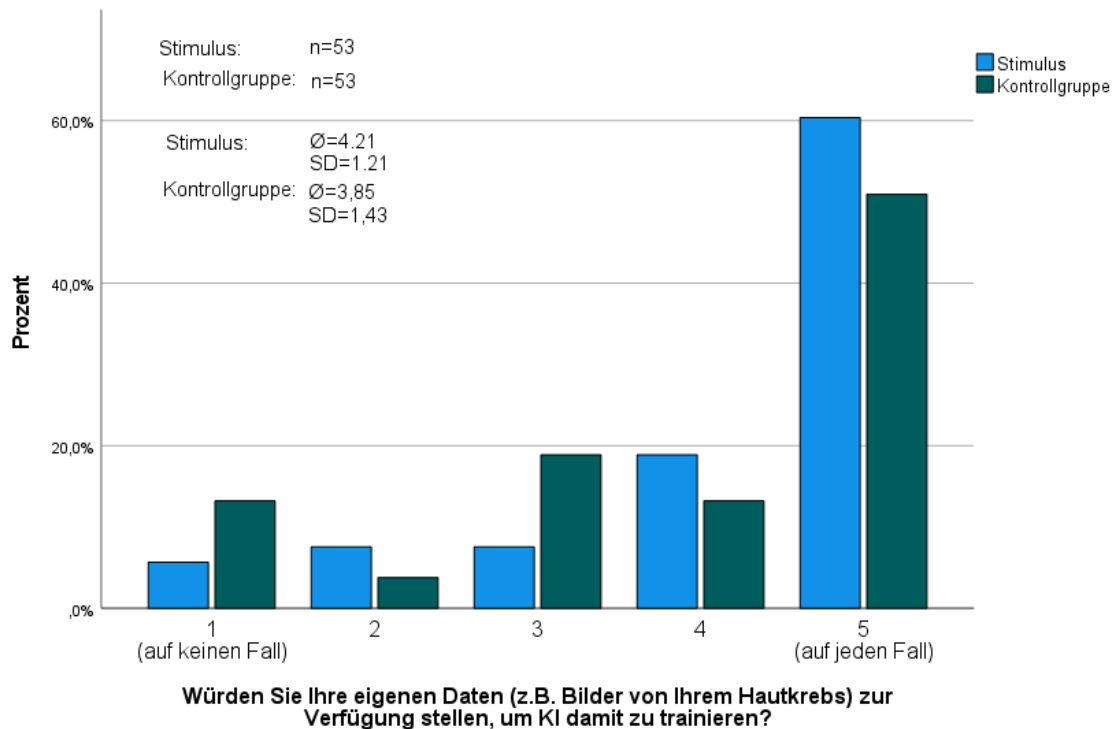


Abbildung 11: Säulendiagramm zu der Bereitschaft der Patienten, Daten zum Training von KI zur Verfügung zu stellen, in den Subgruppen "Stimulus" und "Kontrollgruppe"

Insgesamt hinterließen 3 (2,8%) Patienten eine Anmerkung. Eine Anmerkung war, dass eine KI für eine Diagnose hilfreich sei, jedoch keinen Ersatz darstelle. Ein anderer Patient sagte, dass zusätzlich zur KI ergänzende Untersuchungen durch ein Labor und einen Arzt nötig seien. Schließlich gab ein Patient an, Dachdecker zu sein und da er häufig der Sonne ausgesetzt sei, begrüße er den Einsatz von KI.

4.2 Ergebnisse der Ärztebefragung

Insgesamt haben 63 Ärzte an der Befragung teilgenommen. 3 Personen sind von der statistischen Bewertung ausgeschlossen worden, da diese eine andere Fachrichtung als Hausarzt/Internist oder Dermatologe angegeben haben. Aus diesem Grund sind die Daten von 60 befragten Ärzten in die statistische Analyse eingeflossen.

Unter den befragten Ärzten waren 30 (50,0%) Männer und 30 (50,0%) Frauen. Das Durchschnittsalter der Ärzte betrug 52 Jahre (SD=7,89). Dabei waren die Männer mit 54 Jahren (SD=5,82) im Schnitt 4 Jahre älter als die Frauen mit einem Durchschnittsalter von 50 Jahren (SD=9,27). Die jüngste befragte Person war 26 Jahre alt. Als höchstes Alter sind 67 Jahre angegeben worden (siehe **Abbildung 12**).

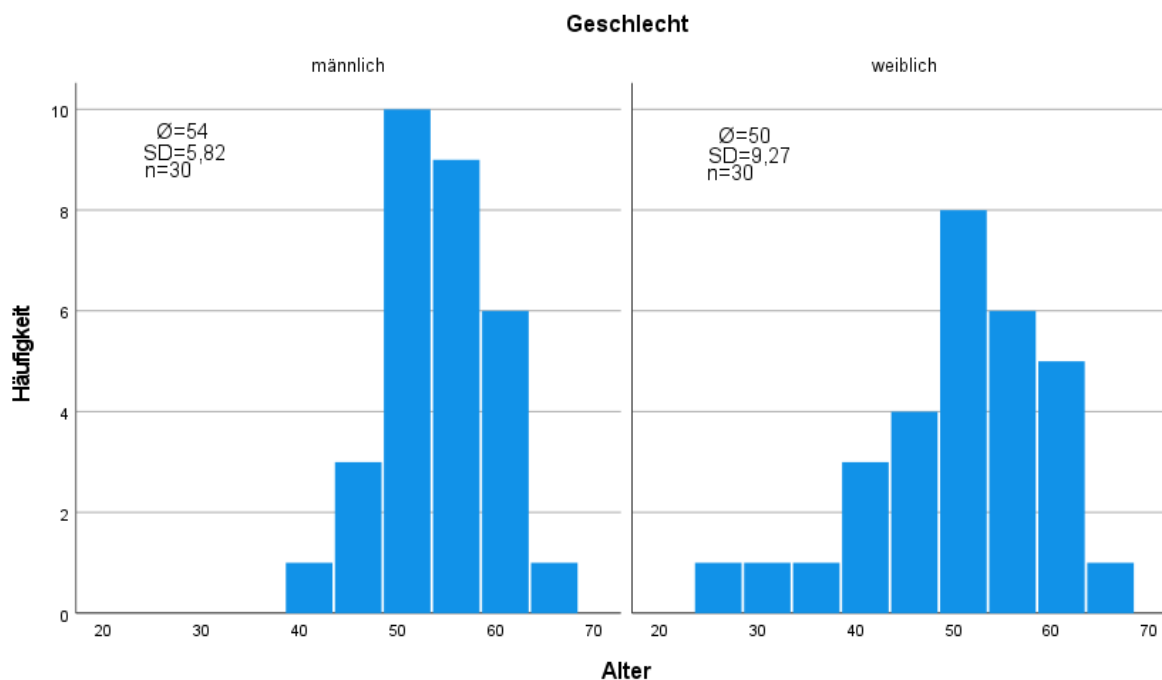


Abbildung 12: Säulendiagramme zu der geschlechtsspezifischen Altersverteilung der Ärzte

Unter den befragten Ärzten waren mit jeweils 30 (50,0%) Ärzten gleichviele Dermatologen wie Hausärzte vertreten. Die Hausärzte gaben ausnahmslos an, ambulant beschäftigt zu sein. Von den Dermatologen waren 22 (73,3%) im ambulanten Versorgungsbereich und 3 (10,0%) stationär tätig. 5 (16,7%) Dermatologen waren sowohl im ambulanten als auch im stationären Bereich beschäftigt (siehe **Tabelle 22**). Im Schnitt gaben die Ärzte eine Berufserfahrung von 22,03 Jahren (SD=7,41) in ihrer zum Zeitpunkt der Befragung ausgeübten Fachrichtung an. Die minimale Erfahrung betrug ein Jahr, die maximale Erfahrung betrug 45 Jahre. Insgesamt haben 58 (96,7%) Ärzte eine abgeschlossene Facharztausbildung in ihrer zu dem Zeitpunkt der Befragung ausgeübten Fachrichtung angegeben (siehe **Tabelle 23**).

Bei der Frage, wie häufig die Ärzte in ihrem Alltag mit Hautkrebsdiagnostik konfrontiert sind, ist im Schnitt ein Punktwert von 4,18 (SD=1,03) angegeben worden. Alle Ärzte gaben an, bis zu einem gewissen Umfang von der Hautkrebsdiagnostik in ihrem Alltag tangiert zu werden. Kein Arzt hat eine 1 angegeben. 10 (16,7%) Ärzte würden häufig mit Hautkrebsdiagnostik konfrontiert werden und 33 (55,0%) Ärzte gaben an, täglich in die Hautkrebsdiagnostik involviert zu sein. 14 (46,7%) Hausärzte und 29 (96,6%) Dermatologen gaben an, häufig bis täglich mit Hautkrebsdiagnostik zu tun zu haben (Werte von 4 und 5) (siehe **Tabelle 2**).

		Fachrichtung		Gesamt	
		Dermatolo ge	Hausarzt (Allgemeinmediziner /Internist)		
Hautkrebsdiagnostik allgemein: Wie häufig werden Sie in Ihrem Alltag mit der Hautkrebsdiagnostik konfrontiert?	1 (nie)	Anzahl	0	0	0
		% von Fachrichtung	0,0%	0,0%	0,0%
	2	Anzahl	0	5	5
		% von Fachrichtung	0,0%	16,7%	8,3%
	3	Anzahl	1	11	12
		% von Fachrichtung	3,3%	36,7%	20,0%
	4	Anzahl	1	9	10
		% von Fachrichtung	3,3%	30,0%	16,7%
	5 (täglich)	Anzahl	28	5	33
		% von Fachrichtung	93,3%	16,7%	55,0%
	Gesamt	Anzahl	30	30	60
		% von	100,0%	100,0%	100,0%
Fachrichtung					

Tabelle 2: Häufigkeit der Hautkrebsdiagnostik im Alltag der Ärzte

59 (98,3%) Ärzte gaben an, schonmal etwas von KI gehört zu haben. Im Schnitt schätzten die Ärzte ihr Wissen auf eine 3,23 (SD=1,14). Dabei sagten 5 (8,3%) Ärzte, dass sie keine Kenntnisse von der Funktionsweise von KI haben, während 10 (16,7%) Ärzte angaben, wenig dazu zu wissen. 18 (30,0%) Ärzte gaben an, eher mehr zu dem Thema zu wissen und 8 (13,3%) Ärzte sagten, dass sie ein sehr gutes Verständnis von der Funktionsweise von KI haben (siehe **Tabelle 24**). Betrachtet man die Analyse der einzelnen Fachrichtungen, lassen sich deutliche Unterschiede erkennen. Die Dermatologen schätzten ihr Wissen im Schnitt auf eine 3,77 (SD=0,94), die Hausärzte hingegen auf eine 2,70 (SD=1,09) ein (siehe **Abbildung 13**).

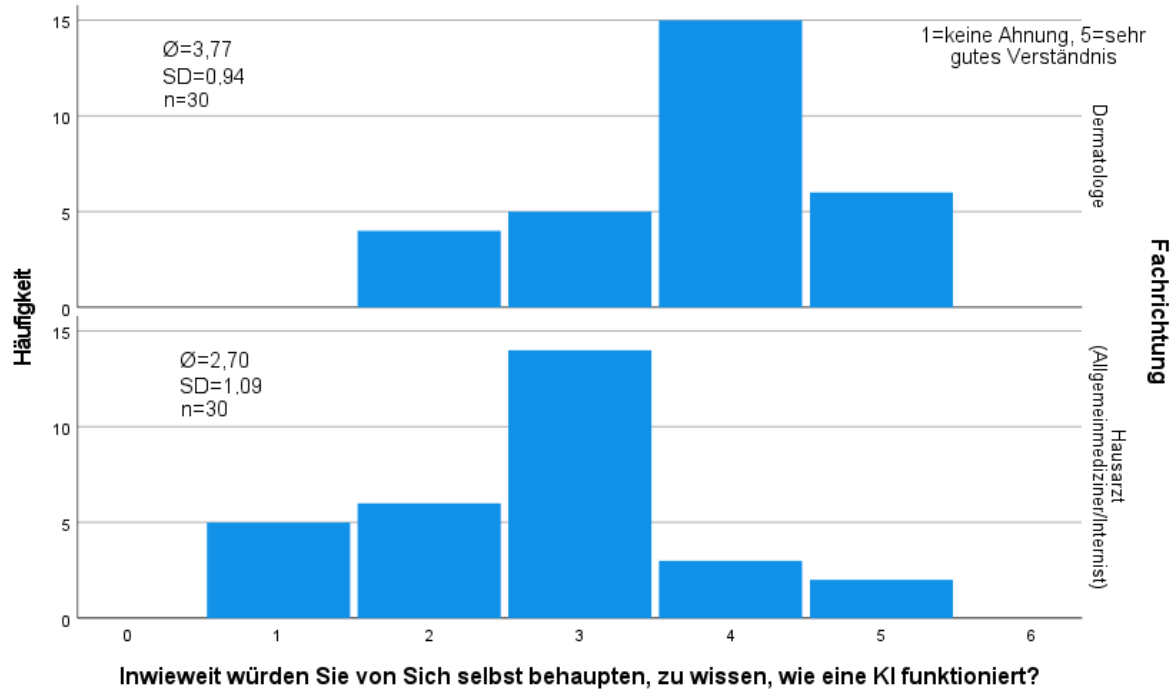


Abbildung 13: Säulendiagramme zu der Einschätzung der Ärzte in Bezug auf ihr Wissen um die Funktionsweise von KI

11 (18,3%) Ärzte gaben an, im Rahmen der Hautkrebsdiagnostik KI einzusetzen. 2 (3,3%) Ärzte diagnostizierten im Rahmen ihrer ärztlichen Tätigkeit keinen Hautkrebs (siehe **Tabelle 25**). Diejenigen Ärzte, welche KI einsetzten, gehörten alle dem Fachbereich der Dermatologie an. Insgesamt gaben die befragten Ärzte im Durchschnitt eine 4,02 (SD=1,09) bei der Frage an, ob sie den generellen Einsatz von KI im Rahmen der Hautkrebsdiagnostik befürworten würden. 19 (76,0%) Dermatologen und 15 (68,2%) Hausärzte gaben an, den Einsatz von KI auf jeden Fall oder eher zu befürworten. Kein Dermatologe sagte, den Einsatz von KI auf keinen Fall zu befürworten (siehe **Abbildung 14**).

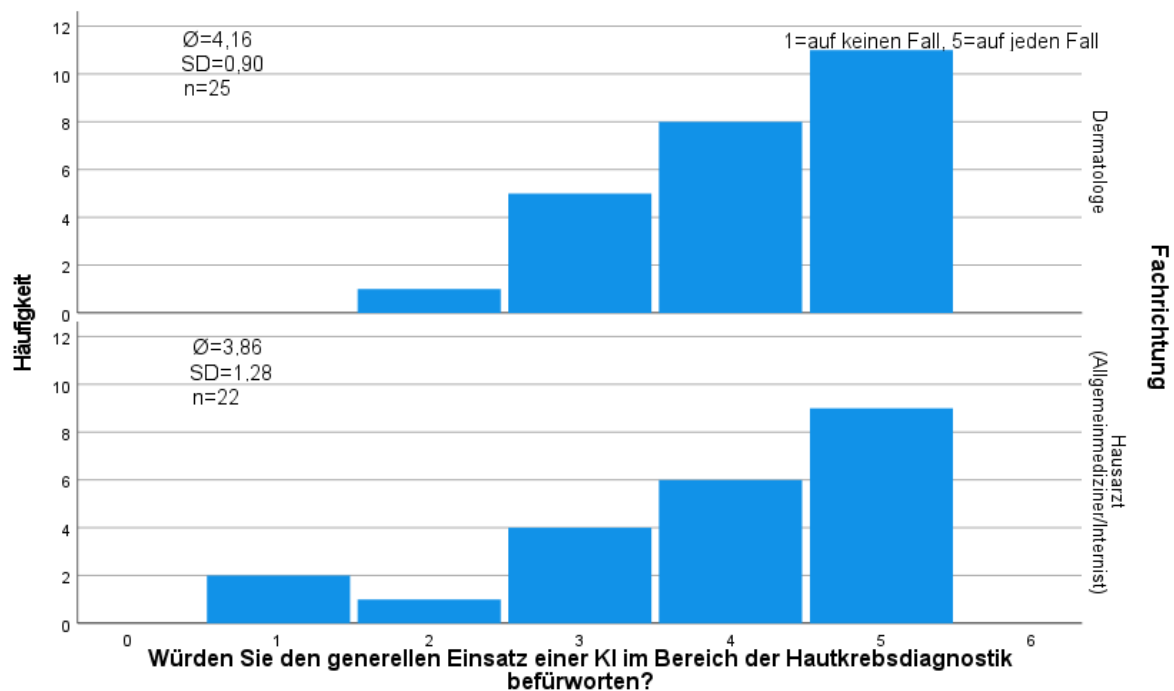


Abbildung 14: Säulendiagramme zu dem Einsatz von KI in der Hautkrebsdiagnostik aus Sicht der Ärzte

26 (43,3%) Ärzte erhielten bei ihrer Befragung einen Aufklärungstext zu KI, die restlichen 34 (56,7%) Ärzte erhielten keine Aufklärung. Die aufgeklärten Ärzte schätzen ihr Wissen in Bezug auf KI im Schnitt mit einer 3,42 (SD=0,81) ein. Die Ärzte ohne Aufklärung gaben im Mittel eine 3,09 (SD=1,33) an.

58 (96,7%) Ärzte gaben an, ein Smartphone oder ein anderes internetfähiges Endgerät zu besitzen. 41 (70,7%) Ärzte sagten in diesem Zusammenhang, dass sie am häufigsten ein Smartphone in ihrer Freizeit einsetzen. Das Tablet und der Computer wurden von jeweils 8 (13,8%) Ärzten am häufigsten genutzt. Insgesamt setzten 54 (94,4%) Ärzte dieses Medium ein, um sich über Krankheitssymptome zu informieren. Die befragten Ärzte gaben im Durchschnitt eine Häufigkeit von 3,10 (SD=1,02) an, das ausgewählte Medium zu nutzen, um sich über Krankheitssymptome zu informieren (siehe **Abbildung 15**).

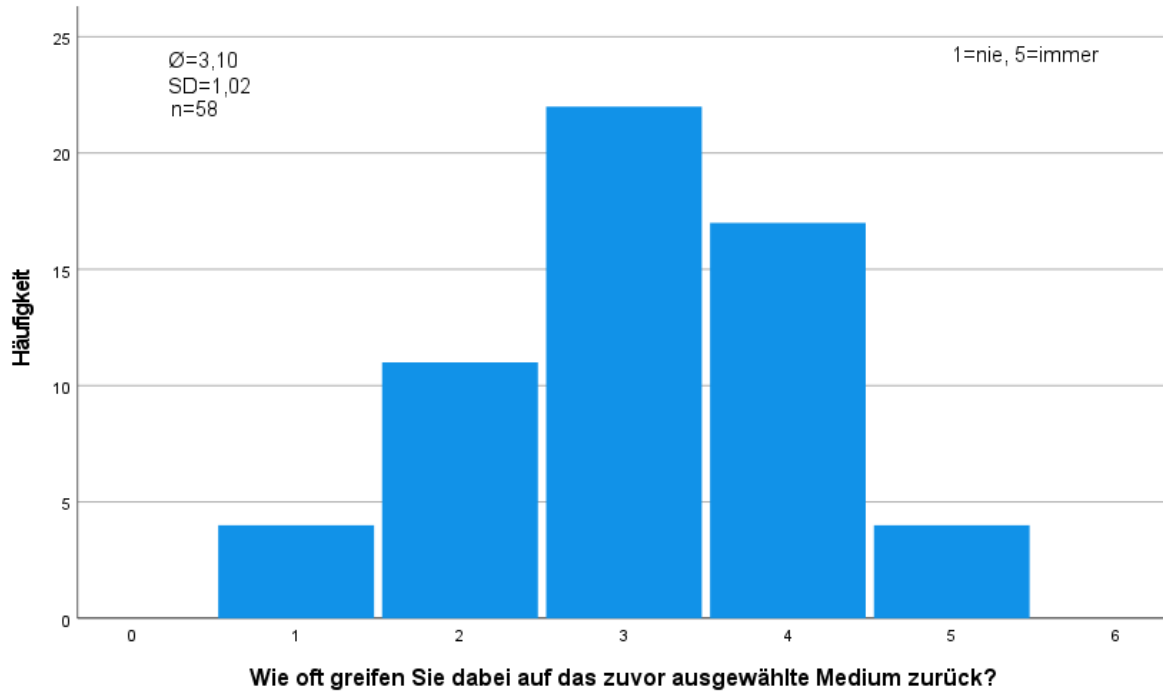


Abbildung 15: Säulendiagramm zu der Mediennutzung der Ärzte bei der Recherche von Krankheitssymptomen

Am häufigsten wurde auf Internetsuchmaschinen wie zum Beispiel Google zurückgegriffen. Diese werden von 51 (94,4%) der befragten Ärzte genutzt. 9 (16,7%) Ärzte nutzten spezielle Diagnoseprogramme für den Computer und 6 (11,1%) Ärzte gaben an, Gesundheitsapps zu nutzen. Die Kategorie „sonstiges“ wurde von 20 (37,0%) Ärzten ausgewählt (siehe **Tabelle 26**).

Die befragten Dermatologen würden Patienten im Mittel mit einer 2,37 (SD=1,27) dazu raten, KI selbst einzusetzen. Bei den Hausärzten lag der Mittelwert bei 2,03 (SD=1,13) (siehe **Abbildung 16**). 23 (38,3%) Ärzte sagten, dass sie Patienten auf keinen Fall dazu raten würden, KI selbstständig einzusetzen, 15 (25,0%) rieten eher davon ab. 9 (15,0%) rieten Patienten eher dazu, KI einzusetzen und 2 (3,3%) Ärzte sagten, dass sie Patienten den Gebrauch von KI auf jeden Fall empfehlen würden (siehe **Tabelle 27**).

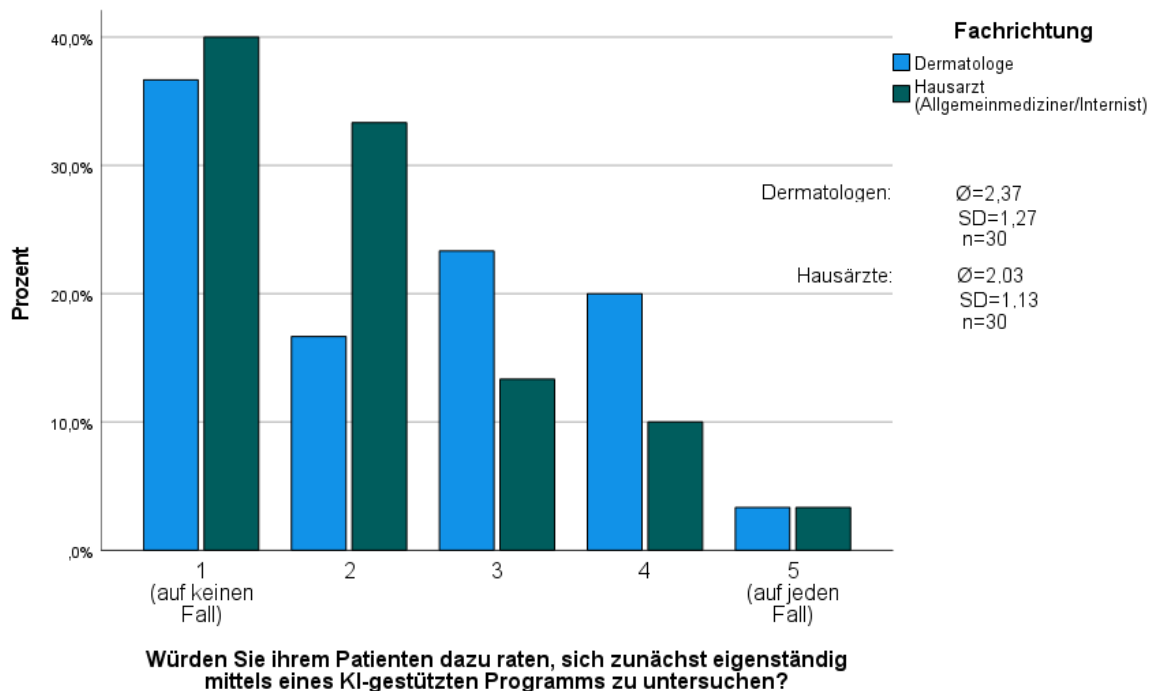


Abbildung 16: Säulendiagramm zu der Zustimmung der Ärzte zum selbstständigen Einsatz von KI durch Patienten

Für den Fall, dass die KI im häuslichen Einsatz durch den Patienten zu einer bösartigen Diagnose gelangt, würden die Ärzte mit einem Mittelwert von 4,81 (SD=0,62) den Patienten dazu raten, sie trotzdem aufzusuchen. Im Falle einer gutartigen Diagnose wurde im Schnitt eine 3,97 (SD=0,99) angegeben. 33 (89,2%) Ärzte gaben an, dass sie im Falle einer bösartigen Diagnose den Patienten auf jeden Fall dazu raten würden, sie aufzusuchen. Für den Fall, dass die Diagnose gutartig wäre, sagten 15 (40,5%) Ärzte, dass sie ihren Patienten auf jeden Fall dazu raten würden, sie aufzusuchen und 8 (21,6%) beziehungsweise 12 (32,4%) Ärzte sagten, dass sie dies eher ihren Patienten empfehlen würden, beziehungsweise unentschlossen seien (siehe **Tabelle 28**). Betrachtet man die Subgruppen, welche durch die Aufklärung entstanden sind, so wird vor allem in Bezug auf die gutartige Diagnose durch die KI ein Unterschied deutlich. Mit einem Mittelwert von 4,36 (SD=1,01) in der Stimulusgruppe, zeigt sich, dass die aufgeklärten Ärzte eine stärkere Tendenz dazu hatten, die Patienten trotz der gutartigen Diagnose durch die KI zusätzlich persönlich zu untersuchen, als die Ärzte aus der Kontrollgruppe (Mittelwert=3,74, SD=0,92). Diese Fragen schlossen nur diejenigen Ärzte ein, welche den Gebrauch von KI durch Patienten nicht kategorisch ablehnten.

25 (41,7%) Ärzte sagten in Bezug auf die Möglichkeit, KI als Assistenzsystem einzusetzen, dass sie diese auf jeden Fall nutzen würden, 15 (25,0%) Ärzte tendierten ebenfalls dazu, KI unterstützend einzusetzen (siehe **Tabelle 29**). Im Mittel gaben die Ärzte in Bezug auf den Einsatz von KI als Assistenzsystem einen Wert von 4,0 (SD=1,04) an. Die Verteilungen in der Stimulus- und der Kontrollgruppe unterschieden sich nicht voneinander, Kolmogorov-Smirnov $p > 0,05$. Der Mann-Whitney-U-Test ergab keinen statistisch signifikanten Unterschied zwischen der Bereitschaft zum Einsatz von KI in der Stimulus- und der Kontrollgruppe ($U = 424,00$, $Z = -0,285$, $p > 0,05$). Bei der Geschlechterverteilung zeigte sich folgendes Bild: 11 (36,7%) Ärzte sagten, dass sie KI auf jeden Fall unterstützend einsetzen würden und jeweils 8 (26,7%) Ärzte gaben an, dass sie zu dem Einsatz tendieren würden oder unentschlossen seien. Von den Ärztinnen sagten 14 (46,7%), dass sie KI auf jeden Fall unterstützend einsetzen würden, 7 (23,3%) tendierten eher zu dem Einsatz und 9 (30,0%) enthielten sich. Keine Ärztin lehnte den unterstützenden Einsatz von KI ab. Der Mittelwert lag bei den Ärzten bei 3,83 (SD=1,18) und bei den Ärztinnen bei 4,17 (SD=0,87) (siehe **Abbildung 17**).

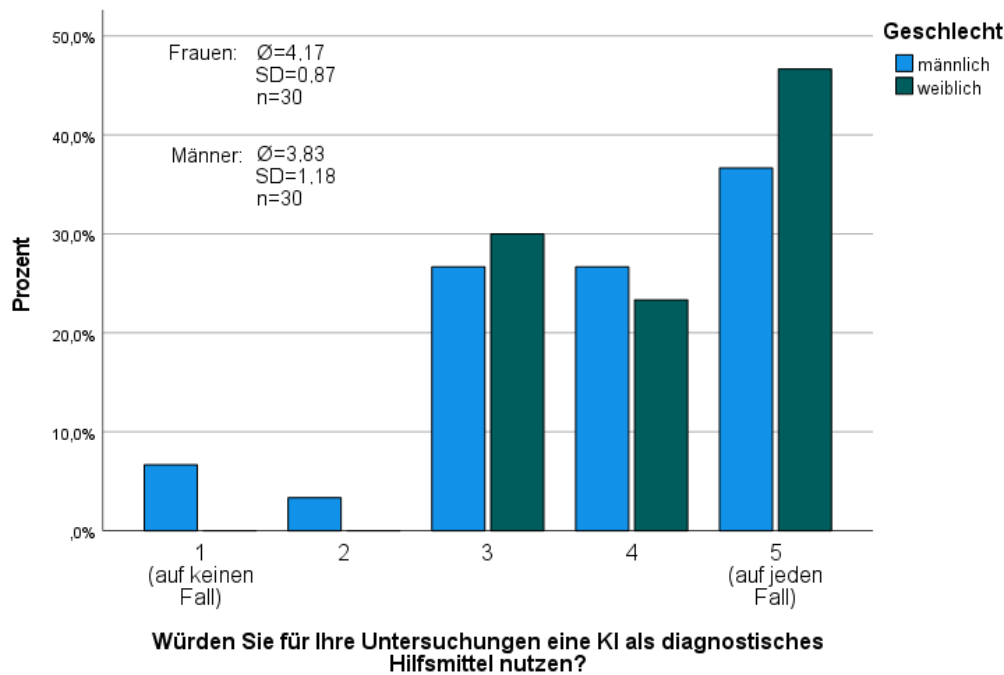


Abbildung 17: Säulendiagramm zu der geschlechterspezifischen Bereitschaft der Ärzte KI einzusetzen

Der Mittelwert lag für die Dermatologen bei 4,17 (SD=0,87) und für die Hausärzte bei 3,83 (SD=1,18). Für die Ärzte jünger gleich 50 Jahre ergab sich im Mittel ein Wert von 4,00 (SD=1,11). Der Mittelwert der Ärzte, welche älter als 50 Jahre waren, betrug 4,00 (SD=1,01). In Bezug auf das Vorwissen der Ärzte hat sich für die Bereitschaft für den Einsatz von KI bei den Ärzten folgendes Bild gezeigt: Ärzte mit niedrigem Vorwissen (Werte von 1 bis 3) gaben in der Stimulusgruppe im Mittel eine 3,54 (SD=0,97) und in der Kontrollgruppe eine 3,71 (SD=1,19) an. Bei den Ärzten mit hohem Vorwissen (Werte von 4 und 5) ergab sich in der Stimulusgruppe ein Mittelwert von 4,62 (SD=0,51) und in der Kontrollgruppe ein Mittelwert von 4,31 (SD=0,95) (siehe **Abbildung 18**). Insgesamt sagten unter den Dermatologen mit hohem Vorwissen 80,9%, dass sie KI als Assistenzsystem einsetzen würden. Dieser Anteil betrug unter den Hausärzten 100% (siehe **Tabelle 30**).

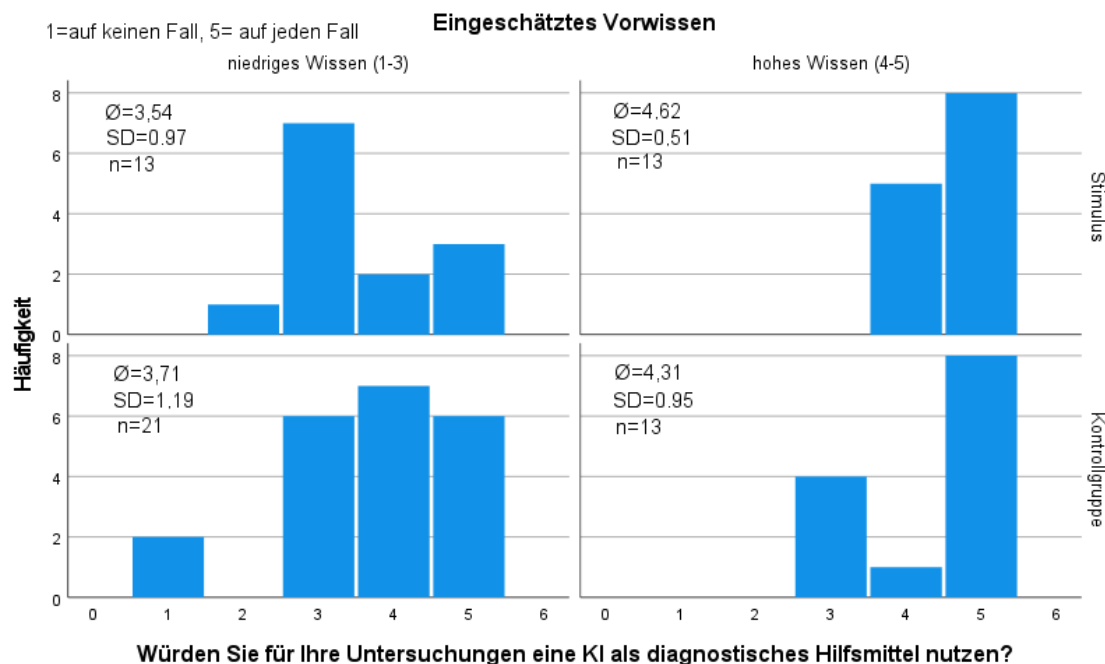


Abbildung 18: Säulendiagramme zu der Bereitschaft der Ärzte KI einzusetzen, in Abhängigkeit des Vorwissens, in den Subgruppen "Stimulus" und "Kontrollgruppe"

Die folgenden 3 Fragen wurden nur von denjenigen Ärzten beantwortet, welche den unterstützenden Einsatz von KI nicht ablehnten.

Für den Fall, dass der Arzt selbst eine therapeutisch relevante Diagnose fällen würde und die KI diese verneinte, haben 38 (65,5%) Ärzte angegeben, dass sie eine Biopsie nehmen würden. 9 (15,5%) Ärzte sagten, dass sie dennoch therapieren würden und 11 (19,0%) Ärzte wählten „eine andere Option wählen“ aus. Für den umgekehrten Fall, dass die KI eine therapeutische Diagnose fällen würde und der Arzt diese verneinte, haben 40 (69,0%) Ärzte angegeben, dass sie eine Biopsie nehmen würden. 2 (3,4%) Ärzte sagten, dass sie dennoch therapieren würden und 16 (27,6%) Ärzte wählten eine andere Option (siehe **Tabelle 31**).

Die befragten Ärzte würden mit einem Mittelwert von 1,31 (SD=0,60) allein auf die Diagnose einer KI vertrauen, ohne den Patienten selbst untersucht zu haben. 43 (74,1%) Ärzte sagten, dass sie auf keinen Fall allein auf eine KI vertrauen würden, 13 (22,4%) Ärzte tendierten ebenso dazu. Ein Arzt (1,7%) enthielt sich und ein weiterer (1,7%) tendierte dazu, allein auf eine KI zu vertrauen (siehe **Abbildung 19**).

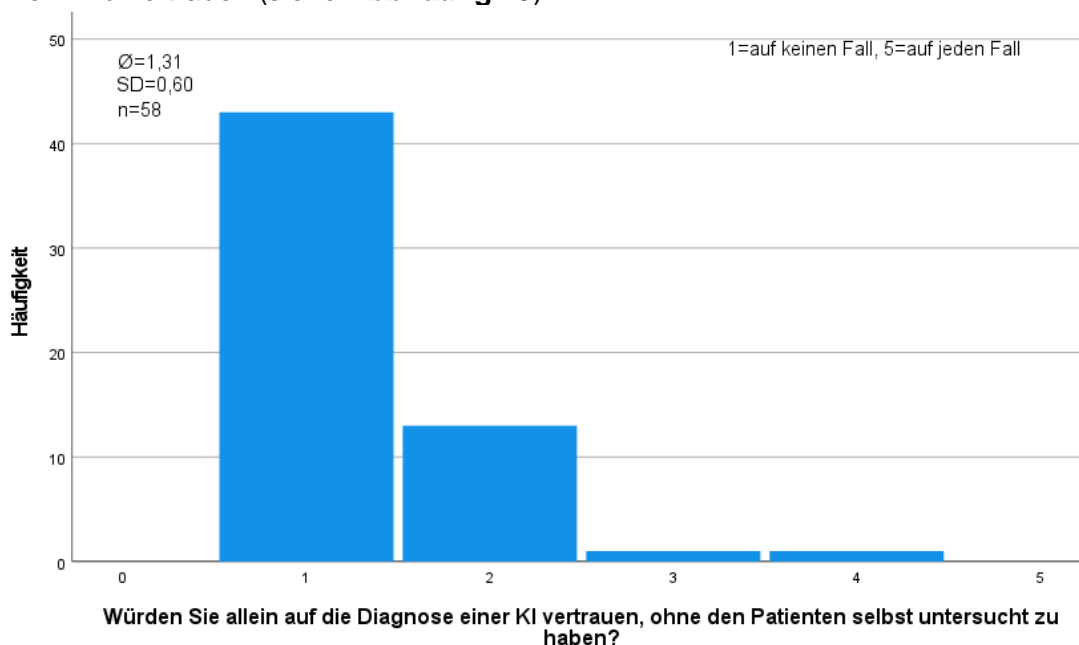


Abbildung 19: Säulendiagramm zu dem Vertrauen der Ärzte in den stand-alone Einsatz von KI

14 (23,3%) Ärzte sagten bei der Befragung, dass sie auf jeden Fall dem Patientenwunsch nachkommen würden, KI einzusetzen und 19 (31,7%) Ärzte tendierten ebenfalls dazu (siehe **Tabelle 32**). Insgesamt ergab sich für die Zustimmung ein Mittelwert von 3,6 (SD=1,12).

Ob das Verständnis um die Funktionsweise einer KI als Voraussetzung für deren Einsatz betrachtet werden sollte, wurde von den Ärzten im Mittel mit einer 3,64 (SD=1,33) beantwortet. Es zeigte sich kein Unterschied zwischen der Stimulus- und der Kontrollgruppe (Mittelwert=3,67 (SD=1,31) vs. Mittelwert=3,62 (SD=1,37)). Die Verteilungen in der Stimulus- und der Kontrollgruppe unterschieden sich nicht voneinander, Kolmogorov-Smirnov $p > 0,05$. Der Mann-Whitney-U-Test ergab keinen statistisch signifikanten Unterschied, ob das Verständnis um die Funktionsweise einer KI als Voraussetzung für deren Einsatz betrachtet werden sollte, zwischen der Stimulus- und der Kontrollgruppe ($U = 344,50$, $Z = -0,065$, $p > 0,05$). 21 (39,6%) Ärzte stimmten zu, dass das Verständnis für die Funktionsweise von KI als Voraussetzung für deren Einsatz betrachtet werden sollte und 9 (17,0%) Ärzte tendierten ebenfalls zu dieser Ansicht (siehe **Tabelle 33**). In den Subgruppen der Fachrichtungen zeigten sich keine relevanten Unterschiede. Dahingegen lag der Mittelwert bei den Ärzten bei 3,19 (SD=1,47) und bei den Ärztinnen bei 4,12 (SD=0,99). In Bezug auf das Vorwissen zeigte sich, dass die Zustimmung bei höherem Wissen größer war. Diese lag bei höherem Vorwissen bei 66,6% (siehe **Tabelle 3**).

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 (auf keinen Fall)	1	3,8	4,2	4,2
	2	4	15,4	16,7	20,8
	3	3	11,5	12,5	33,3
	4	5	19,2	20,8	54,2
	5 (auf jeden Fall)	11	42,3	45,8	100,0
	Gesamt	24	92,3	100,0	
Fehlend	nicht beantwortet	2	7,7		
Gesamt		26	100,0		

Tabelle 3: Ansicht der Ärzte mit hohem Vorwissen, ob das Wissen um die Funktionsweise als Voraussetzung für deren Einsatz zu betrachten ist

14 (24,1%) Ärzte sagten, dass KI auf jeden Fall Teil der Ausbildung zu ihrer Facharzttrichtung sein sollte. 16 (27,6%) Ärzte stimmten eher zu (siehe **Tabelle 34**). Dabei zeigte sich kein relevanter Unterschied zwischen den Dermatologen und den Hausärzten. Insgesamt ergab sich ein Mittelwert von 3,69 (SD=0,98).

50 (83,3%) Ärzte waren der Meinung, dass KI das Potential hat, die Hautkrebsversorgung zu verbessern. 9 (15,0%) Ärzte glaubten, dass die Versorgung von der KI unbeeinflusst bleibt und ein Arzt (1,7%) meinte, dass KI die Versorgung von Hautkrebs verschlechtert (siehe **Tabelle 35**).

Bei dieser Frage lag eine Likertskala mit Werten von 1 bis 3 zugrunde. Der Mittelwert lag in der Stimulusgruppe bei 2,85 (SD=0,37), in der Kontrollgruppe bei 2,79 (SD=0,48) und für beide Gruppen zusammen bei 2,82 (SD=0,43). Es zeigten sich in diesem Zusammenhang keine relevanten Unterschiede in den Subgruppen „Alter“, „Geschlecht“ und „Facharzttrichtung“. Die Verteilungen in der Stimulus- und der Kontrollgruppe unterschieden sich nicht voneinander, Kolmogorov-Smirnov $p > 0,05$. Der Mann-Whitney-U-Test ergab keinen statistisch signifikanten Unterschied zwischen dem geschätzten Potential in der Stimulus- und der Kontrollgruppe ($U = 430,00$, $Z = -0,277$, $p > 0,05$).

Der überwiegende Anteil der Ärzte würde die im Rahmen der Diagnostik anfallenden Daten zum Training von KI zur Verfügung stellen (Mittelwert=4,07, SD=1,21). 30 (50%) Ärzte würden dies auf jeden Fall machen und 15 (25,0%) weitere Ärzte tendierten dazu (siehe **Tabelle 36**). In der Stimulusgruppe fiel die Zustimmung höher aus als in der Kontrollgruppe. 16 (61,5%) Ärzte aus der Stimulusgruppe würden die Daten auf jeden Fall zur Verfügung stellen und 7 (26,9%) Ärzte tendierten ebenfalls dazu. Dahingegen waren es in der Kontrollgruppe nur 14 (41,2%) Ärzte, welche die Daten auf jeden Fall zur Verfügung stellen würden und 8 (23,5%) Ärzte, die dies eher tun würden (siehe **Abbildung 20**).

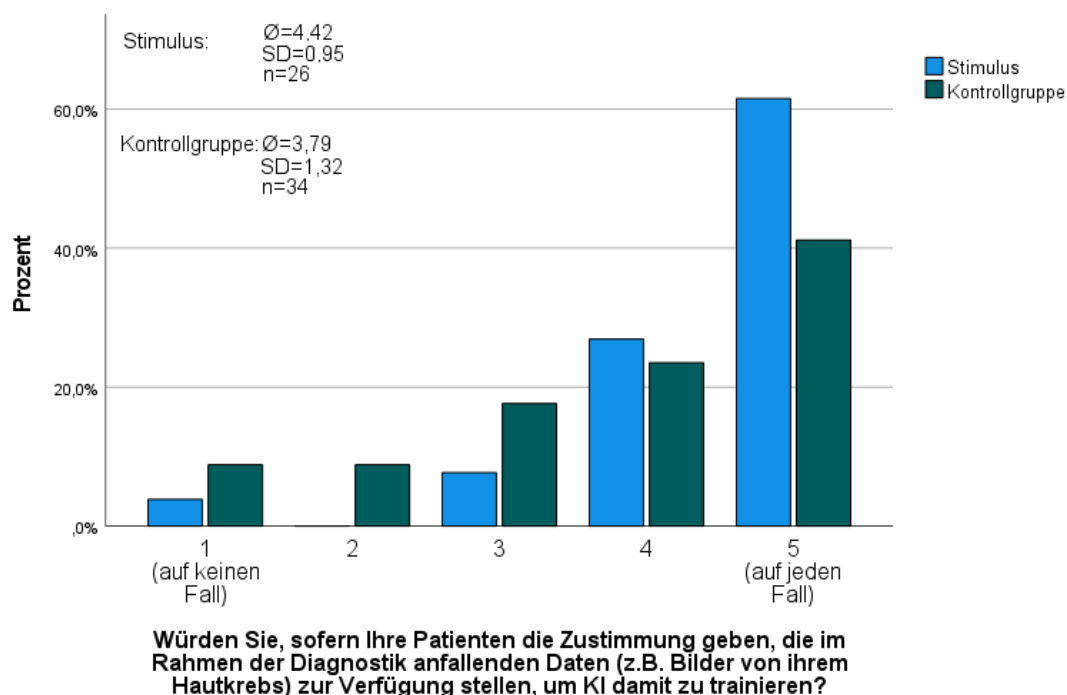


Abbildung 20: Säulendiagramm zu der Bereitstellung von Daten zum Training von KI in Abhängigkeit von der Aufklärung

49 (81,7%) Ärzte gaben an, dass die Diagnosesicherheit die wichtigste Eigenschaft einer KI ist. 7 (11,7%) Ärzte sagten, dass ihnen das Erkennen möglichst vieler Diagnosen bei der Anwendung einer KI am wichtigsten ist und 4 (6,7%) Ärzte wählten die Freitextantwort (siehe **Tabelle 37**). Hier wurden zum einen die Zeitersparnis beim Hautkrebscreening und der Beurteilung genannt, der positive Effekt durch das Vorscreenen von vielen Patienten und schließlich der Preis und die Wahlfreiheit.

Der Aspekt „Nachvollziehbarkeit des Ergebnisses“ hatte für 27 (46,6%) Ärzte in Bezug auf die Anwendung von KI die höchste Priorität. An zweiter Stelle folgt die „leichte Bedienbarkeit“, welche von 23 (39,7%) Ärzten auf Rang 2 gewählt worden ist. Die „Analysegeschwindigkeit“ und der damit vermutet Zeitgewinn wurde von 25 (43,1%) Ärzten auf Rang 3 gewählt und 42 (42,4%) Ärzte platzierten die „Delegierbarkeit“ auf Rang 4 (siehe **Tabelle 38**).

In den Freitextantworten sind zahlreiche Stärken der KI benannt worden. Eine relativ häufig angeführte Eigenschaft war die Objektivität der KI (15%). Ebenso wurde die Verfügbarkeit der KI von mehreren Ärzten benannt (8%). Diese bezog sich sowohl auf die zeitliche Verfügbarkeit, also die Möglichkeit, KI Tag und Nacht einzusetzen, als auch auf die räumliche Verfügbarkeit, welche die Hautkrebsdiagnostik in Regionen mit geringer Arztdichte ermöglicht. Weitere Stärken, welche von den Ärzten aufgeführt wurden, waren zum einen die Möglichkeit der Zweitmeinung (10%), welche vor allem der ambulanten Versorgung zugutekäme, und zum anderen die diagnostische Sicherheit (5%), welche durch KI gefördert würde. Auch wurde ein Vorteil für das Hautkrebscreening gesehen, bei welchem KI ein gutes Screeningtool darstelle (7%). Allgemein wurde die KI als gute Hilfestellung für verschiedene Fachärzte aufgeführt, auch in Bezug auf unklare Diagnosen, seltene Diagnosen und Differentialdiagnosen, die sie vor allem auch für weniger erfahrene Ärzte biete (8%). Als weitere Stärke wurde der Datenspeicher aufgeführt, welcher einer KI zur Verfügung steht (10%). Dieser sei größer als das Wissen und die Erfahrung der Ärzte. Auch sei die KI nicht beeinflussbar, nicht ablenkbar, nicht emotional, vorurteilsfrei und könne den menschlichen Bias vermeiden (7%). Weitere Stärken der KI seien deren Schnelligkeit (5%) und Präzision (2%). Ebenso biete die KI Raum für interdisziplinäre Kommunikation (2%), nicht zuletzt auch aufgrund ihrer Standardisierbarkeit (2%), welche ebenfalls als Stärke angeführt wurde. Eine andere Stärke bestehe in der Möglichkeit, Kosten einzusparen, da weniger Fachärzte benötigt würden (2%).

Auch wurde eine Stärke der KI in der Umsetzung neuer Erkenntnisse (2%) und dem ständigen Zuwachs an Wissen gesehen (2%).

Eine häufig aufgezählte Schwäche war die fehlende menschliche Interaktionsfähigkeit des Algorithmus, um auf den Patienten zu reagieren (10%). Eine weitere Schwäche sei die fehlende Verlässlichkeit, sei es aufgrund mangelnder trainierter Differentialdiagnosen, der zum Trainingsdatenset veränderten Kontextsituationen, des Einflusses von Artefakten, der Erkennung nur eindeutiger Befunde, ihrer Manipulierbarkeit, der Notwendigkeit hoher Datenqualität des Inputs oder der fehlenden Kontext- und Untersuchungsinformationen (22%). Die Folge daraus seien falsche Ergebnisse und eine hohe Fehleranfälligkeit (7%). Als weitere Schwäche wurde die Möglichkeit der Fehlerfortpflanzung aufgrund der selbstlernenden Programmierung benannt (2%). Auch wurden Nachteile in dem technischen Aufwand, der Notwendigkeit von Updates, der komplizierten Anwendung, den Kosten und der geringen Verfügbarkeit gesehen (8%). Zudem brauche KI große Datenmengen (2%), müsse zu den Prozessen in der Praxis passen (2%) und Akzeptanz unter den Patienten finden (2%). Ein Nachteil sei auch die Angst vor KI (2%). Auch läge die Verantwortung nach wie vor beim Arzt, welcher trotz des Einsatzes von KI geschult sein müsse (4%). Auch solle KI nicht ohne ärztliche Begleitung eingesetzt werden, da die Gefahr bestehe, dass diese als eine Art Diagnoseautomat missbraucht würde (2%). Eine andere Schwäche sei das Gefühl der falschen Sicherheit, sodass Diagnosen unkritisch angenommen würden (2%). Auch würde durch den Einsatz von KI die Diskussionsfähigkeit (2%) herabgesetzt werden und es käme zu einer Einsparung von Fachärzten (2%). Nicht zuletzt bestehe auch eine Schwäche darin, dass die KI nicht zum Therapieren in der Lage sei (2%). Schließlich ist auch ein Preisverfall für ärztliche Leistungen angemerkt worden (2%).

Außerdem sind auch Anmerkungen gemacht worden, welche zum Teil bereits bei den Stärken oder Schwächen Erwähnung gefunden haben, sodass an dieser Stelle nicht alle aufgelistet sind. Zwei Ärzte merkten an, dass die Anwendung einer KI finanziell zumeist nicht vergütet wird. Ein weiterer Arzt stellte fest, dass sich die KI erst dann verbreiten werde, wenn die Ergebnisse verlässlich sind und sich kein finanzieller Verlust aus deren Einsatz ergibt. Zudem wurde angemerkt, dass KI bereits eingesetzt wird. Außerdem wurde gesagt, dass ein Ganzkörperscan durch die KI im Rahmen des Hautkrebsscreening sinnvoll sei und wieder ein anderer stellte fest, dass er gern eine KI nutzen würde. Des Weiteren ist aufgeführt worden, dass für die Etablierung einer KI eine interdisziplinäre Schnittstelle zwischen dermatologischen Einrichtungen und Hausarztpraxen nötig sei, sowie die Tatsache, dass die KI aufgrund der mit sich bringenden Zeitersparnis von der Politik zur Pflicht gemacht werden würde. Ein Arzt gab auch an, bereits mehrere Fehldiagnosen mit einer KI gehabt zu haben. Außerdem fand Erwähnung, dass viele digitale Anwendungen den Praxisablauf hemmen würden. Zudem müsse zuerst geklärt werden, wer die Verantwortung für Fehldiagnosen der KI übernehme. Schließlich stellte ein Arzt fest, dass mehr Aufklärungsarbeit für KI sinnvoll sei.

4.3 Kombinierte Ergebnisse der Patienten- und Ärztebefragung

59 (98,3%) Ärzte und 88 (82,2%) Patienten gaben in der Befragung an, schonmal etwas von KI gehört zu haben (siehe **Tabellen 5, 24**).

Die Ärzte schätzten ihr Wissen um die Funktionsweise einer KI im Mittelwert mit einer 3,23 (SD=1,14) ein, während die Patienten im Mittel eine 2,39 (SD=1,20) angaben. Ein Blick auf die statistische Verteilung zeigt, dass 26 (43,3%) Ärzte ein sehr gutes oder gutes Verständnis um die Funktionsweise von KI angegeben haben. Bei den Patienten waren es hingegen nur 21 (20,2%). 30 (28,8%) Patienten gaben an, keine Kenntnisse zu der Funktionsweise von KI zu haben (siehe **Abbildung 21**). Die Verteilungen in der Arzt- und der Patientengruppe unterschieden sich voneinander, Kolmogorov-Smirnov $p < 0,05$. Der Mann-Whitney-U-Test ergab einen statistisch signifikanten Unterschied zwischen dem Verständnis zu der Funktionsweise von KI in der Arzt- und der Patientengruppe ($U = 1921,00$, $Z = -4,201$, $p < 0,001$).

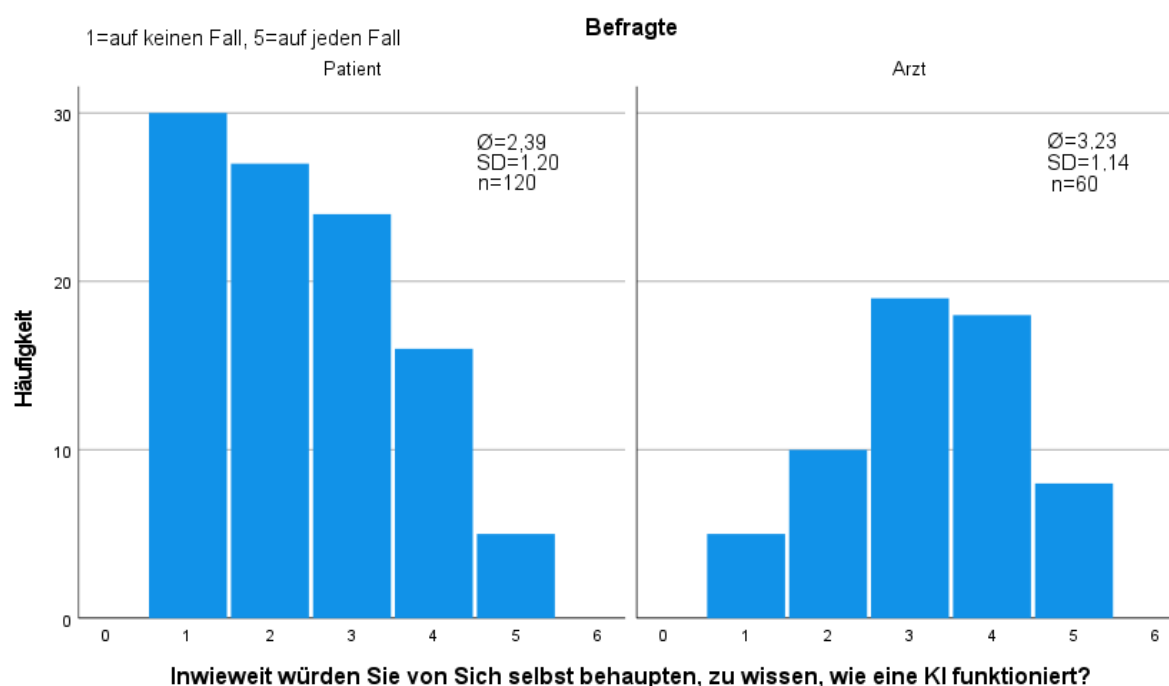


Abbildung 21: Säulendiagramme zu der Einschätzung des Wissens von Ärzten und Patienten um die Funktionsweise von KI

11 (18,3%) Ärzte haben angegeben, dass sie ihren Patienten auf jeden Fall oder eher dazu raten würden, sich selbstständig mittels KI zu untersuchen. 23 (38,3%) würden auf keinen Fall und 15 (25,0%) eher nicht zu dem Einsatz von KI raten. Dahingegen waren es unter den Patienten 35 (45,5%), die sich auf jeden Fall oder eher selbstständig mittels KI untersuchen würden (siehe **Tabelle 12, 27**). Im Mittel gaben die Ärzte eine 2,20 (SD=1,21) und die Patienten eine 3,27 (SD=1,60) an.

Für den Fall, dass eine KI bei einem Patienten eine Hautläsion als bösartig einstuft, hat die Mehrheit der Patienten und Ärzte gesagt, dass ein Arzt aufgesucht werden sollte (Patienten: Mittelwert=4,93, SD=0,41 vs. Ärzte: Mittelwert=4,81, SD=0,62). Für den Fall, dass die KI eine Hautläsion als gutartig klassifiziert, zeigte sich hingegen, dass die Patienten signifikant häufiger ihren Arzt aufsuchen würden, als die befragten Ärzte es empfehlen würden (Patienten: Mittelwert=4,43, SD=1,04 vs. Ärzte: Mittelwert=3,97, SD=0,99), (siehe **Abbildung 22**). Hier unterschieden sich die Verteilungen in der Arzt- und der Patientengruppe voneinander, Kolmogorov-Smirnov $p < 0,05$. Der Mann-Whitney-U-Test ergab einen statistisch signifikanten Unterschied zwischen dem Bestreben der Patienten, im Falle einer gutartigen Diagnose durch eine KI zum Arzt zu gehen und der Empfehlung der Ärzte, sie in diesem Fall aufzusuchen ($U = 809,00$, $Z = -2,665$, $p < 0,009$). 25 Personen ließen diese Frage unbeantwortet.

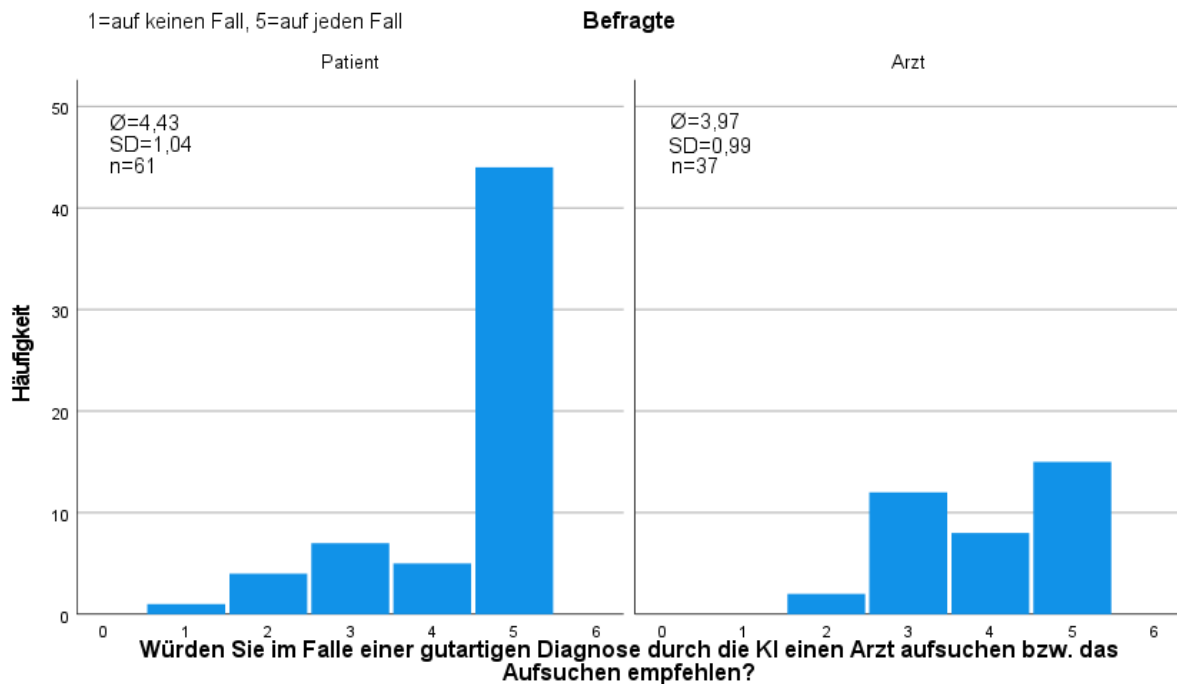


Abbildung 22: Säulendiagramme zu der Ansicht von Ärzten und Patienten, im Falle einer gutartigen Diagnose durch die KI den Arzt aufzusuchen

Im Mittel gaben die Patienten eine 1,94 (SD=1,12) an, ob sie allein auf die Diagnose einer KI vertrauen würden. Bei den Ärzten ergab sich ein Mittelwert von 1,31 (SD=0,60). Es zeigt sich, dass das Vertrauen der Patienten signifikant höher war als das Vertrauen der Ärzte. Die Verteilungen in der Arzt- und der Patientengruppe unterschieden sich voneinander, Kolmogorov-Smirnov $p < 0,05$. Der Mann-Whitney-U-Test ergab einen statistisch signifikanten Unterschied zwischen dem Vertrauen der Patienten und dem Vertrauen der Ärzte in die KI ($U = 2062,50$, $Z = -3,665$, $p < 0,001$). Insgesamt fiel das Vertrauen beider Gruppen niedrig aus. 70 (68,0%) Patienten gaben an, kein oder kaum Vertrauen in die alleinige Diagnose einer KI zu haben und bei den Ärzten waren es 56 (96,6%). 8 (7,8%) Patienten sagten, dass sie auf jeden Fall oder eher auf die Diagnose der KI vertrauen würden (siehe **Abbildung 23**).

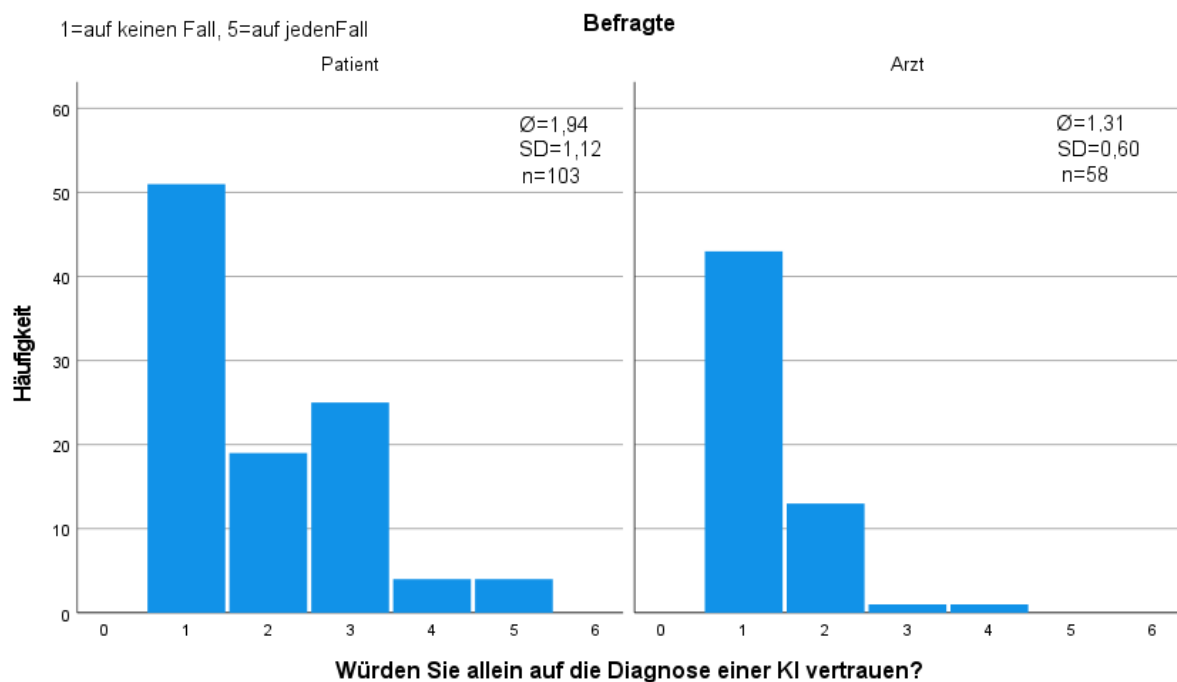


Abbildung 23: Säulendiagramme zu dem Vertrauen von Ärzten und Patienten in die Diagnose einer KI

30 (29,2%) Patienten gaben an, auf jeden Fall oder eher auf den zusätzlichen Einsatz von KI durch den Arzt bestehen zu würden. Im Vergleich dazu haben 33 (55,0%) Ärzte gesagt, dass sie dem Wunsch des Patienten für den Einsatz von KI auf jeden Fall oder eher nachkommen würden (siehe **Tabelle 15, 32**). Dabei ergab sich für die Patientengruppe ein Mittelwert von 2,99 (SD=1,26) und für die Ärztegruppe ein Mittelwert von 3,60 (SD=1,12).

Sowohl bei den Ärzten als auch bei den Patienten war die Bereitschaft, Daten zum Training von KI bereitzustellen, hoch. 76 (71,7%) Patienten gaben an, dass sie auf jeden Fall oder eher dazu tendierten, ihre Daten zur Verfügung zu stellen. Unter den Ärzten waren es 45 (75,0%), die die Daten zur Verfügung stellen würden (siehe **Abbildung 24**). Insgesamt ergab sich für die Gruppe der Patienten ein Mittelwert von 4,02 (SD=1,34) und für die Ärzte ein Mittelwert von 4,07 (SD=1,21).

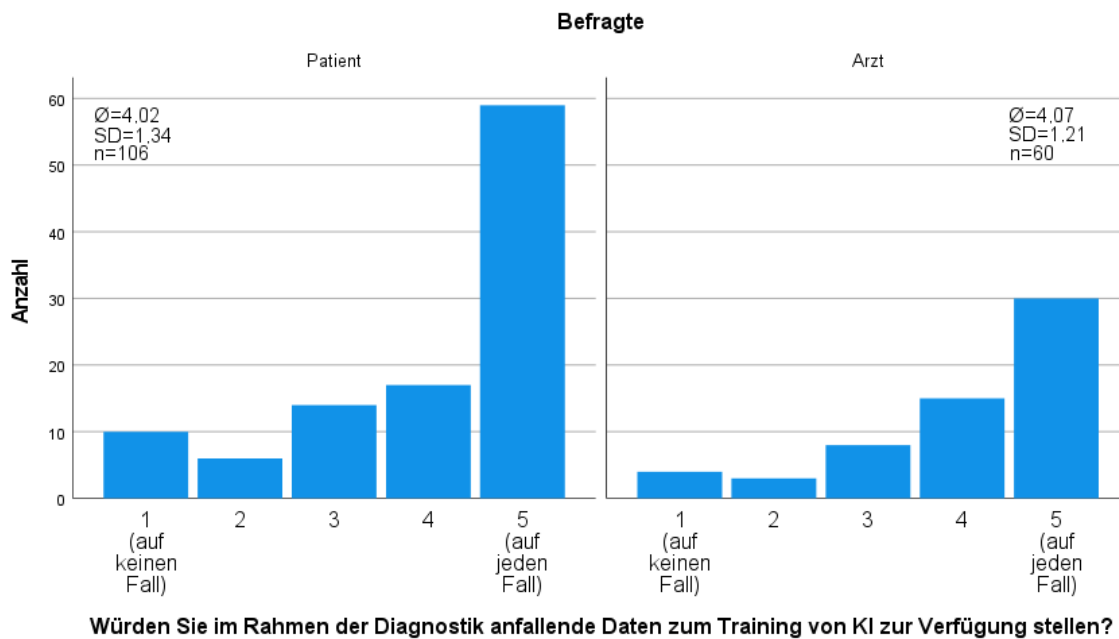


Abbildung 24: Säulendiagramme zu der Bereitschaft der Ärzte und Patienten Daten zum Training von KI zur Verfügung zu stellen

5 Diskussion

Das primäre Ziel der Dissertation bestand darin, die Einstellung von Patienten zum eigenständigen Einsatz von KI zur Selbstdiagnostik sowie von Haut- und Hausärzten zum unterstützenden Einsatz von KI im Rahmen der Hautkrebsdiagnostik zu erörtern. Dabei sind sowohl der aktuelle Stand der Technik als auch die ethischen und rechtlichen Voraussetzungen für den Einsatz von KI in der Medizin berücksichtigt worden. Daraus sollen Hürden abgeleitet werden können, welche einem flächendeckenden Einsatz von KI im Wege stehen.

5.1 Einsatz von KI aus Sicht der Patienten

Der primäre Endpunkt in der Patientenbefragung wurde durch die Frage nach dem eigenständigen Einsatz eines KI-gestützten Programms zur Selbstdiagnose, vor dem Besuch bei einem Arzt, abgebildet. An dieser Stelle konnte kein signifikanter Unterschied zwischen der Stimulus- und der Kontrollgruppe gezeigt werden ($p > 0,05$). In Erwartung, dass das Vorwissen der Patienten den Effekt der Aufklärung beeinflusst hat, sind anschließend selektiv die Patienten betrachtet worden, welche ein vergleichsweise geringes Vorwissen angegeben haben (Wert von 1 bis 3). Es ist anzunehmen, dass der Effekt der Aufklärung durch diese Einschränkung stärker hervortritt. Entgegen der Erwartung war dies nicht der Fall. Mögliche Erklärungen für diese Beobachtung sind zum einen eine mangelnde Qualität der Aufklärung sowie der nicht abzuschätzende Grad des Verständnisses vonseiten der Patienten. Zum anderen kann nicht nachvollzogen werden, wer von den Patienten die Aufklärung tatsächlich gelesen hat. Es zeigt sich jedoch, dass das Vorwissen der Patienten in Bezug auf KI die Einstellung zu dem selbstständigen Einsatz von KI in der Hautkrebsdiagnostik beeinflusst. Betrachtet man die Werte von 4 und 5 als Zustimmung für den selbstständigen Einsatz von KI, so haben unabhängig vom Vorwissen 45,5% der Befragten gesagt, dass sie sich selbst mittels KI untersuchen würden. In der Literatur findet sich eine Zustimmung von 56% für den Einsatz von KI im häuslichen Umfeld (82). Dabei muss jedoch berücksichtigt werden, dass in der zum Vergleich herangezogenen Studie 88% der Befragten angegeben haben, eine ungefähre Vorstellung von einer KI zu haben (82). In der Befragung von den Patienten haben jedoch 28,8% angegeben, keine Vorstellung von der Funktionsweise von KI zu haben. Betrachtet man jedoch nur diejenigen Patienten, welche ein Vorwissen von 4 oder 5 angegeben haben, so ergibt sich mit einer Zustimmung von 58,8% ein mit der Angabe in der Literatur vergleichbarer Wert. Es ist jedoch einschränkend zu sagen, dass diese Frage nur von denjenigen Patienten beantwortet worden ist, welche zuvor den Einsatz von KI nicht komplett ausschlossen.

Ein weiterer wichtiger Aspekt wurde durch die Frage nach der Befürwortung des Einsatzes von KI in der Klinik abgebildet. In diesem Fall konnte kein signifikanter Unterschied zwischen der Befürwortung in der Stimulusgruppe und der Befürwortung in der Kontrollgruppe gezeigt werden ($p > 0,05$). Jedoch hat sich insgesamt eine deutliche Zustimmung für den Einsatz von KI in der Klinik gezeigt (Stimulusgruppe, Mittelwert=3,94 vs. Kontrollgruppe, Mittelwert=3,78). Insgesamt unterstützen 61,2% der Patienten (Werte von 4 und 5) den Einsatz von KI zur Hautkrebsdiagnostik in der Klinik. Bei der systematischen Literaturrecherche haben sich keine unmittelbar vergleichbaren Werte finden können. Betrachtet man nur die Patienten mit hohem Vorwissen, liegt die Zustimmung bei 65%. Jedoch haben in einer anderen Befragung etwa 90% der Befragten gesagt, dass sie den unterstützenden Einsatz von KI in der Hautkrebsdiagnostik befürworten (82). Ein möglicher Grund für die Diskrepanz zwischen diesen Werten beruht darin, dass die Zustimmung zu einem unterstützenden Einsatz eher gegeben wird, als allein auf künstliche Intelligenz zu setzen (stand-alone System). Dennoch zeigt sich in beiden Fällen die Zustimmung der Mehrheit zum Einsatz von KI zur Hautkrebsdiagnostik.

Ein wichtiger Faktor für die unkomplizierte Etablierung einer für Patienten zur Verfügung stehenden KI stellt die Nutzung digitaler Endgeräte dar. Da bereits KI-Systeme zur Hautkrebsdiagnostik in Form von Apps für Smartphones existieren, bietet das Smartphone

optimale Voraussetzungen für einen flächendeckenden Einsatz von KI zur Hautkrebsdiagnostik. 93,5% der Patienten haben angegeben ein Smartphone oder ein anderes internetfähiges Endgerät zu besitzen. In einer vergleichbaren Befragung haben 90% der Befragten angegeben, wenigstens ein elektronisches Endgerät zu besitzen (83). Es zeigt sich, dass viele Patienten über ein KI-System für das Smartphone erreicht werden können. Ebenso hat sich aus der Umfrage ergeben, dass das Smartphone bereits ein fester Bestandteil zur Eigendiagnostik der Patienten ist. Insgesamt haben 87,2% der Patienten angegeben, ihr mobiles Endgerät zum Nachschlagen von Krankheitssymptomen einzusetzen. Somit zeigt sich eine grundlegende Bereitschaft, das Smartphone auch zur eigenen Gesundheit einzusetzen. Schließlich haben 32,0% der Patienten explizit angegeben, KI in Form einer App für das Smartphone oder das Tablet einsetzen zu wollen. Folglich könnte etwa ein Drittel aller Patienten über diesen Weg erreicht werden. 29,9% der Befragten gaben an, KI als ein selbstständiges Endgerät einsetzen zu wollen. Hier zeigt sich möglicherweise eine Tendenz, die Hautkrebsdiagnostik von persönlichen Datenträgern trennen zu wollen. Ein möglicher Grund für die Entscheidung hin zu dem selbstständigen Endgerät könnte die Angst um die Sicherheit persönlicher Daten auf dem Smartphone der Patienten sein. Auch ist ein erhöhtes Zutrauen in ein eigens für den Einsatz von KI entwickeltes Endgerät denkbar. An dieser Stelle ist zu überlegen, ob solche selbstständigen Endgeräte zusätzlich zu den bereits den Patienten zugänglichen Apps entwickelt werden sollten, um den Einsatz von KI weiter voranzutreiben. Im Zusammenhang mit der Frage nach der Form des selbstständigen Einsatzes von KI hat sich gezeigt, dass Patienten aus der Kontrollgruppe signifikant häufiger angegeben haben, unter keinen Umständen eine KI einsetzen zu wollen, als es in der Stimulusgruppe der Fall gewesen ist ($p < 0,05$). Es ist davon auszugehen, dass die Aufklärung einen positiven Einfluss auf die Bereitschaft der Patienten hatte, KI einzusetzen. Jedoch ist zu berücksichtigen, dass sich diese Signifikanz bei der zuvor diskutierten Frage, bei der explizit nach dem selbstständigen Einsatz von KI gefragt worden ist, nicht gezeigt hat. Es ist davon auszugehen, dass weitere Faktoren diese Frage beeinflusst haben.

Dass der Einsatz von KI zur Hautkrebsdiagnostik nicht nur für den Fall der hautärztlichen Konsultation sinnvoll ist, zeigen die Befragungsergebnisse in **Tabelle 1**. In lediglich 32% der Fälle ist der Hautkrebs durch den Hautarzt entdeckt worden. In 24% der Fälle haben die Patienten den Hautkrebs selbst entdeckt und in 24% der Fälle ist der Hautkrebs von Hausärzten gefunden worden. Es zeigt sich, dass sowohl die Patienten selbst als auch die Hausärzte wichtige Initiatoren für die Diagnostik von Hautkrebs sind. Folglich scheint es sinnvoll, an dieser Stelle unterstützend mit KI einzugreifen.

Durch die vorhergehende Aufklärung konnte in unserer Untersuchungskohorte die Verunsicherung über den Einsatz von KI durch Ärzte reduziert werden. Patienten aus der Stimulusgruppe waren signifikant weniger verunsichert als es bei Patienten aus der Kontrollgruppe der Fall gewesen ist ($p < 0,05$). Insgesamt gaben in der Kontrollgruppe 20,0% der Patienten an, auf jeden Fall durch den Einsatz von KI verunsichert zu werden. Dieser Anteil lag in der Stimulusgruppe bei 5,7%. Auch wenn der Anteil dieser verunsicherten Patienten am Gesamtkollektiv vergleichsweise niedrig ist, gilt dies vor dem Einsatz von KI durch Ärzte zu berücksichtigen, vor allem auch in Hinblick auf eine gute Arzt-Patienten-Beziehung, in welcher Vertrauen eine wichtige Rolle spielt. Hieraus lässt sich ableiten, dass eine Aufklärung zu der Thematik „KI“ vor deren diagnostischen Einsatz durchaus empfehlenswert ist. Dadurch kann die Verunsicherung der Patienten gegenüber dieser Technik deutlich gesenkt werden. Es bedarf jedoch weiterer Studien über Art und Umfang einer effektiven Aufklärung über KI. Bei einer systematischen Literaturrecherche ließen sich keine vergleichbaren Werte finden.

Für den Fall, dass die Diagnosen von Arzt und künstlicher Intelligenz voneinander abweichen und damit eine wesentliche therapeutische Konsequenz einhergeht (z.Bsp. Nävus entfernen oder belassen), liegt das Vertrauen der Patienten deutlich auf Seiten der Ärzte.

Keiner der befragten Patienten hat angegeben, auf jeden Fall auf die KI vertrauen zu wollen. Die Mittelwerte lagen bei 1,62 (Kontrollgruppe) und 1,88 (Stimulusgruppe), wobei 1 für das Vertrauen in die Ärzte und 5 für das Vertrauen in die KI steht. Hier zeigt sich, dass das Vertrauen in den Arzt als menschlichen Experten deutlich stärker ist als das Vertrauen in die Maschine in Form einer KI. Vor dem Hintergrund, KI als eine Art Hilfsmittel für den Arzt einzusetzen, ist es auf den ersten Blick durchaus sinnvoll, wenn die höchste Autorität beim Arzt liegt, der die Verantwortung trägt. Diese Art der Entscheidung ist jedoch als ein vereinfachtes und künstliches Szenario zu betrachten, welches nicht die Komplexität der Realität widerspiegelt. So stellt die Entnahme einer Biopsie beispielsweise eine gängige Maßnahme bei unklarer diagnostischer Lage dar. Die überwiegende Mehrheit der Patienten hat bei vergleichbaren Befragungen diesen Weg für den Fall einer Konfliktsituation gewählt (82, 83). Die Entnahme einer Biopsie stand in dieser Befragung allerdings nicht als Auswahloption zur Verfügung. Auf diese Weise sollten die Befragten sich bewusst zwischen KI und Arzt unterscheiden, um die Vertrauensverhältnisse realistischer abzubilden. Die Compliance ist eng mit dem Thema Vertrauen verknüpft. Bei einer Onlinebefragung hat sich gezeigt, dass die Compliance größer war, wenn die Empfehlung von einem Arzt oder der Kombination aus Arzt und KI ausgesprochen worden ist, als wenn dies nur durch eine KI geschah (85). Hier zeigt sich ebenfalls, dass das alleinige Vertrauen bzw. die Compliance gegenüber der Empfehlung, welche lediglich von einer KI ausgesprochen worden ist, geringer ist als das Vertrauen in einen Arzt.

Basierend auf einer Auswahl verschiedener Optionen hat sich die „Diagnosesicherheit“ als das wichtigste Kriterium einer KI für den Patienten herausgestellt (42,1%). Dieses Verlangen nach Sicherheit lässt sich analog zu der bereits zuvor diskutierten Unsicherheit der Patienten betrachten, welche durch den Einsatz von KI hervorgerufen werden kann. Diese Entscheidung ist rational nachvollziehbar, schließlich hat eine Fehldiagnose das Potential einer tödlichen Konsequenz. An zweiter Stelle wurde die „anschließend schnelle Terminvermittlung zum Arzt“ genannt. Diese Vermittlung scheint eine logische Konsequenz auf das Treffen einer Therapie bedürftigen Diagnose zu sein. Dies ist vor allem für diejenigen Fälle von äußerster Wichtigkeit, bei denen Tage oder Wochen entscheidend für das Überleben der Patienten sind. In diesem Zusammenhang ist eine mit dem Einsatz von KI automatische Therapievermittlung in Betracht zu ziehen. Jedoch wird man möglicherweise an ethische Grenzen stoßen, wenn es darum geht, einen Algorithmus über die Priorisierung von Patiententerminen entscheiden zu lassen. Hier stellt sich unter anderem die Frage, ob es gerecht ist, die Leiden zweier Individuen gegeneinander abzuwägen. Eine Lösung für dies Frage sollte nur gemeinsam und im Einklang mit allen Beteiligten gefunden werden. Es folgt die Kategorie „jederzeit Verfügbarkeit“, welche ein starkes Argument für die Etablierung von KI im Rahmen der Hautkrebsdiagnostik ist. Auf diese Weise können Patienten erreicht werden, welche aufgrund zahlreicher Gründe nicht oder nur erschwert einen Arzt aufsuchen können. KI bietet folglich die Möglichkeit, zusätzliche Zielgruppen zu erreichen. Zu diesen gehören unter anderem Menschen, die weite Wege zur ärztlichen Versorgung zurücklegen müssen, aus zeitlicher Sicht keinen Arzt aufsuchen würden oder Hemmungen haben, zum Arzt zu gehen. Dieser Vorteil der Verfügbarkeit wird in der Literatur ebenfalls angeführt (83). An letzter Stelle kommt die „Erstattung durch die Krankenkasse“. Auch wenn keine Preise genannt worden sind, haben sich die Patienten am wenigsten für eine Erstattung der Leistung durch den Einsatz von KI entschieden. Dieser Punkt war aus Sicht der Patienten eher nachrangig. Bisher lässt sich keine KI basierte App im Verzeichnis für digitale Gesundheitsanwendungen des Bundesinstituts für Arzneimittel und Medizinprodukte finden (26). Einschränkend ist zu nennen, dass dies eine vordefinierte Liste mit vier Kriterien war und es wahrscheinlich weitere Kriterien in Bezug auf den Einsatz von KI gibt, welche den Patienten ebenfalls wichtig sind.

Eine geringe Mehrheit der Patienten tendierte dazu, den Gebrauch von KI Freunden und Verwandten weiterzuempfehlen, wobei diese Tendenz in der Stimulusgruppe stärker ausfiel als es in der Kontrollgruppe der Fall war (Mittelwert=3,51 vs. Mittelwert=3,16). Auch wenn dies eine hypothetische Frage ist und diese keine effektive Erfahrung mit dem Umgang von KI voraussetzt, lässt sich hieran die Tendenz der Patienten erkennen, KI weiterzuempfehlen, welche für einen großflächigen Einsatz von KI von Vorteil sein könnte. Insgesamt gaben in der

Stimulusgruppe 54,8% und in der Kontrollgruppe 31,4% der Patienten an, dass sie auf jeden Fall oder eher den Einsatz von KI an Freunde und Verwandte weiterempfehlen würden. Dieser Anteil stellt sich als deutlich geringer heraus, im Vergleich zu den in der Literatur angegebenen 75% (bzw. 71% für stand-alone Systeme) (83). Eine mögliche Erklärung ist die Tatsache, dass die Vergleichsdaten im Rahmen eines halbstrukturierten Interviews erhoben worden sind. Hierbei kann die Tendenz der Zustimmung stärker ausgeprägt gewesen sein, als es bei der ausschließlich schriftlichen Befragung der Fall gewesen ist.

Schließlich zeigen die Anmerkungen der Patienten, dass deren Vorstellung von dem Einsatz von KI in der Hautkrebsdiagnostik sich mit den Thesen der aktuellen Studienlage decken. Wie in der Literaturdiskussion bereits dargelegt worden ist, besteht die Tendenz dazu, KI als eine Art diagnostisches Hilfsmittel einzusetzen. Das primäre Ziel ist nicht, den Arzt vollkommen zu ersetzen. Dies haben zwei Patienten in ihren Anmerkungen widerspiegelt, bei denen sie zum einen klarstellten, dass KI hilfreich sei, jedoch kein Ersatz darstelle und zudem anmerkten, dass neben dem Einsatz von KI zusätzliche Maßnahmen ergriffen werden sollten, wie die Abnahme eines Labors oder die Untersuchung durch einen Arzt. Ein weiterer Patient wies auch darauf hin, dass er als Dachdecker oft der Sonne ausgesetzt sei und aus diesem Grund den Einsatz von KI begrüße. Aufgrund seiner überproportional hohen Sonnenexposition kann er zu einer Risikogruppe gezählt werden und bedarf deshalb einer besonderen Aufmerksamkeit in Hinblick auf die Entstehung von Hauttumoren. Regelmäßige und selbstständige Screening Maßnahmen unter Zuhilfenahme von KI könnten in diesem Fall eine Erleichterung für den Patienten sowie eine Entlastung für den Arzt darstellen. Der zeitliche Aufwand, welcher mit den regelmäßigen Arztbesuchen einhergehen würde, kann auf diese Weise reduziert werden.

5.2 Einsatz von KI aus Sicht der Ärzte

Der primäre Endpunkt in der Ärztebefragung wurde durch die Frage nach der Bereitschaft, KI als Assistenzsystem einzusetzen, abgebildet. Hierbei konnte kein signifikanter Unterschied zwischen der Bereitschaft in der Stimulus- und der Kontrollgruppe gezeigt werden. Jedoch konnte gezeigt werden, dass die Bereitschaft unter den Ärzten grundlegend hoch war (Mittelwert=4,0). Diese Bereitschaft wurde durch die Aufklärung, das Alter und die Fachrichtung nicht wesentlich beeinflusst. Insgesamt sagten 41,7% der Ärzte, dass sie auf jeden Fall KI einsetzen würden und 25,0% gab an, KI eher einsetzen zu wollen. In der Literatur lagen diese Werte bei einer Befragung unter Hausärzten bei 45% und 41% (88). Schließt man nur diejenigen Ärzte ein, welche ein hohes Vorwissen (Werte von 4 und 5) angegeben haben, so lag die Bereitschaft für den Einsatz von KI unter den Dermatologen bei 80,9%. Bei den Hausärzten betrug der Anteil 100%, wobei zu berücksichtigen ist, dass dieser Wert aus lediglich 5 Fällen zusammengesetzt ist. Es zeigt sich, dass mit genauerer Kenntnis über KI die Bereitschaft unter den Ärzten für deren Einsatz steigt. Auch wenn die Aufklärung keinen signifikanten Einfluss auf die Bereitschaft der Ärzte hatte, KI als Assistenzsysteme einzusetzen, ein möglicher Grund stellt in diesem Zusammenhang die allgemein hohe Bereitschaft der Ärzte für den Einsatz von KI dar, stellt sich jedoch deutlich heraus, dass das Vorwissen der Ärzte einen Einfluss auf die Bereitschaft für den Einsatz von KI hatte. Entsprechend der Überlegungen aus der Diskussion der Patientenbefragung kann hier eine mangelnde Qualität der Aufklärung sowie eine fehlende Überprüfbarkeit der korrekten Bearbeitung der Befragung als Erklärung zugrunde gelegt werden. Der zum Vergleich herangezogene Wert wurde im Rahmen einer Befragung erhoben, bei welcher explizit die diagnostischen Schwierigkeiten von Hausärzten angesprochen worden sind. In diesem Zusammenhang ist die KI als Lösungsansatz eingeführt worden. Es ist anzunehmen, dass dies die Zustimmung für den Einsatz von KI erhöht hat.

Vor diesem Hintergrund hat sich die Frage gestellt, ob das Verständnis für die Funktionsweise von KI eine Voraussetzung für deren Einsatz sein sollte. Entgegen der Erwartung zeigte sich kein signifikanter Unterschied zwischen dem Mittelwert der Stimulusgruppe und dem Mittelwert

der Kontrollgruppe ($p > 0,05$). In der Subgruppe „Fachrichtung“ zeigten sich hingegen deutliche Unterschiede. Die Dermatologen gaben in der Kontrollgruppe im Mittel eine 3,08 und in der Stimulusgruppe eine 4,13 an. Bei den Hausärzten zeigte sich jedoch ein umgekehrtes Muster. Hier ergab sich für die Ärzte in der Stimulusgruppe ein Mittelwert von 2,75 und für die Kontrollgruppe ein Mittelwert von 4,06. Aus diesem Grund ist davon auszugehen, dass weitere Faktoren einen wesentlichen Einfluss auf die Einstellung der Ärzte hatten, sodass der Effekt der Aufklärung schwer zu beurteilen ist. Grundsätzlich zeigte sich ein deutlicher Unterschied zwischen den Geschlechtern der Befragten. Die Ärztinnen sahen das Verständnis um die Funktionsweise von KI eher als Voraussetzung für deren Einsatz an, als die Ärzte (Mittelwert=4,12 vs. Mittelwert=3,19). Ein Grund dafür könnte das stärkere Bewusstsein der Ärztinnen für die Gefahren, welche mit dem unachtsamen Einsatz von KI einhergehen, sein (90). Ebenso spielte das Alter der Befragten eine Rolle. Die jüngeren Ärzte (jünger gleich 50 Jahre) erachteten das Verständnis als wichtiger als die älteren Ärzte (älter als 50 Jahre) (Mittelwert=3,80 vs. Mittelwert= 3,17). Eine Erklärung dafür ist die möglicherweise größere Sensibilisierung und ein größeres Interesse der jüngeren Generationen für moderne Technologie. Bei der systematischen Literaturrecherche ließen sich keine direkt vergleichbaren Werte finden.

Im Hinblick auf die Zukunft wurde zudem nach dem Potential von KI in Bezug auf die Hautkrebtsdiagnostik gefragt. Auch wenn sich kein statistisch signifikanter Unterschied zwischen den Mittelwerten der Stimulus- und der Kontrollgruppe zeigte ($p > 0,05$), stellte sich deutlich das von der überwiegenden Mehrheit zugesprochene Potential der KI heraus (83,3%), welches mit den Angaben in der Literatur vergleichbar ist (77%) (87). Berücksichtigt man, dass bei dem zum Vergleich herangezogenen Wert nur Dermatologen befragt worden sind, nähern sich die Ergebnisse noch weiter an (80% vs. 83,3%). Die Subgruppen „Alter“, „Geschlecht“ und „Fachrichtung“ ließen dieses Ergebnis unbeeinflusst.

Warum neben Dermatologen auch Hausärzte in diese Befragung eingeschlossen worden sind, ergibt sich aus der Tatsache, dass diese ebenfalls einen wichtigen Faktor in Bezug auf die Hautkrebtsdiagnostik darstellen. Es hat sich gezeigt, dass 46,7% der Hausärzte häufig bis täglich mit der Diagnostik von Hautkrebs konfrontiert sind. Dieser Anteil lag bei den Dermatologen bei 96,6%. Der unterstützende Einsatz von KI würde folglich nicht nur den Hautarzt, sondern auch den Hausarzt betreffen, welcher von dem Einsatz von KI ebenfalls profitieren würde. Dies gilt es unter anderem vor dem Hintergrund zu berücksichtigen, dass bei einer Befragung 98% aller Hausärzte angaben, Schwierigkeiten zu haben, nicht melanozytären Hautkrebs zu diagnostizieren (88).

Des Weiteren hat sich gezeigt, dass KI bereits von einigen Dermatologen eingesetzt wird. 18,3% der Dermatologen gaben in der Befragung an, KI aktuell zur Hautkrebtsdiagnostik zu verwenden. Bei den Hausärzten hingegen gab keiner der Befragten an, KI einzusetzen. Ein möglicher Grund dafür ist, dass die Hautkrebtsdiagnostik primär zum Fachgebiet der Dermatologie gezählt werden kann und nur einen sehr geringen Anteil an der Arbeit eines Hausarztes ausmacht. Zudem hat eine Umfrage unter Hausärzten in Frankreich gezeigt, dass 68% der Ärzte nicht für eine KI zahlen würden (88). Möglicherweise waren die Hausärzte der Meinung, dass der Nutzen die Kosten nicht aufwiegen würde. Allerdings sind keine konkreten Kosten für eine KI benannt worden.

Insgesamt hat mit einem Mittelwert von 4,02 die Mehrheit der Ärzte angegeben, dass sie den generellen Einsatz von KI im Bereich der Hautkrebtsdiagnostik befürworten würde. Der Anteil derjenigen, die den Einsatz auf jeden Fall oder eher befürworten würden, lag bei den Dermatologen bei 76,0% und bei den Hausärzten bei 68,2%. Damit liegt die Zustimmung niedriger als es bei einer vergleichbaren Studie unter Hausärzten mit etwa 86% der Fall gewesen ist. Für diese Studie ist jedoch direkt die Schwäche der Ärzte in Bezug auf die Diagnostik nicht melanozytärer Hauttumore angesprochen worden mit der Absicht, diese durch den Einsatz von KI zu minimieren (88).

Ein wichtiger Aspekt stellt auch die Überlegung dar, wie im Falle eines Konfliktes zwischen Arzt und KI vorgegangen werden sollte. Dies gilt sowohl für den Fall, dass die Entscheidung

der KI zu einer therapeutisch relevanten Diagnose führt als auch für den Fall, dass der Arzt diese trifft, während die andere Partei diese jeweils verneint. Für diese Fälle haben 69,0% bzw. 65,5% der Ärzte gesagt, dass sie eine Biopsie nehmen würden. Jedoch gaben für den ersten Fall 3,4% der Ärzte an, dennoch therapieren zu würden, während es im zweiten Fall 15,5% waren. Es ist anzunehmen, dass das Vertrauen der Ärzte in die eigenen Fähigkeiten größer war als das in KI. Um diesen Konflikt zu lösen, entschied sich die überwiegende Mehrheit für die Entnahme einer Biopsie. Die Entnahme einer Biopsie ist nicht nur aus Sicht der Ärzte, sondern auch aus der, der Patienten eine wichtige Lösung (82, 83). Dieser Konsens in Bezug auf das therapeutische Vorgehen ist ein wichtiger Punkt für die Befürwortung des Einsatzes von KI aus Sicht beider Parteien.

Eine wichtige Frage ist auch, ob KI Teil der Facharztausbildung sein sollte. Man kann diskutieren, ob dieses Thema bereits in dem Studium behandelt werden sollte, schließlich ist der Einsatz von KI keineswegs allein auf den Fachbereich der Dermatologie beschränkt. Vielleicht kann KI auch schon als Grundlagenwissen für das Studium vorausgesetzt werden. Dass dies nicht der Fall ist, hat eine Studie in Südkorea gezeigt, bei welcher 6% der befragten Medizinstudenten und Ärzte angegeben haben, gut mit dem Thema KI vertraut zu sein (86). Dieser Anteil lag in der Ärztebefragung bei 13,3%. Ein möglicher Grund dafür ist, dass in der Datenerhebung keine Studenten angesprochen worden sind. Der Anteil an Studenten war jedoch in der verglichenen Studie mit 18,1% relativ gering (86). Insgesamt haben 51,7% der Ärzte gesagt, dass sie die Aufnahme von KI in die Facharztausbildung auf jeden Fall oder eher befürworten würden, der Mittelwert lag bei 3,69. Somit lag eine geringe Tendenz hin zu der Aufnahme in die Ausbildung vor. In einer vergleichbaren Studie, bei der von einer allgemeinen Aufnahme von KI in die medizinische Ausbildung die Rede war, haben knapp 80% der Ärzte gesagt, dass sie diese unterstützen würden (87). Als eine Erklärung für diese Abweichung ist die unterschiedliche Adressierung des medizinischen Ausbildungsabschnitts anzunehmen. Während die zuvor aufgeführte Studie generell auf die medizinische Ausbildung Bezug genommen hat, ist bei der Erhebung nur die Facharztausbildung angeführt worden. Insgesamt scheint eine Mehrheit der Befragten KI in der medizinischen Ausbildung zu befürworten, es bleibt zu klären, wann das am besten geschehe.

Wie die Patienten hat auch unter den Ärzten die Mehrheit (81,7%) nach Wahlmöglichkeit angegeben, dass ihnen die Diagnosesicherheit für den Einsatz von KI am wichtigsten ist. Wäre diese nicht gegeben, könnte man den Sinn des Einsatzes von KI in Frage stellen. 11,7% der Ärzte sagten, dass die KI möglichst viele Differentialdiagnosen unterscheiden können sollte. Dies gilt es vor allem vor dem Hintergrund zu berücksichtigen, dass die KI bei ihrer unbekanntem Hauttumoren schnell fehleranfällig wird. Zudem stand eine Freitextantwort zur Verfügung, bei der die Zeitersparnis durch den Einsatz von KI erwähnt worden ist. Würde der Einsatz zu einem zusätzlichen Zeitaufwand führen, könnte ein ethisches Dilemma entstehen, sofern durch den Einsatz von KI eine genauere Diagnostik ermöglicht werden würde. Dann müsste abgewogen werden, ob der diagnostische Mehrwert den Zeitverlust und eine möglicherweise damit verbundene kürzere Betreuung anderer Patienten überwiegt. Ebenso ist der Effekt des „Vorscreenens“ durch den Einsatz von KI angeführt worden, welcher ebenfalls auf die Idee der Zeitersparnis anspielt. Des Weiteren ist auch der Preis der KI erwähnt worden. Auch wenn diese Angabe nicht weiter ausgeführt worden ist, kann davon ausgegangen werden, dass der Einsatz einer KI erschwinglich sein sollte. Schließlich ist der Aspekt der Wahlfreiheit aufgegriffen worden. Hiermit ist mutmaßlich die Freiheit des Arztes gemeint ist, sich für oder gegen die Nutzung einer KI zu entscheiden. Dies ist eine auf längere Sicht wichtige Frage. Sollte der Einsatz von KI zur Pflicht werden, wenn dadurch die Diagnostik verbessert werden würde? Viele Aspekte wie der bereits erwähnte Preis und der mögliche Zeitgewinn oder auch Zeitverlust spielen dabei eine wichtige Rolle.

Den Ärzten standen vier Aspekte zur Auswahl, welche sie nach ihrer Wichtigkeit für die Anwendung einer KI priorisieren sollten. Hier hat sich die „Nachvollziehbarkeit der Ergebnisse“ als wichtigster Punkt herausgestellt. Dies ist eine wichtige Voraussetzung für das Vertrauen in die Anwendung einer KI (46, S. 16). Zudem können bei nachvollziehbaren Analysen Fehler der KI besser erkannt und somit verhindert werden, wenn beispielsweise der Fokus einer

Hautanalyse nicht auf der zu untersuchenden Hautläsion lag, sondern fälschlicherweise auf einem Haar. An zweiter Stelle wurde die „leichte Bedienbarkeit“ angeführt. Diese ist wichtig, um eine KI unkompliziert im ärztlichen Alltag zu etablieren. An dritter Stelle kam die „Analysegeschwindigkeit“. Es ist anzunehmen, dass eine schnelle Analysegeschwindigkeit und ein damit einhergehender Zeitgewinn einen positiven Einfluss auf die Bereitschaft zum Einsatz von KI hat. Es ist nachvollziehbar, dass der Einsatz von KI nicht mit einem erheblichen Mehraufwand verbunden sein sollte, schließlich soll die Hautkrebsdiagnostik durch deren Einsatz unterstützt werden. An letzter Stelle steht die „Delegierbarkeit“. Während die ersten drei genannten Punkte als Spezifikation der KI betrachtet werden können, bildet der letzte Punkt eine eher rechtliche Frage ab. Wenn sich der Einsatz von KI aufgrund deren diagnostischen Qualität und Effizienz lohnt, ist nicht auszuschließen, dass die Handhabung der KI delegiert werden könnte. Dies könnte mit einem weiteren Zeitgewinn für das ärztliche Personal einhergehen und ließe beispielsweise mehr Zeit für Patientengespräche und die Versorgung weiterer Patienten.

In der Freitextaufgabe haben die Ärzte zahlreichen Stärken einer KI identifiziert. Dabei war die Objektivität eine häufig genannte Eigenschaft. Diese Objektivität ermöglicht es, eine sachlich und emotional unabhängige Diagnose zu stellen und kann mit den Ergebnissen einer Laboruntersuchung verglichen werden, welche durch den Arzt in den Kontext der Patientengeschichte gerückt wird. Ein ebenfalls mehrfach angeführter Punkt war die Verfügbarkeit der KI, welche mehrdimensional zu betrachten ist. Als Beispiel lässt sich hier eine MRT (Magnetresonanztomographie) heranziehen. Ein MRT ist teuer und nicht überall verfügbar und zudem nicht portabel (91). Eine KI hingegen kann unkompliziert auf Smartphones etabliert werden (53). Sie ist portabel und kann überall und jederzeit eingesetzt werden. Als weitere Stärke, in diesem Fall vor allem in Bezug auf die ambulante Versorgung, ist die Einholung einer Zweitmeinung durch die KI benannt worden. Während in der Klinik erfahrenere Kollegen gefragt werden können, stehen diese im Praxisalltag häufig nicht zur Verfügung. Hier bietet die KI eine unkomplizierte Möglichkeit, eine Diagnose zu „diskutieren“. An diesen Punkt knüpft der Aspekt der „diagnostischen Sicherheit“ an. Stellt sich die KI als robust für den medizinischen Alltag heraus, kann sie die Hautkrebsdiagnostik optimieren und dem Arzt Sicherheit beim Fällen einer Diagnose geben. Dies ist vor allem für weniger erfahrene Ärzte hilfreich sowie für diejenigen, welche nicht direkt Kollegen in ihrem Arbeitsumfeld haben, mit denen sie sich beraten können. Als weitere Stärke ist die Standardisierbarkeit genannt worden, welche Diagnosen vergleichbarer macht. Dies wäre auch vor dem Hintergrund der interdisziplinären Kommunikation von Bedeutung, für den Fall, dass die KI auch in anderen Fachbereichen eingesetzt würde. Auch ist festgestellt worden, dass der Einsatz von KI für das Hautkrebscreening hilfreich wäre. Dabei könnten aufgrund der Schnelligkeit Kosten eingespart werden. Schließlich ist die KI lernfähig und kann sich aufgrund dessen gut neues Wissen aneignen und somit ständig optimiert werden.

Die befragten Ärzte haben zudem einige Schwächen der KI benannt. Zum einen ist die fehlende menschliche Interaktion genannt worden und die damit einhergehende fehlende emotionale Betreuung der Patienten. Dieser Kritikpunkt bezieht sich auf die stand-alone Anwendung einer KI und ist bereits bekannt (83). Auch wird eine fehlende Verlässlichkeit in die Fähigkeiten einer KI angeprangert. Sie sei fehleranfällig, was veränderte Kontextfaktoren, Artefakte und Manipulationen betrifft. Dies hat sich auch in mehreren Studien gezeigt, sodass nach Lösungsansätzen gesucht wird (79, 90). Vor diesem Hintergrund gilt es den Aspekt der Schadensverhütung zu wahren. Der Einsatz von KI soll keinen Schaden verursachen (46). Logisch ist, dass Läsionen oder Hauttypen, welche nicht oder kaum beim Training eingesetzt worden sind, schwer erkannt werden können. Aus diesem Grund ist es wichtig, verschiedene Ethnien ins Training einer KI einzuschließen (76). Auch wurde gesagt, dass die Leistung der KI sinkt, wenn die realen Kontextbedingungen beim Einsatz einer KI von denen des Trainings abweichen. Das hat sich in Studien ebenfalls gezeigt (79, 90). Eine getestete Methode, um die Varianz der Ergebnisse bei abweichenden Kontextfaktoren zu verringern, ist die Vergrößerung des Trainingssets (90). Zudem ist die Gefahr der Fehlerfortpflanzung aufgrund der Fähigkeit des Selbstlernens benannt worden. Um dies zu verhindern, ist auf die Qualität des Trainingsmaterials zu achten. Hier bietet sich an, nur histologisch gesichertes Datenmaterial

zu verwenden. Zudem sollte die AOI auf der zu untersuchenden Hautläsion liegen und nicht auf einer unwichtigen, miterfassten Hautstruktur. Als weiterer Kritikpunkt ist angeführt worden, dass KI von den Patienten akzeptiert werden muss, um diese einsetzen zu können. Dass der Einsatz von KI im Bereich der Dermatologie auf große Akzeptanz trifft, hat sich in der Datenerhebung sowie in weiteren Patientenstudien gezeigt (82, 83). Ein Arzt hat zudem die Angst vor KI benannt. Die Tatsache, dass eine gewisse Angst vor dem Einsatz von KI in der Hautkrebsdiagnostik unter Ärzten vorhanden ist, ist nicht unbekannt. In einer Studie unter Ärzten haben etwa 20% angegeben, Angst in Bezug auf den Einsatz von KI in der Dermatologie zu haben (87). Dennoch hat sich bei der Befragung gezeigt, dass die Mehrheit der Ärzte bereit ist, KI einzusetzen, was auch in weiteren Studien bestätigt worden ist (87, 88). Dies bedeutet jedoch nicht, dass die Angst unberücksichtigt bleiben sollte. Stattdessen sollte sie gezielt adressiert werden und nach Lösungsansätzen gesucht werden, wie diese reduziert werden kann. Eine mögliche Ursache für die Angst kann die Ungewissheit in Bezug auf die Funktionsweise einer KI sein. In diesem Zusammenhang ist es auch aus ethischer Sicht wichtig, die Erklärbarkeit der KI-Prozesse nicht zu vernachlässigen (46). Es ist anzunehmen, dass die Nachprüfbarkeit von Entscheidungsprozessen der KI-Anwendungen zu einer Reduktion der Angst führen kann.

Als weitere Schwäche ist die Tatsache angeführt worden, dass trotz des Einsatzes von KI die Verantwortung immer noch beim Arzt läge. Dass das Stellen von Diagnosen grundsätzlich in den ärztlichen Aufgabenbereich fällt und die Ärzte dafür die Verantwortung übernehmen, ist ein gängiges Prinzip. In diesem Zusammenhang kann hinterfragt werden, inwieweit der Arzt der KI vertrauen kann und Verantwortung für den unterstützenden Einsatz von KI übernehmen will. Es gilt an dieser Stelle zu berücksichtigen, dass in der digitalen Medizin und auch in Bezug auf KI eine Multiakteurverantwortung vorliegt (45, S. 249, 46, S. 18-19). Außerdem steht bislang keine eindeutige Haftungsregelung fest. Jedoch existieren Empfehlungen des EU-Parlaments für die Regelung der zivilrechtlichen Haftung beim Einsatz künstlicher Intelligenz (35). Das Gefühl der „falschen Sicherheit“ ist ebenfalls benannt worden. Es ist anzunehmen, dass damit gemeint ist, der KI nicht blind zu vertrauen. Ausgehend von der Idee, KI als unterstützendes Hilfsmittel für Ärzte zu etablieren, ist vorauszusetzen, dass diese sich zumindest bedingt mit der Diagnostik von Hautkrebs auskennen. Versteht sich der Anwender nicht darauf oder ist ungeübt im Umgang mit der KI, kann die Leistung der KI abnehmen (92). Ein weiterer Arzt hat angemerkt, dass KI zu einer Einsparung von Fachärzten führen könnte. Es kann als Vorteil angesehen werden, dass KI den zeitlichen Vorgang der Hautkrebsdiagnostik optimiert. Jedoch ist fraglich, inwieweit daraus eine personelle Einsparung resultiert, wenn man bedenkt, dass die KI lediglich als Diagnostikum dient. Die Therapie des Patienten bleibt nach wie vor dem Arzt überlassen. Abgesehen davon stellt der Hautkrebs nur einen geringen Anteil der dermatologischen Krankheiten dar. Vielmehr kann überlegt werden, wie der zeitliche Gewinn genutzt werden kann, die allgemeine Patientenversorgung zu verbessern.

Eine andere Schwäche der KI sei ihre Unfähigkeit, zu therapieren. Wenn die KI neben der Diagnostik auch die Therapie übernehmen könnte, würde das den Arbeitsaufwand von Dermatologen verringern, sofern dieser Prozess nicht überwacht werden müsste und der zeitliche Gewinn würde zusätzlich gesteigert werden. In diesem Fall wäre eine Personaleinsparung zunehmend wahrscheinlicher. Die Übernahme der Therapie spiegelt allerdings nicht den Anspruch der aktuellen Forschung wider. Schließlich wurde auch ein „Preisverfall der ärztlichen Leistung“ bemängelt. Wie dieser im Zusammenhang mit den Schwächen einer KI zu verstehen ist, lässt sich nur mutmaßen. Der Einsatz von KI wird nach dem aktuellen Stand den Arzt bei der Hautkrebsdiagnostik nicht ersetzen, sondern lediglich unterstützen. KI für den Gebrauch durch Patienten soll unnötige Konsultationen verhindern und bietet dem Arzt Raum, sich auf andere Krankheitsfälle zu konzentrieren. Möglicherweise spielt dieser Kritikpunkt darauf an, dass die KI die Hautkrebsdiagnostik übernehmen könnte und der Arzt deshalb seinen Wert verliert. Dies ist jedoch nach aktuellem Stand nicht absehbar.

Eine weitere Anmerkung war, dass viele digitale Anwendungen den Praxisablauf hemmen würden. In diesem Zusammenhang spielen viele Faktoren wie beispielsweise das technische

Knowhow der Mitarbeiter, die Ausstattung der Praxis und die Integration der KI in den Praxisablauf eine wichtige Rolle. Prinzipiell bieten digitale Anwendungen die Möglichkeit, Daten schnell zu verarbeiten und weiterzugeben. Dies könnte sich im Praxisalltag als hilfreich erweisen. Es ist jedoch auch denkbar, dass es je nach Umsetzung zu Störungen im Praxisablauf kommen kann. Zuletzt wurde auch darauf hingewiesen, dass mehr Aufklärungsarbeit für KI notwendig sei. Diesem Aspekt liegt mutmaßlich der Gedanke zugrunde, dass die Angst vor oder die Misskonzeption von KI zu einer Ablehnungshaltung führt. Zugleich sollten die Anwender von KI für damit einhergehende Risiken sensibilisiert werden. Diese Idee hat Einzug in diese Arbeit gefunden und wurde in dieser näher untersucht.

5.3 Vergleich der Ansichten von Patienten und Ärzten

98,3% der Ärzte und 82,2% der Patienten gaben in der Befragung an, schonmal etwas von KI gehört zu haben. Damit wird deutlich, dass das Thema „KI“ den Ärzten vertrauter war als es bei den Patienten der Fall gewesen ist, auch wenn KI insgesamt für beide Parteien kein unbekanntes Thema zu sein schien. In der Literatur wird für die Patienten ein vergleichbarer Wert von 88% angegeben, der sich jedoch auch auf das Verständnis von KI bezieht (82).

In Bezug auf die persönliche Einschätzung zu dem Wissen um die Funktionsweise von KI zeigt sich, dass dieses Wissen von den Ärzten signifikant höher eingeschätzt worden ist, als es bei den Patienten der Fall war ($p < 0,05$). Dieser Unterschied deckt sich mit der allgemeinen Bekanntheit von KI unter diesen beiden Parteien. Da KI bereits seit einigen Jahren im Bereich der Hautkrebsdiagnostik getestet wird, ist anzunehmen, dass die Ärzte eher mit dieser Thematik konfrontiert werden als es bei den Patienten der Fall ist. Insgesamt schätzten 43,3% der Ärzte ihr Wissen als sehr gut oder gut ein. In der Literatur gaben bei einer Befragung lediglich 6% der Ärzte an, gut mit dem Thema KI vertraut zu sein (86). In einer weiteren Befragung haben 24% der Ärzte gesagt, sich gut mit dem Thema KI im Bereich der Dermatologie auszukennen (87). Während die zweite Befragung direkt auf das spezielle Gebiet der Dermatologie abzielt und somit nur bedingt zum Vergleich herangezogen werden kann, erscheint die Diskrepanz zu dem Ergebnis der zuvor angeführten Studie deutlich. Es ist anzunehmen, dass dies unter anderem darauf zurückzuführen ist, dass auch Medizinstudenten in diese Studie eingeschlossen worden sind, welche noch keinen beruflichen Kontakt zu KI hatten. Des Weiteren ist diese Studie bereits über drei Jahre alt, sodass auch dieser zeitliche Faktor bedacht werden muss. Bei den Patienten machte der Anteil derjenigen, welche eine sehr gute oder gute Vorstellung um die Funktionsweise von KI besitzen, 20,2% aus, während hingegen 28,8% sagten, keine Kenntnisse von KI zu haben. Diese Ergebnisse sind mit Zahlen aus der Literatur in etwa vergleichbar. Hier gaben 21% der Befragten an, keine Vorstellung von KI zu haben (83). Eine möglichere Erklärung für diesen Unterschied ist, dass der in der Literatur erhobene Wert in Form eines Interviews erhoben worden ist. In diesem Zusammenhang ist denkbar, dass die Patienten eher dazu neigten, zu behaupten, dass sie sich mit KI auskennen würden.

Deutlich divergierende Meinungen zeigen sich in Bezug auf den selbstständigen Einsatz von KI durch den Patienten. Während sich 45,5% der Patienten auf jeden Fall oder eher vorstellen konnten, diese einzusetzen, rieten lediglich 18,3% der Ärzte zu diesem Einsatz. Von der Erkenntnis aus der Befragung ausgehend, dass Ärzte stärker mit dem Thema KI vertraut sind, lässt sich vermuten, dass sich diese eher der eingeschränkten Aussagekraft einer KI bewusst sind, für den Fall, dass diese unsachgemäß gehandhabt wird. Im Gegenzug ist nicht auszuschließen, dass eine Tendenz dazu vorliegt, sich nicht durch eine Maschine ersetzen lassen zu wollen und das Vertrauen in die eigenen Fähigkeiten überwiegt. Für Patienten stellt der selbständige Einsatz von KI einen deutlichen Freiheits- und Selbstständigkeitsgewinn dar. Zudem sind diese vermutlich weniger für die Fehleranfälligkeit von KI im unsachgemäßen Umgang sensibilisiert. 38,3% der Ärzte sagten, dass sie diesen Einsatz auf keinen Fall empfehlen würden und 25,0% würden diesen eher nicht empfehlen. Somit haben geringfügig mehr Ärzte von dem selbständigen Einsatz von KI durch Patienten abgeraten, als es in der

Literatur mit 20% (sehr kritisch) und 37% (kritisch) der Fall gewesen ist (88). An dieser Stelle gilt es zu berücksichtigen, dass sowohl aufgeklärte als auch nicht aufgeklärte Ärzte in die Berechnung eingeschlossen worden sind. Für die Patienten wird in der Literatur angegeben, dass 56% KI selbständig einsetzen würden (82).

Im Falle einer bösartigen Diagnose durch eine KI beim selbstständigen Gebrauch durch den Patienten, waren sich Patienten und Ärzte einig, dass ein Arzt aufgesucht werden sollte (Mittelwert=4,93 vs. Mittelwert=4,81). Dies ist ohne Frage eine logische Konsequenz, unter der Voraussetzung, dass bei der überwiegenden Mehrheit der Fälle tatsächlich eine bösartige Hautveränderung vorliegt. Im Falle einer gutartigen Diagnose zeigte sich hingegen, dass Patienten signifikant häufiger den Arzt aufsuchen würden als Ärzte dies empfehlen würden (Mittelwert=4,43 vs. Mittelwert=3,97, $p < 0,05$). Es ist anzunehmen, dass bei den Patienten trotz der gutartigen Diagnose Bedenken vorliegen bzw. diese von der KI nicht ausgeräumt werden. Dennoch lässt sich auch in diesem Fall festhalten, dass sowohl Patienten als auch Ärzte eher dazu raten würden, einen Arzt aufzusuchen. Unabhängig von der Diagnose durch die KI scheint für Patienten wie auch Ärzte die Konsequenz zu sein, einen Arzt aufzusuchen. An dieser Stelle lässt sich der allgemeine häusliche Gebrauch einer KI durch den Patienten infrage stellen. Dennoch gibt es viele gute Gründe für den Einsatz von KI durch den Patienten, wie z. B. der erhöhte Zeitaufwand, welcher mit einem Arztbesuch in Verbindung steht und durch die KI reduziert werden könnte, eine weite Entfernung zu Versorgungszentren, sowie die Hemmschwelle, einen Arzt aufzusuchen. Es ist jedoch auch klar, dass, wenn alle Patienten mit einer gutartigen Diagnose zum Arzt gehen sollten, es zu vielen überflüssigen Konsultationen kommen würde. Bei der systematischen Literaturrecherche ließen sich keine unmittelbar vergleichbaren Werte finden.

Des Weiteren hat sich gezeigt, dass das Vertrauen von Patienten und Ärzten in die alleinige Diagnose durch eine KI gering ausfällt. 68,0% der Patienten sagten, kein oder kaum Vertrauen in das stand-alone System einer KI zu haben. Bei den Ärzten sind es 96,6% gewesen. Das Vertrauen der Ärzte fiel signifikant geringer aus als das Vertrauen der Patienten ($p < 0,05$). Lediglich 7,8% der Patienten sagten, dass sie auf jeden Fall oder eher auf die Diagnose der KI vertrauen würden, was sich auch in der Literatur widerspiegelt (12%) (84). Eine Erklärung für das geringe Vertrauen, ist das Wissen, um die Fehleranfälligkeit einer KI, ebenso wie das fehlende Wissen um die Entscheidungsfindungsprozesse der KI. Zudem ist eine gewisse psychologische Hemmschwelle zu erwarten, eher einer Maschine als einem Menschen zu vertrauen.

Würde die KI für den Einsatz der Hautkrebsdiagnostik flächendeckend eingesetzt werden, stellt sich die Frage, inwieweit dieser Einsatz verpflichtend wäre und wie die Patienten und Ärzte damit umgehen würden. Es hat sich gezeigt, dass 29,2% auf den zusätzlichen Einsatz von KI durch den Arzt bestehen würden. Zugleich gaben 55,0% der Ärzte an, dass sie diesem Wunsch der Patienten nachkommen würden. Hieraus lässt sich ein Druck auf die Ärzte ableiten, welcher durch die Erwartungshaltung der Patienten aufgebaut würde. Zugleich könnte eine Weigerung von Seiten der Ärzte, KI einzusetzen, die Arzt-Patienten-Beziehung negativ beeinflussen. In diesem Fall tritt ein Konflikt zwischen der Wahlfreiheit des Patienten, sich mittels KI untersuchen zu lassen und der Freiheit des Arztes, KI für die Untersuchungen einzusetzen, auf.

Um die KI weiter voranzutreiben, werden große Datenmengen an Trainingsmaterial benötigt. Es hat sich gezeigt, dass Patienten wie auch Ärzte mehrheitlich dazu bereit sind, Datenmaterial aus ihren Untersuchungen zur Verfügung zu stellen. 71,7% der Patienten gaben an, auf jeden Fall oder eher Daten zur Verfügung stellen zu wollen. Unter den Ärzten betrug dieser Anteil 75,0%. Diese Bereitschaft fiel im Vergleich zu den Referenzwerten, welche sich aus der Literaturrecherche ergaben, niedriger aus. Hier ergaben sich in zwei verschiedenen Studien unter Patienten Werte von 88% und 87% (82, 84). Eine Erklärung für diese Abweichung kann auf der Tatsache beruhen, dass bei der Befragung nicht explizit darauf hingewiesen worden ist, dass die Daten anonym verarbeitet werden würden. Zudem ist in der zweiten aufgeführten Studie lediglich nach dem Verständnis, dass die eigenen Daten zum

Training von KI genutzt werden könnten und nicht explizit nach der Bereitschaft, diese dazu freizugeben, gefragt worden. Die Angst vor einem potenziellen Missbrauch der eigenen Daten kann dazu beigetragen haben, dass die Bereitschaft geringer ausfiel. Dennoch bietet die große Bereitschaft Gesundheitsdaten freizugeben das Potential, KI weiter voranzutreiben.

5.4 Limitationen

Limitationen ergeben sich zum einen aus der Durchführung der Befragung. In Bezug auf die Ärzte ist hier allen voran die Auswahl der Befragten zu nennen. Es sind nur Praxen eingeschlossen worden, welche auf „Google Maps“ eingezeichnet waren, eine Homepage hatten und auf dieser eine E-Mail-Adresse angegeben haben. Diese Bedingungen legen den Schluss nahe, dass die Befragten offener für digitale Medien waren als diejenigen Ärzte, welche sich nicht auf diese digitale Weise präsentierten. Möglicherweise sind auf diese Art eher Ärzte angesprochen worden, die sich für das Thema KI interessierten. Zugleich kann aufgrund der Anonymität der Befragung nicht ausgeschlossen werden, dass die Umfragen nicht auch von nicht-ärztlichem Personal ausgefüllt worden sind.

Zum anderen wird die Aussagekraft der Patientenbefragung dadurch eingeschränkt, dass diese auf einem monozentrisch erhobenen Datensatz zurückzuführen ist. Möglicherweise unterscheidet sich das Patientenklientel an der Universitätsmedizin Mainz von dem der durchschnittlichen Regelversorgung und es ist nicht ausgeschlossen, dass einige Patienten über verschiedene Studienprogramme schon Kontakt zu KI im Bereich der Dermatologie hatten. Dennoch ist an dieser Stelle hervorzuheben, dass durch die Befragung stationärer sowie ambulanter Patienten eine maximale Diversität des zu erreichenden Patientenklientels angestrebt worden ist.

Im Hinblick auf die Aufklärung zu dem Thema KI kann davon ausgegangen werden, dass es sowohl in der Befragung der Patienten als auch in der Befragung der Ärzte zu einer Wirkabschwächung gekommen ist. In beiden Fällen bestand die Möglichkeit, diese zu übergehen und es ist zu erwarten, dass einige der Befragten diesen Text aufgrund mangelnder Zeit oder fehlenden Interesses bewusst ausgelassen haben.

Auch einzelne Fragen sind im Hinblick auf ihre Aussagekraft limitiert. Dies betrifft einerseits die Frage PF8 („Welches Gerät nutzen Sie zeitlich gesehen am häufigsten in Ihrer Freizeit?“). An dieser Stelle ist nicht explizit angegeben worden, dass es sich um eine Einfachauswahl, handelte, sodass ein nicht unerheblicher Anteil der Patienten mehrere Antworten ausgewählt hat. Diese Frage war ebenso Bestandteil der Ärztebefragung (AF12). Aufgrund der Programmierung der Ärztebefragung war eine Mehrfachauswahl an dieser Stelle nicht möglich. Um die Vergleichbarkeit der Befragung zu gewährleisten, sind die Mehrfachantworten von der statistischen Auswertung ausgeschlossen worden.

Andererseits hat die Frage PF18 unerwartet Schwierigkeiten bereitet. In diesem Fall waren 4 Antwortmöglichkeiten vorgegeben, welche priorisiert werden sollten, mit 1 als höchste und 4 als niedrigste Priorität. Die Absicht war, dass jede Ziffer einmal vergeben werden sollte, was in den Testdurchläufen kein Problem darstellte. Bei der Befragung hat jedoch ein Großteil der Befragten Ziffern mehrfach vergeben, sodass diese Antworten von der statistischen Auswertung ausgeschlossen werden mussten. Den Ärzten lag eine ähnliche Frage mit dem gleichen Antwortprinzip vor (AF26). Jedoch ließ, analog zu AF12, die Programmierung keine Mehrfachauswahl der gleichen Ziffer zu.

Auffallend ist auch, dass bei der Frage PF15a, der Frage nach dem selbstständigen Einsatz von KI, einige Patienten angekreuzt haben, auf keinen Fall KI einzusetzen zu wollen, während sie dies bei der Frage PF14 davor nicht getan haben. Ein möglicher Grund dafür könnte sein, dass die Antwortmöglichkeiten bei PF14 nicht sorgfältig gelesen worden sind. Dies wiederum schränkt die Aussagekraft dieser beiden Fragen ein.

Insgesamt waren einige der Patientenfragebögen unvollständig ausgefüllt, sodass die Summe aller Antworten zwischen den verschiedenen Fragen zum Teil deutlich schwankt. In diesem Fall sind die Anteile jeweils in Bezug auf die Anzahl derjenigen, welche die Frage beantwortet haben, berechnet worden und nicht auf das gesamte Befragungskollektiv, wodurch eine geringe Verzerrung der Ergebnisse entstehen kann.

6 Zusammenfassung

In Bezug auf die primären Endpunkte konnte sowohl bei den Ärzten als auch bei den Patienten kein statistisch signifikanter Unterschied zwischen der Stimulus- und der Kontrollgruppe gezeigt werden. Dies ist unter anderem durch das hohe Vorwissen, welche sich vor allem in dem befragten Ärztekollektiv darstellte, zu erklären.

Insgesamt haben 45,5% der Patienten gesagt, dass sie sich selbstständig mittels KI untersuchen würden. Dieser Wert fiel etwas niedriger aus als die in der Literatur angegeben 56%, was jedoch auf das höhere Vorwissen der Patienten in der zum Vergleich herangezogenen Befragung zurückzuführen ist (82).

Noch größer war die Zustimmung der Patienten in Bezug auf den Einsatz von KI in der Klinik. In diesem Fall haben 61,2% der Patienten angegeben, diesen zu unterstützen.

Des Weiteren konnte gezeigt werden, dass die Aufklärung der Patienten die Angst durch den unterstützenden Einsatz von KI durch den Arzt reduziert ($p < 0,05$), im Falle eines Konfliktes würden jedoch mehr als zwei Drittel der Patienten eher dem Arzt als einer KI vertrauen.

Zwei Drittel der Ärzte gaben an, dass sie KI für die Hautkrebsdiagnostik unterstützend einsetzen würden. Unter Berücksichtigung des Vorwissens der Ärzte hat sich gezeigt, dass bei hohem Vorwissen die Bereitschaft zu dem Einsatz von KI größer ist. In diesem Fall haben 80,9% der befragten Dermatologen angegeben, KI als Assistenzsystem einsetzen zu wollen. Unter den Hausärzten betrug die Zustimmung 100%. Jedoch ist aufgrund der geringen Fallzahl in der Subgruppe die Aussagekraft dieses Wertes unter Vorbehalt zu beurteilen. Eine Befragung unter Hausärzten in Frankreich hat eine geringfügig niedrigere Zustimmung von 86% ergeben (88).

Mit 56,6% gab etwa die Hälfte der Ärzte an, der Ansicht zu sein, dass das Wissen um die Funktionsweise von KI als Voraussetzung für deren Einsatz zu betrachten ist. Bei größerem Vorwissen lag dieser Anteil bei zwei Dritteln. Schließlich haben 83,3% der Ärzte die These unterstützt, dass KI das Potential besitzt, die Hautkrebsdiagnostik zu verbessern. In der Literatur lässt sich ein vergleichbarer Wert von 77% finden (87).

Insgesamt waren sich die befragten Patienten und Ärzte mehrheitlich einig, dass nach der selbstständigen KI gestützten Untersuchung durch den Patienten ein Arzt aufgesucht werden sollte, unabhängig davon, ob die KI die untersuchte Läsion als gutartig oder bösartig klassifiziert. Diese Tatsache sollte vor einem großflächigen Einsatz von KI durch Patienten bedacht werden, um entsprechende Maßnahmen zu ergreifen, eine Flut unnötiger Konsultationen zu verhindern.

Entsprechend gering ist das Vertrauen der Ärzte und der Patienten in die alleinige Anwendung einer KI ausgefallen. Knapp zwei Drittel der Patienten und weit über 90% der Ärzte haben gesagt, wenig Vertrauen in die KI allein zu haben. Lediglich 7,8% der Patienten haben Vertrauen in die KI als stand-alone System angegeben. Mit 12% findet sich ein vergleichbarer Wert in der Literatur (84).

Außerdem hat sich gezeigt, dass etwa ein Drittel der Patienten auf den Einsatz von KI beim Arzt bestehen würden, während ca. 50% der Ärzte angegeben haben, dem Wunsch nach dem Einsatz von KI nachkommen zu würden. Auch diese Erkenntnis zeigt, dass der Einsatz von KI überlegt sein sollte und Potential für Konflikte in der Arzt-Patienten-Interaktion bietet.

Schließlich haben sowohl die Ärzte als auch die Patienten nach Wahlmöglichkeit angegeben, dass ihnen die Diagnosesicherheit die wichtigste Eigenschaft einer KI zur Hautkrebsdiagnostik sei. Dass die KI Verbesserungspotential in Bezug auf ihre Robustheit besitzt, ist bekannt (79, 90) und es gibt es bereits verschiedene Ansätze, diese Robustheit weiter zu fördern und die Diagnosesicherheit zu verbessern (77, 78, 90). Um weitere, verlässliche Aussagen in Bezug auf die Robustheit von KI treffen zu können, sollten diese im realen Umfeld validiert werden.

7 Literaturverzeichnis

1. Photographer U. National Cancer Institute Visuals Online Website. Melanoma [Internet]. 08.09.2011 [zitiert am 24.10.2022]. URL: <https://visualsonline.cancer.gov/details.cfm?imageid=9186>.
2. Lutterotti, N. FAZ.net, Eins zu Null für Dr. Computer [Internet]. 30.05.2018 [zitiert am 04.04.2022]. URL: <https://www.faz.net/aktuell/wissen/medizin-ernaehrung/dermatologen-uebertrumpft-eins-zu-null-fuer-den-computer-15614845.html>.
3. Maier, M. Berliner Zeitung, KI in der Medizin: Algorithmus erkennt Hautkrebs präziser als Dermatologen [Internet]. 17.04.2019 [zitiert am 04.04.2022]. URL: <https://www.berliner-zeitung.de/zukunft-technologie/ki-in-der-medizin-algorithmus-erkennt-hautkrebs-praeziser-als-dermatologen-li.74794?pid=true>.
4. Hillienhof, A. KI bei der medizinischen Bildanalyse – Expertenblick weiter unverzichtbar [Internet]. 18.02.2021, [zitiert am 04.04.2022]. URL: <https://www.aerzteblatt.de/nachrichten/120725/KI-bei-der-medizinischen-Bildanalyse-Expertenblick-weiter-unverzichtbar>.
5. Apt W, Botthof A, Bovenschulte M, Büching C, Dwertmann A, Ferdinand J-P, et al. Künstliche Intelligenz Technologie | Anwendung | Gesellschaft [E-Book] Berlin Springer-Verlag GmbH; 2019. Available from: DOI 10.1007/978-3-662-58042-4.
6. RKI. Krebs in Deutschland für 2017/2018 [Internet] 13. Auflage. 2021 [zitiert am 05.04.2022]. URL: https://www.krebsdaten.de/Krebs/DE/Content/Publikationen/Krebs_in_Deutschland/kid_2021/krebs_in_deutschland_2021.pdf?__blob=publicationFile.
7. Hillienhof A, Bundesärztekammer. Hillienhof. Dermatologen warnen vor Nachwuchsmangel [Internet]. 07.02.2022 [zitiert am 03.04.2022]. URL: <https://www.aerzteblatt.de/nachrichten/131611/Dermatologen-warnen-vor-Nachwuchsmangel>.
8. Turing, A. M. Computing Machinery and Intelligence [Internet]. 1950 [zitiert am 12.04.2022]. URL: <https://courses.cs.umbc.edu/471/papers/turing.pdf>.
9. Russell SJ, Norvig P. Artificial Intelligence A Modern Approach [E-Book], 3. Auflage. New Jersey: Prentice Hall; 2010. Available from: <https://people.engr.tamu.edu/guni/csce625/slides/AI.pdf>.
10. Lämmel U, Cleve J. Künstliche Intelligenz : Wissensverarbeitung - neuronale Netze [E-Book]. München: Carl Hanser Verlag; 2020. Available from: DOI 10.3139/9783446463639.
11. Searle, J. Minds, Brains and Programs, Behavioral and Brain Sciences 3 (3): 417-457 [Internet]. 14.07.2003 [zitiert am 12.04.2022]. URL: <https://web-archive.southampton.ac.uk/cogprints.org/7150/1/10.1.1.83.5248.pdf1980>.
12. Huss R. Künstliche Intelligenz, Robotik und Big Data in der Medizin [E-Book]. Berlin: Springer-Verlag GmbH 2019. Available from: DOI 10.1007/978-3-662-58151-3.
13. Mainzer K. Künstliche Intelligenz – Wann übernehmen die Maschinen? [E-Book]. Heidelberg, Berlin: Springer-Verlag; 2016. Available from: DOI 10.1007/978-3-662-48453-1.
14. Wennker P. Künstliche Intelligenz in der Praxis [E-Book]. Wiesbaden: Springer Gabler, Springer Fachmedien Wiesbaden GmbH; 2020. Available from: DOI 10.1007/978-3-658-30480-5.
15. Mitchell TM. Machine Learning [E-Book]: McGraw-Hill Science/Engineering/Math; 1997. Available from: <https://www.cs.cmu.edu/~tom/files/MachineLearningTomMitchell.pdf>.
16. Hady MFA, Schwenker F. Combining Committee-Based Semi-Supervised Learning and Active Learning. Journal of Computer Science and Technology. 06.2010;25:681-98.
17. Brauner, M. Was bedeutet Data Labeling für das Maschinelle Lernen? [Internet]. [zitiert am 09.12.2022]. URL: <https://datalabeling.de/wissenswertes-zum-data-labeling/>.
18. Ärzteblatt. Stellungnahme der Zentralen Kommission zur Wahrung ethischer Grundsätze in der Medizin und ihren Grenzgebieten (Zentrale Ethikkommission) bei der Bundesärztekammer "Entscheidungsunterstützung ärztlicher Tätigkeit durch Künstliche Intelligenz" [Internet]. 23.08.2021 [zitiert am 04.04.2022]. URL: <https://www.zentrale-ethik.de/Dateien/2021/08/23/20210823-Stellungnahme-Zentrale-Ethikkommission-zur-Wahrung-ethischer-Grundsätze-in-der-Medizin-und-ihren-Grenzgebieten-bei-der-Bundesärztekammer-Entscheidungsunterstützung-ärztlicher-Tätigkeit-durch-Künstliche-Intelligenz.pdf>.

ethikkommission.de/fileadmin/user_upload/_old-files/downloads/pdf-Ordner/Zeko/ZEKO_SN_CDSS_Online_final.pdf.

19. Bundesärztekammer. (Muster-)Berufsordnung für die in Deutschland tätigen Ärztinnen und Ärzte, Bundesärztekammer, Deutsches Ärzteblatt [Internet]. 15.06.2018 [zitiert am 09.04.2022]. URL: https://www.uni-ulm.de/fileadmin/website_uni_ulm/ethkom/MBO_BAEK_neu__2018-06-15_.pdf.
20. Verletzung von Privatgeheimnissen, § 203 Absatz 3 Satz 2 StGB (01.08.2022).
21. Jorzig A, Sarangi F. Digitalisierung im Gesundheitswesen : Ein kompakter Streifzug durch Recht, Technik und Ethik [E-Book]. Berlin: Springer-Verlag GmbH; 2020. Available from: doi.org/10.1007/978-3-662-58306-7.
22. Aufklärungspflichten, § 630e Absatz 1 BGB (26.02.2013).
23. Körperverletzung, § 223 Absatz 1 StGB (16.8.2023).
24. Förderung der digitalen Gesundheitskompetenz, § 20k Absatz 1 SGB V (19.12.2019).
25. BFARM. Für DiGA-Nutzende [Internet]. [zitiert am 11.12.2022]. URL: <https://diga.bfarm.de/de/diga-nutzende>.
26. BFARM. Bundesinstitut für Arzneimittel und Medizinprodukte, Das DiGA-Verzeichnis [Internet]. [zitiert am 11.12.2022]. URL: <https://diga.bfarm.de/de/verzeichnis?category=%5B%2274%22%5D>.
27. Grundsätze für die Verarbeitung personenbezogener Daten, Artikel 5 Absatz 1 EU-DSGVO (25.05.2018).
28. Widerspruchsrecht, Artikel 21 Absatz 1 EU-DSGVO (25.05.2018).
29. Recht auf Löschung ("Recht auf Vergessenwerden"), Artikel 17 Absatz 1 EU-DSGVO (25.05.2018).
30. Widerspruchsrecht, Artikel 21 Absatz 3 d) EU-DSGVO (25.05.2018).
31. Zweck des Gesetzes; Anwendungsbereich, § 1 Absatz 1 Gesetz zur Nutzung von Gesundheitsdaten zu gemeinwohlorientierten Forschungszwecken und zur datenbasierten Weiterentwicklung des Gesundheitswesens (Gesundheitsdatennutzungsgesetz – GDNG) (22.3.2024).
32. BMBF. Konsultation zum Forschungsdatengesetz gestartet [Internet]. 06.03.2023 [zitiert am 22.11.2023] URL: <https://www.bmbf.de/bmbf/shareddocs/kurzmeldungen/de/2023/03/230306-forschungsdatengesetz.html>.
33. Haftung, § 1 Absatz 1 ProdHaftG (17.7.2017).
34. Haftung, § 1 Absatz 2 Nr. 2 ProdHaftG (17.07.2017).
35. EUR-Lex. Amtsblatt der Europäischen Union, Regelung der zivilrechtlichen Haftung beim Einsatz künstlicher Intelligenz [Internet]. 06.10.2021 [zitiert am 18.11.2022]. URL: <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A52020IP0276&qid=1645340747377>.
36. Haftung für den Verrichtungsgehilfen, § 831 Absatz 1 BGB (14.3.2023).
37. Verordnung 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz, Präambel (179) (13.06.2024).
38. Verordnung 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz, Präambel (26) (13.06.2024).
39. Verordnung 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz, Präambel (27) (13.06.2024).
40. Verordnung 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz, Präambel (50) (13.06.2024).
41. Bundesministerium für Wirtschaft und Klimaschutz. Konformitätsbewertung [Internet]. [zitiert am 24.06.2023]. URL: <https://www.bmwk.de/Redaktion/DE/Artikel/Technologie/konformitaetsbewertung.html#:~:text=Konformit%C3%A4tsbewertungen%20sind%20T%C3%A4tigkeiten%20wie%20das,beispielweise%20der%20Schadstofffreiheit%20von%20Trinkwasser>.

42. Verordnung 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz, Präambel (65) (13.06.2024).
43. Verordnung 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz, Präambel (67) (13.06.2024).
44. Verordnung des Europäischen Parlaments und des Rates vom 5. April 2017 über Medizinprodukte, Anhang VIII, Kapitel III, Regel 11 (05.4.2017).
45. Dabrock P, Amunts K, Kruse A, Wiesemann C, Angerer C, Augsburg S, et al. Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung [E-Book]. Berlin: Deutsche Ethikrat; 2017. Available from: <https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-big-data-und-gesundheit.pdf>.
46. Ala-Pietilä P, Lucas P, Bauer W, Martinkenaite I, Bergmann U, Metzinger T, et al. Ethikleitlinien für eine vertrauenswürdige KI [E-Book]. Brüssel: Europäische Kommission; 2019. Available from: <https://op.europa.eu/de/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1>.
47. Recht auf Achtung des Privat- und Familienlebens, Artikel 8 Absatz 1 Europäische Menschenrechtskonvention (01.08.2021).
48. Artikel 2 Absatz 1 GG (19.12.2022).
49. Gegenstand und Ziele, Artikel 1 Absatz 2 DSGVO (25.05.2018).
50. Ploug T, Holm S. Meta consent: a flexible and autonomous way of obtaining informed consent for secondary research. *Bmj*. 2015;350:h2146.
51. Rechtmäßigkeit der Verarbeitung, Artikel 6 Absatz 1 EU-DSGVO (25.05.2018).
52. Udrea A, Mitra GD, Costea D, Noels EC, Wakkee M, Siegel DM, et al. Accuracy of a smartphone application for triage of skin lesions based on machine learning algorithms. *J Eur Acad Dermatol Venereol*. 2020;34(3):648-655.
53. Pangti R, Mathur J, Chouhan V, Kumar S, Rajput L, Shah S, et al. A machine learning-based, decision support, mobile phone application for diagnosis of common dermatological diseases. *J Eur Acad Dermatol Venereol*. 2021;35(2):536-545.
54. Huang HW, Hsu BW, Lee CH, Tseng VS. Development of a light-weight deep learning model for cloud applications and remote diagnosis of skin cancers. *J Dermatol*. 2020;48:310-316.
55. Tenzer, F. Anzahl der Smartphone-Nutzer weltweit von 2016 bis 2020 und Prognose bis 2024 [Internet]. 24.01.2022 [zitiert am 20.07.2022]. URL: <https://de.statista.com/statistik/daten/studie/309656/umfrage/prognose-zur-anzahl-der-smartphone-nutzer-weltweit/>.
56. Tenzer, F. Anteil der Smartphone-Nutzer* in Deutschland in den Jahren 2012 bis 2021 [Internet]. 28.11.2022 [zitiert am 20.02.2023]. URL: <https://de.statista.com/statistik/daten/studie/585883/umfrage/anteil-der-smartphone-nutzer-in-deutschland/#:~:text=Der%20Anteil%20der%20Smartphone%2DNutzer,oder%20Handy%20im%20Haushalt%20besitzen.&text=Fast%20jeder%20Deutsche%2C%20der%20unter,alt%20ist%2C%20nutzt%20ein%20Smartphone>.
57. Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of Oncology*. 2018;29:1836-1842.
58. Marchetti MA, Codella NCF, Dusza SW, Gutman DA, Helba B, Kalloo A, et al. Results of the 2016 International Skin Imaging Collaboration International Symposium on Biomedical Imaging challenge: Comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images. *J Am Acad Dermatol*. 2018;78:270-277.e1.
59. Brinker TJ, Hekler A, Enk AH, Klode J, Hauschild A, Berking C, et al. A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task. *Eur J Cancer*. 2019;111:148-154.

60. Brinker TJ, Hekler A, Enk AH, Klode J, Hauschild A, Berking C, et al. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *Eur J Cancer*. 2019;113:47-54.
61. Brinker TJ, Hekler A, Enk AH, Berking C, Haferkamp S, Hauschild A, et al. Deep neural networks are superior to dermatologists in melanoma image classification. *Eur J Cancer*. 2019;119:11-17.
62. RKI. Zentrum für Krebsregisterdaten online, Malignes Melanom der Haut [Internet]. 30.09.2022 [zitiert am 03.12.2022]. URL: https://www.krebsdaten.de/Krebs/DE/Content/Krebsarten/Melanom/melanom_node.html.
63. RKI. Zentrum für Krebsregisterdaten online, Nicht-melanotischer Hautkrebs [Internet]. 29.11.2021 [zitiert am 03.04.2022]. URL: https://www.krebsdaten.de/Krebs/DE/Content/Krebsarten/Nicht-melanotischer-Hautkrebs/nicht-melanotischer-hautkrebs_node.html.
64. Tschandl P, Rosendahl C, Akay BN, Argenziano G, Blum A, Braun RP, et al. Expert-Level Diagnosis of Nonpigmented Skin Cancer by Combined Convolutional Neural Networks. *JAMA Dermatol*. 2019;155:58-65.
65. Liu Y, Jain A, Eng C, Way DH, Lee K, Bui P, et al. A deep learning system for differential diagnosis of skin diseases. *Nat Med*. 2020;26:900-908.
66. Jinnai S, Yamazaki N, Hirano Y, Sugawara Y, Ohe Y, Hamamoto R. The Development of a Skin Cancer Classification System for Pigmented Skin Lesions Using Deep Learning. *Biomolecules*. 2020.
67. Maron RC, Weichenthal M, Utikal JS, Hekler A, Berking C, Hauschild A, et al. Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks. *Eur J Cancer*. 2019;119:57-65.
68. Sharma AN, Shwe S, Mesinkovska NA. Current state of machine learning for non-melanoma skin cancer. *Arch Dermatol Res*. 2021;314:325-327.
69. Reiter O, Mimouni I, Gdalevich M, Marghoob AA, Levi A, Hodak E, et al. The diagnostic accuracy of dermoscopy for basal cell carcinoma: A systematic review and meta-analysis. *J Am Acad Dermatol*. 2019;80:1380-1388.
70. Rosendahl C, Cameron A, Argenziano G, Zalaudek I, Tschandl P, Kittler H. Dermoscopy of squamous cell carcinoma and keratoacanthoma. *Arch Dermatol*. 2012;148:1386-1392.
71. Hekler A, Utikal JS, Enk AH, Hauschild A, Weichenthal M, Maron RC, et al. Superior skin cancer classification by the combination of human and artificial intelligence. *Eur J Cancer*. 2019;120:114-121.
72. Tschandl P, Codella N, Akay BN, Argenziano G, Braun RP, Cabo H, et al. Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. *Lancet Oncol*. 2019;20:938-947.
73. Han SS, Park I, Eun Chang S, Lim W, Kim MS, Park GH, et al. Augmented Intelligence Dermatology: Deep Neural Networks Empower Medical Professionals in Diagnosing Skin Cancer and Predicting Treatment Options for 134 Skin Disorders. *J Invest Dermatol*. 2020;140:1753-1761.
74. Deeks JJ, Dinnes J, Williams HC. Sensitivity and specificity of SkinVision are likely to have been overestimated. *J Eur Acad Dermatol Venereol*. 2020;34:582-583.
75. Chuchu N, Takwoingi Y, Dinnes J, Matin RN, Bassett O, Moreau JF, et al. Smartphone applications for triaging adults with skin lesions that are suspicious for melanoma. *Cochrane Database Syst Rev*. 2018;12.
76. Han SS, Kim MS, Lim W, Park GH, Park I, Chang SE. Classification of the Clinical Images for Benign and Malignant Cutaneous Tumors Using a Deep Learning Algorithm. *J Invest Dermatol*. 2018;138:1529-1538.
77. Höhn J, Hekler A, Krieghoff-Henning E, Kather JN, Utikal JS, Meier F, et al. Integrating Patient Data Into Skin Cancer Classification Using Convolutional Neural Networks: Systematic Review. *J Med Internet Res*. 2021;23:e20708.

78. Lucius M, De All J, De All JA, Belvisi M, Radizza L, Lanfranconi M, et al. Deep Neural Frameworks Improve the Accuracy of General Practitioners in the Classification of Pigmented Skin Lesions. *Diagnostics (Basel)*. 2020;10.
79. Young AT, Fernandez K, Pfau J, Reddy R, Cao NA, von Franque MY, et al. Stress testing reveals gaps in clinic readiness of image-based diagnostic artificial intelligence models. *NPJ Digit Med*. 2021;4:10.
80. Foto Finder Systems GmbH. Harmloser Naevus oder gefährliches Melanom? [Internet]. 18.04.2023 [zitiert am 24.10.2023]. URL: <https://www.fotofinder.de/unternehmen/aktuelles/harmloser-naevus-oder-gefaehrliches-melanom>.
81. Health N. Neko Health Launches Website, Neko Health launches a first-of-its-kind full-body scanner to help doctors find and prevent disease [Internet], Aktualisierungsdatum: 02.02.2023, Zitiert am: 22.04.2023, URL: <https://www.nekohealth.com/en/press-release-launch>.
82. Jutzi TB, Krieghoff-Henning EI, Holland-Letz T, Utikal JS, Hauschild A, Schadendorf D, et al. Artificial Intelligence in Skin Cancer Diagnostics: The Patients' Perspective. *Front Med (Lausanne)*. 2020;7:233.
83. Nelson CA, Pérez-Chada LM, Creadore A, Li SJ, Lo K, Manjaly P, et al. Patient Perspectives on the Use of Artificial Intelligence for Skin Cancer Screening: A Qualitative Study. *JAMA Dermatol*. 2020;156:501-512.
84. Lim K, Neal-Smith G, Mitchell C, Xerri J, Chuanromanee P. Perceptions of the use of artificial intelligence in the diagnosis of skin cancer: an outpatient survey. *Clin Exp Dermatol*. 2022;47:542-546.
85. Soellner M, Koenigstorfer J. Compliance with medical recommendations depending on the use of artificial intelligence as a diagnostic method. *BMC medical informatics and decision making*. 2021;21:236.
86. Oh S, Kim JH, Choi SW, Lee HJ, Hong J, Kwon SH. Physician Confidence in Artificial Intelligence: An Online Mobile Survey. *J Med Internet Res*. 2019;21:e12422.
87. Polesie S, Gillstedt M, Kittler H, Lallas A, Tschandl P, Zalaudek I, et al. Attitudes towards artificial intelligence within dermatology: an international online survey. *Br J Dermatol*. 2020;183:159-161.
88. Samaran R, L'Orphelin JM, Dreno B, Rat C, Dompmmartin A. Interest in artificial intelligence for the diagnosis of non-melanoma skin cancer: a survey among French general practitioners. *Eur J Dermatol*. 2021;31:457-462.
89. Leiner, D., Leiner, S. SoSci Survey (Version 3.2.55), SoSci Survey GmbH [Internet], [zitiert am 04.04.2022]. URL: <https://www.soscisurvey.de/de/index>.
90. Maron RC, Haggemüller S, von Kalle C, Utikal JS, Meier F, Gellrich FF, et al. Robustness of convolutional neural networks in recognition of pigmented skin lesions. *Eur J Cancer*. 2021;145:81-91.
91. Grätzel von Grätz, P. Die MRT wird bald 50 – was bringt die Zukunft? [Internet]. 03.12.2020 [zitiert am 16.11.2022]. URL: <https://www.siemens-healthineers.com/deu/perspectives/mso-the-MRI-at-50#:~:text=Trotz%20ihrer%20St%C3%A4rken%20ist%20die,%C3%A4nger%20als%20%C3%BCr%20eine%20CT>.
92. Ferrante di Ruffano L, Takwoingi Y, Dinnes J, Chuchu N, Bayliss SE, Davenport C, et al. Computer-assisted diagnosis techniques (dermoscopy and spectroscopy-based) for diagnosing skin cancer in adults. *Cochrane Database of Systematic Reviews*. 2018:Cd013186.

8 Anhang

8.1 Aushang für die Patientenbefragung



UNIVERSITÄTSmedizin.
MAINZ

Hautklinik und Poliklinik

Patientenbefragung

Sehr geehrte Damen und Herren,

mein Name ist Damian Klaedtke und ich bin Medizinstudent an der Universitätsklinik Mainz. Im Rahmen meiner Doktorarbeit beschäftige ich mich mit dem **Einsatz künstlicher Intelligenz im Bereich der Hautkrebsdiagnostik**. Für diese Arbeit werden sowohl Patienten als auch Ärzte zu dieser Thematik befragt.

Die Datenerhebung erfolgt im Zeitraum vom 16.08.2021 bis zum 16.12.2021, ist einmalig, anonym und dauert etwa 10-15 Minuten.

Um an dieser Befragung teilzunehmen, scannen Sie einfach den nachfolgenden QR-Code mit ihrem Smartphone oder Tablet.

Vielen Dank für Ihre Teilnahme.



8.2 Aufklärungstext

Künstliche Intelligenz in der Medizin

Hautkrebsdiagnostik

Was versteht man unter einer künstlichen Intelligenz und wie kann diese Einfluss auf die Hautkrebsdiagnostik nehmen?

Hinweis: Bitte lesen Sie sich den folgenden Text aufmerksam durch.

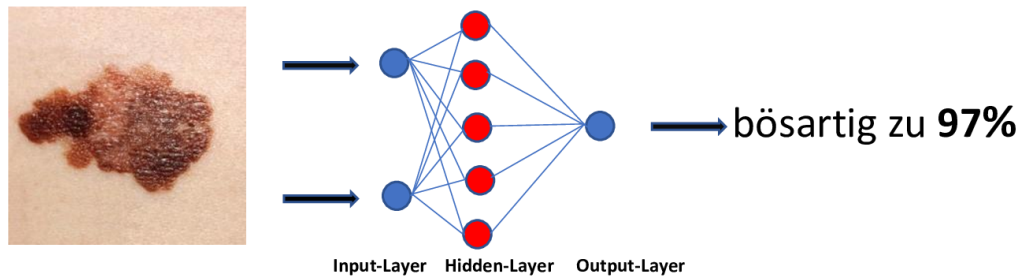
Der Begriff „künstliche Intelligenz“ (im weiteren Verlauf als „**KI**“ abgekürzt) lässt sich am einfachsten als ein Teilgebiet der **Computerwissenschaft** verstehen, welches sich mit der technischen Imitation menschlicher Verhaltens- und Denkweisen befasst. Bisher ist es noch nicht gelungen, ein vollständig autonom agierendes Computersystem zu programmieren. Das gehört noch in den Bereich der science fiction.

Vollkommen real hingegen und bereits in die Funktionsweisen vieler technischer Applikationen integriert, ist die „**schwache**“ **KI**, auf die im weiteren Text näher eingegangen wird.

Diese beschäftigt sich mit konkreten logischen Problemstellungen und zeichnet sich vor allem dadurch aus, dass sie erst auf einen **externen Impuls** hin aktiv wird. Sie arbeitet also weder autonom, noch kann man ihr ein Bewusstsein zusprechen. Der Unterschied zu einem „herkömmlichen“ Programm besteht darin, dass die „schwache“ KI die Fähigkeit besitzt, sich fortlaufend selbst zu optimieren, sie ist also **lernfähig**. Das KI-Programm verarbeitet dazu die ihm angebotenen Daten (INPUT) und lernt, daraus bestimmte Ergebnisse (OUTPUT) zu generieren. Im Zentrum der Methodik steht dabei das **maschinelle Lernen**. Das aktuell gängigste Verfahren ist die Verwendung von neuronalen Netzen (kurz CNNs, aus dem Englischen für Convolutional Neural Nets), die nun kurz beschrieben werden.

Ein **Neuronales Netzwerk** ähnelt in seinem Aufbau grundsätzlich dem **biologischen Nervensystem**. Es ist aus einer Vielzahl kleiner Recheneinheiten, den **Neuronen** aufgebaut. In seiner primitivsten Struktur besteht es aus wenigstens drei Schichten, sogenannten Layern. Die Neuronen der ersten Schicht, dem Input-Layer, nehmen die zu verarbeitenden Informationen auf, geben diese an die Hidden-Layer Schicht weiter, welche die Informationen schließlich an den Output-Layer weiterreicht (siehe Abbildung).

Ähnlich dem Ablauf in einem biologischen Nervensystem durchlaufen die Informationen verschiedene Pfade durch das Netzwerk der Neuronen. Das zunächst naive Netzwerk muss ein Training durchlaufen, um die Beurteilung der Daten zu lernen. Ein gängiges Prinzip ist das „**überwachte Lernen**“. Dabei bekommt das Programm eine Information, wie zum Beispiel ein Bild, gemeinsam mit einem bereits bekannten Ergebnis (Diagnose, Bezeichnung, Unterscheidung gutartig / bösartig, etc.) präsentiert. Stark vereinfacht ausgedrückt „lernt“ das neuronale Netzwerk **Muster** in den angebotenen Daten zu erkennen und damit das gewünschte Ergebnis zu berechnen. Im Fall der Hautkrebsdiagnostik bedeutet das zum Beispiel, anhand eines Bildes einer Hautveränderung zu sagen, ob es sich um eine gutartige oder eine bösartige Veränderung handelt. Als Ergebnis (OUTPUT) erhält der Nutzer einen Wert, der die Wahrscheinlichkeit oder Diagnosesicherheit für den ausgegebenen Wert darstellt.



Komplexere Systeme können mehrere Krankheiten oder Symptome erkennen und diese in einer Wahrscheinlichkeitsverteilung angeben, etwa in Form eines abgestuften Rankings der fünf wahrscheinlichsten Krankheiten.

In zahlreichen Studien konnte bereits überzeugend dargestellt werden, dass KI-gestützte Programme Experten auf dem Fachgebiet der Hautkrebsdiagnostik mindestens ebenbürtig sind. In einigen Studien haben sich diese Programme sogar als überlegen dargestellt.

Allerdings stellen sich dem alltäglichen Einsatz von KIs noch einige Herausforderungen entgegen. Ein wesentlicher Punkt dabei ist das fehlende Trainingsmaterial. Dies hat mehrere Gründe: Die Hautveränderungen, welche zum Training der KIs verwendet werden, müssen in ihrer Diagnose zweifelsfrei bestätigt werden. Dazu wird das Gewebe durch ein Mikroskop histologisch begutachtet, was aufwändig ist und längst nicht auf alle Bildaufnahmen von Hautveränderungen zutrifft. Zudem unterliegt des Trainingsmaterial einer Selektion. Von Seltenen Tumoren und seltenen Lokalisationen gibt es nur wenige Aufnahmen. Das hat zur Folge, dass diese von KIs schlechter erkannt werden. Außerdem gibt es weitere Einflussfaktoren, wie zum Beispiel den Hauttyp. Die meisten Bildaufnahmen, welche aktuell zum Training der KIs zur Verfügung stehen, wurden von hellen Hauttypen gemacht. Wenn die KIs anschließend Aufnahmen von anderen Hauttypen analysieren sollten, schnitten diese deutlich schlechter ab.

Dennoch hat man bereits zeigen können, dass die Kombination aus KI und Dermatologe dem Dermatologen allein überlegen ist. Besonders deutlich ist dieses Ergebnis bei Ärzten mit geringer Erfahrung im Bereich der Hautkrebsdiagnostik.

Als ein noch junges Gebiet im Bereich der Hautkrebsdiagnostik, unterliegt die Entwicklung und der Einsatz der KIs einer wachsenden Dynamik und es bleibt offen, welche Auswirkungen sich daraus auf die Hautkrebsdiagnostik von morgen ergeben.

8.3 Fragebogen: Patienten



Hautklinik und Poliklinik

Sehr geehrte Damen und Herren,

mein Name ist Damian Klaedtke und ich bin Medizinstudent an der Universitätsklinik Mainz. Im Rahmen meiner Doktorarbeit beschäftige ich mich mit dem Einsatz künstlicher Intelligenz im Bereich der Hautkrebsdiagnostik. Für diese Arbeit werden sowohl Patienten als auch Ärzte zu dieser Thematik befragt. Die Umfrage ist einmalig, anonym und dauert etwa 5-15 Minuten.

Hinweis: Um den Lesefluss zu erleichtern und möglichen Verwechslungen vorzubeugen, ist diese Befragung nicht gegendert. Dies hat unter keinen Umständen einen abwertenden Hintergrund. Ich danke Ihnen für Ihr Verständnis.

Bevor Sie beginnen:

Bitte achten Sie darauf, die Reihenfolge der Blätter und der Fragen einzuhalten.

Dies ist für die Auswertung außerordentlich wichtig.

1. Bitte geben Sie Ihr Geschlecht an:

- männlich
 - weiblich
 - divers
-

2. Bitte geben Sie Ihr Alter an:

_____ (in Jahren)

3. Bitte geben Sie Ihren höchsten Bildungsstand an:

- Schüler
 - Hauptschulabschluss
 - Realschulabschluss
 - Abitur/Fachabitur
 - Ausbildung
 - abgeschlossenes Studium
 - Sonstiges
-

4. Haben Sie schon einmal etwas von „künstlicher Intelligenz“ (KI) gehört?

- ja
 - nein
-

Im Folgenden wird der Ausdruck „künstliche Intelligenz“ durch **KI** ersetzt!

5. Inwieweit würden Sie von Sich selbst behaupten, zu wissen, wie eine KI funktioniert?

(keine Ahnung) (sehr gutes
Verständnis)

6. Benutzen Sie in Ihrem Alltag KI?

- ja
- nein
- das weiß ich nicht

7. Besitzen Sie ein Smartphone oder ein anderes internetfähiges Endgerät?

- ja
 - nein
-

Hinweis: Wenn Sie diese Frage mit „**nein**“ beantwortet haben, fahren Sie bitte mit **Frage 11** fort.

8. Welches Gerät nutzen Sie zeitlich gesehen am häufigsten in Ihrer Freizeit?

- Smartphone
 - Tablet
 - Computer
 - Sonstiges
-

9. Wenn Sie sich über Krankheitssymptome informieren: Wie oft greifen Sie dabei auf das zuvor ausgewählte Medium zurück?

- (nie) (immer)
-

Hinweis: Wenn Sie diese Frage mit „**nie**“ beantwortet haben, fahren Sie bitte mit **Frage 11** fort.

10. Worauf haben Sie dabei zurückgegriffen? (**Mehrfachauswahl**)

- Internetsuchmaschinen (z.B. Google)
 - spezielle Diagnoseprogramme für den Computer
 - Gesundheits-Apps
 - Sonstiges
-

11. Haben oder hatten Sie eine Form von Hautkrebs?

- ja
 - nein
-

Hinweis: Wenn Sie diese Frage mit „**nein**“ beantwortet haben, fahren Sie bitte mit **Frage 14** fort.

12. Wie sind Sie auf den Hautkrebs aufmerksam geworden?

- selbständig
- durch Bekannte/Verwandte
- durch meinen Hausarzt
- durch meinen Hautarzt
- durch einen anderen Facharzt
- Sonstiges

13. Welchen Hautkrebs hatten/haben Sie? (**Mehrfachauswahl möglich**)

- Basalzellkarzinom (Basaliom)

- Plattenepithelkarzinom (Spinaliom)
 - Malignes Melanom
 - ich kann mich nicht erinnern
 - Freitext: _____
-

14. In welcher Form würden Sie eine KI zur eigenständigen Hautkrebsdiagnostik bevorzugt nutzen?

- App für Smartphone/Tablet
 - Computerprogramm
 - selbstständiges Endgerät nur für Hautkrebsdiagnostik
 - ich würde unter keinen Umständen eine KI verwenden
-

Hinweis: Wenn Sie keine KI verwenden würden, fahren Sie bitte mit **Frage 16** fort.

Versuchen Sie sich in folgende hypothetische Situation hineinzusetzen:

15. Angenommen, Sie entdecken eine auffällige Hautveränderung an Ihrem Körper:

- a. Würden Sie sich zunächst eigenständig mittels eines KI-gestützten Programms untersuchen?

(auf keinen Fall) (auf jeden Fall)

Hinweis: Wenn Sie diese Frage mit „auf keinen Fall“ beantwortet haben (das Kreuz ganz links gesetzt haben), fahren Sie bitte mit **Frage 16** fort.

- b. Die KI diagnostiziert die Hautveränderung als Krebs: Würden Sie in diesem Fall Ihren Arzt aufsuchen?

(auf keinen Fall) (auf jeden Fall)

- c. Die KI diagnostiziert die Hautveränderung als gutartig: Würden Sie in diesem Fall Ihren Arzt aufsuchen?

(auf keinen Fall) (auf jeden Fall)

16. Sie haben vor kurzem eine auffällige Hautveränderung an Ihrem Körper entdeckt und besuchen nun Ihren Arzt (ohne sich zuvor mit Hilfe einer KI untersucht zu haben):

- a. Würde es Sie verunsichern, wenn der Arzt zusätzlich eine KI zur Diagnostik zurate ziehen würde, um Sie zu untersuchen?

(auf keinen Fall) (auf jeden Fall)

- b. Würden Sie drauf bestehen, dass Ihr Arzt zusätzlich eine KI nutzt, um Sie zu untersuchen?

(auf keinen Fall) (auf jeden Fall)

c. Würden Sie auf die Diagnose einer KI vertrauen, ohne von Ihrem Arzt untersucht worden zu sein?

(auf keinen Fall) (auf jeden Fall)

d. Sofern die KI und Ihr Arzt zu unterschiedlichen Ergebnissen gelangen: Was würden Sie tun?

(dem Arzt vertrauen) (der KI vertrauen)

17. Würden Sie den Einsatz einer KI zur Hautkrebsdiagnostik in der Klinik befürworten?

(auf keinen Fall) (auf jeden Fall)

18. Was wäre Ihnen für eine KI im „Hausgebrauch“ wichtig? **(Bitte priorisieren Sie mit 1=höchste Priorität, 4=niedrigste Priorität -> bitte Zahlen in die Kästchen unter die Textfelder einfügen)**

Diagnosesicherheit	Jederzeit Verfügbarkeit	Anschließend schnelle Terminvermittlung zum Arzt	Erstattung durch die Krankenkassen
--------------------	-------------------------	--------------------------------------------------	------------------------------------

19. Würden Sie den eigenständigen Gebrauch der KI Ihren Verwandten/Bekanntem weiterempfehlen?

(auf keinen Fall) (auf jeden Fall)

20. Hätten Sie Bedenken um die Sicherheit Ihrer Daten, die die KI im Rahmen der Diagnostik von Ihnen ansammeln würde?

(auf keinen Fall) (auf jeden Fall)

21. Würden Sie Ihre eigenen Daten (z.B. Bilder von Ihrem Hautkrebs) zur Verfügung stellen, um KIs damit zu trainieren?

(auf keinen Fall) (auf jeden Fall)

22. Eigene Anmerkungen...

Ich bedanke mich herzlich für Ihre Unterstützung.

8.4 Fragebogen: Ärzte



UNIVERSITÄTS**medizin.**
MAINZ

Hautklinik und Poliklinik

Sehr geehrte Damen und Herren,

mein Name ist Damian Klaedtke und ich bin Medizinstudent an der Universitätsklinik Mainz. Im Rahmen meiner Doktorarbeit beschäftige ich mich mit dem Einsatz künstlicher Intelligenz im Bereich der Hautkrebsdiagnostik. Für diese Arbeit werden sowohl Patienten als auch Ärzte zu dieser Thematik befragt.

Diese Datenerhebung erfolgt im Zeitraum vom 16.08.2021 bis zum 16.12.2021. Sie ist einmalig, anonym, dauert etwa 5-15 Minuten und richtet sich an Dermatologen und Hausärzte.

Hinweis: Um den Lesefluss zu erleichtern und möglichen Verwechslungen vorzubeugen, ist diese Befragung nicht gegendert. Dies hat unter keinen Umständen einen abwertenden Hintergrund. Ich danke Ihnen für Ihr Verständnis.

Bevor Sie beginnen:

Bitte achten Sie darauf, die Reihenfolge der Blätter und der Fragen einzuhalten.

Dies ist für die Auswertung außerordentlich wichtig.

23. Bitte geben Sie Ihr Geschlecht an:

- männlich
 - weiblich
 - divers
-

24. Bitte geben Sie Ihr Alter an:

_____ (in Jahren)

25. Bitte geben Sie Ihre Fachrichtung an:

- Dermatologe
 - ambulant
 - stationär
 - beides
 - Hausarzt (Allgemeinmediziner/Internist)
 - andere Fachrichtung
-

26. Bitte geben Sie Ihre Berufserfahrung in Ihrer aktuell ausgeübten Fachrichtung an:

_____ (in Jahren)

27. Haben Sie eine abgeschlossene Facharztausbildung in Ihrer derzeitig ausgeübten Fachrichtung?

- ja
 - nein
-

28. Wie häufig werden Sie in Ihrem Alltag mit der Hautkrebsdiagnostik konfrontiert?

(nie) (täglich)

29. Haben Sie schon einmal etwas von „künstlicher Intelligenz“ (KI) gehört?

- ja

- nein
-

Im Folgenden wird der Ausdruck „künstliche Intelligenz“ durch **KI** ersetzt!

30. Inwieweit würden Sie von Sich selbst behaupten, zu wissen, wie eine KI funktioniert?
Verständnis (keine Ahnung) (sehr gutes

31. Benutzen Sie aktuell eine KI im Rahmen der Hautkrebsdiagnostik?

- ja
 nein
 ich diagnostiziere keinen Hautkrebs
-

32. Würden Sie den generellen Einsatz einer KI im Bereich der Hautkrebsdiagnostik befürworten?

- (auf keinen Fall) (auf jeden Fall)

33. Besitzen Sie ein Smartphone oder ein anderes internetfähiges Endgerät?

- ja
- nein

Hinweis: Wenn Sie diese Frage mit „**nein**“ beantwortet haben, fahren Sie bitte mit **Frage 15** fort.

34. Welches Gerät nutzen Sie zeitlich gesehen am häufigsten in Ihrer Freizeit?

- Smartphone
- Tablet
- Computer
- Sonstiges

35. Wenn Sie sich über Krankheitssymptome informieren: Wie oft greifen Sie dabei auf das zuvor ausgewählte Medium zurück?

- (nie) (immer)
-

Hinweis: Wenn Sie diese Frage mit „**nie**“ beantwortet haben, fahren Sie bitte mit **Frage 15** fort.

36. Worauf haben Sie dabei zurückgegriffen? (**Mehrfachauswahl**)

- Internetsuchmaschinen (z.B. Google)
- spezielle Diagnoseprogramme für den Computer
- Gesundheits-Apps
- Sonstiges

37. Dem Patienten fällt eine für ihn ungewöhnliche Hautveränderung auf:

- a. Würden Sie ihm dazu raten, sich zunächst eigenständig mittels eines KI-gestützten Programms zu untersuchen?

(auf keinen Fall) (auf jeden Fall)

Hinweis: Wenn Sie diese Frage mit „**auf keinen Fall**“ beantwortet haben (das Kreuz ganz links gesetzt haben), fahren Sie bitte mit **Frage 16** fort.

- b. Die KI diagnostiziert die Hautveränderung als bösartig: Würden Sie dem Patienten dazu raten, Sie aufzusuchen?

(auf keinen Fall) (auf jeden Fall)

- c. Die KI diagnostiziert keine bösartige Hautveränderung: Würden Sie dem Patienten dazu raten, Sie aufzusuchen?

(auf keinen Fall) (auf jeden Fall)

38. Würden Sie für Ihre Untersuchungen eine KI als diagnostisches Hilfsmittel nutzen?

(auf keinen Fall) (auf jeden Fall)

Hinweis: Wenn Sie diese Frage mit „auf keinen Fall“ beantwortet haben (das Kreuz ganz links gesetzt haben), fahren Sie bitte mit **Frage 20** fort.

39. Wie würden Sie vorgehen, wenn Sie eine therapeutisch relevante Diagnose fällen, die KI diese jedoch verneint?

- therapieren
 - nicht therapieren
 - Biopsie nehmen
 - andere Optionen wählen
-

40. Wie würden Sie vorgehen, wenn die KI eine therapeutisch relevante Diagnose fällen, Sie diese jedoch ausschließen?

- therapieren
 - nicht therapieren
 - Biopsie nehmen
 - andere Optionen wählen
-

41. Würden Sie allein auf die Diagnose einer KI vertrauen, ohne den Patienten selbst untersucht zu haben?

(auf keinen Fall) (auf jeden Fall)

42. Angenommen, der von Ihnen betreute Patient bestünde darauf, dass Sie für Ihre Diagnostik zusätzlich eine KI einsetzen. Würden Sie diesem Wunsch nachkommen?

(auf keinen Fall) (auf jeden Fall)

43. Würden Sie es als Voraussetzung betrachten, dass ein Arzt die Funktionsweise einer KI verstehen muss, bevor er diese verwenden darf?

(auf keinen Fall) (auf jeden Fall)

44. Sind Sie der Meinung, dass KI ein Teil Ihrer Facharztausbildung sein sollte?

(auf keinen Fall) (auf jeden Fall)

45. Wie schätzen Sie das Potential der KIs im Rahmen der Hautkrebsdiagnostik ein?

- verbessert Versorgung
 - verschlechtert Versorgung
 - hat keinen Einfluss auf die Versorgung
-

46. Würden Sie, sofern Ihre Patienten die Zustimmung geben, die im Rahmen der Diagnostik anfallenden Daten (z.B. Bilder von ihrem Hautkrebs) zur Verfügung stellen, um KIs damit zu trainieren?

(auf keinen Fall) (auf jeden Fall)

47. Welche Eigenschaft einer KI wäre Ihnen am wichtigsten?

- Diagnosesicherheit (hohe Sensitivität)
 - Erkennen möglichst vieler Diagnosen
 - eigener Vorschlag: _____
-

48. Welche Aspekte bei der Anwendung eines KI-gestützten Programms erscheinen Ihnen am wichtigsten?

(Bitte priorisieren Sie von 1=höchste Priorität – 4=niedrigste Priorität -> in die Kästchen)

Leichte Bedienbarkeit	Nachvollziehbarkeit der Ergebnisse	Analysegeschwindigkeit (kein Zeitverlust)	Delegierbarkeit
-----------------------	------------------------------------	-------------------------------------------	-----------------

Priorität:

49. Was glauben Sie, ist die größte Stärke der KIs?

50. Was glauben Sie, ist die größte Schwäche der KIs?

51. Eigene Anmerkungen...

Ich bedanke mich herzlich für Ihre Unterstützung.

8.5 Ergebnisse

		Häufigkeit	Prozent	Gültige	Kumulierte
Gültig	Schüler	1	,9	,9	,9
	Hauptschulabschluss	14	12,8	13,0	13,9
	Realschulabschluss	12	11,0	11,1	25,0
	Abitur/Fachabitur	13	11,9	12,0	37,0
	Ausbildung	30	27,5	27,8	64,8
	abgeschlossenes Studium	30	27,5	27,8	92,6
	Sonstiges	8	7,3	7,4	100,0
	Gesamt	108	99,1	100,0	
Fehlend	System	1	,9		
Gesamt		109	100,0		

Tabelle 4: Angabe zum höchsten Bildungsabschluss der Patienten

		Anzahl	Anzahl in %
KI generell gehört	ja	88	82,2%
	nein	19	17,8%
KI im Alltag	ja	32	30,8%
	nein	38	36,5%
	das weiß ich nicht	34	32,7%

Tabelle 5: Angabe der Patienten, ob sie schonmal etwas von KI gehört haben und ihre Einschätzung, ob sie diese im Alltag einsetzen

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 (keine Ahnung)	30	27,5	28,8	28,8
	2	29	26,6	27,9	56,7
	3	24	22,0	23,1	79,8
	4	16	14,7	15,4	95,2
	5 (sehr gutes Verständnis)	5	4,6	4,8	100,0
	Gesamt	104	95,4	100,0	
Fehlend	nicht beantwortet	1	,9		
	System	4	3,7		
	Gesamt	5	4,6		
Gesamt		109	100,0		

Tabelle 6: Einschätzung der Patienten zum eigenen Wissen um die Funktionsweise von KI

Besitzen Sie ein Smartphone?	ja	Anzahl	100	
		Anzahl in %	93,5%	
	nein	Anzahl	7	
		Anzahl in %	6,5%	
Welches Medium nutzen Sie in ihrer Freizeit am häufigsten?	Smartphone	Anzahl	58	
		Anzahl in %	74,4%	
	Tablet	Anzahl	5	
		Anzahl in %	6,4%	
	Computer	Anzahl	13	
		Anzahl in %	16,7%	
	Sonstiges	Anzahl	2	
		Anzahl in %	2,6%	
	Wie oft greifen Sie bei der Informationssuche über Krankheitssymptome auf das zuvor ausgewählte Medium zurück?	1 (nie)	Anzahl	10
			Anzahl in %	12,8%
2		Anzahl	16	
		Anzahl in %	20,5%	
3		Anzahl	16	
		Anzahl in %	20,5%	
4		Anzahl	21	
		Anzahl in %	26,9%	
5 (immer)		Anzahl	15	
		Anzahl in %	19,2%	

Tabelle 7: Anzahl der Smartphonebesitzer, Häufigkeit der Nutzung des Mediums und Häufigkeit des Einsatzes

		Anzahl	Anzahl in %
Internetsuchmaschinen (z.B. Google)	nicht gewählt	12	12,1%
	ausgewählt	87	87,9%
spezielle Diagnoseprogramme für den Computer	nicht gewählt	96	97,0%
	ausgewählt	3	3,0%
Gesundheits-Apps	nicht gewählt	87	87,9%
	ausgewählt	12	12,1%
Sonstiges	nicht gewählt	88	88,9%
	ausgewählt	11	11,1%

Tabelle 8: Angabe der Patienten, worauf sie bei der Recherche der Symptome zurückgegriffen haben

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	ja	58	53,2	54,2	54,2
	nein	49	45,0	45,8	100,0
	Gesamt	107	98,2	100,0	
Fehlend	System	2	1,8		
Gesamt		109	100,0		

Tabelle 9: Anzahl der Patienten mit und ohne Hautkrebs

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig:	Hautkrebs haben	58	100,0	100,0	1,7
	Malignes Melanom	28	48,3	48,3	87,9
	Basaliom	14	24,1	24,1	31,0
	Spinaliom	3	5,2	5,2	100,0
	Basaliom, Malignes Melanom	3	5,2	5,2	93,1
	Basaliom, Spinaliom	1	1,7	1,7	32,8
	Aktinische Keratose	2	3,4	3,4	5,2
	Merkelzellkarzinom	1	1,7	1,7	94,8
	B-Zell-Lymphom	1	1,7	1,7	6,9
	kann sich nicht erinnern	4	6,9	6,9	39,7
	Gesamt	58	100,0	100,0	

Tabelle 10: Hautkrebsverteilung bei den Patienten

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	App für Smartphone/Tablet	31	28,4	32,0	32,0
	selbständiges Endgerät nur für Hautkrebsdiagnostik	29	26,6	29,9	76,3
	Computerprogramm	14	12,8	14,4	46,4
	ich würde unter keinen Umständen eine KI verwenden	23	21,1	23,7	100,0
	Gesamt	97	89,0	100,0	
Fehlend	System	12	11,0		
Gesamt		109	100,0		

Tabelle 11: Präferenz der Patienten, auf welche Weise sie KI einsetzen würden

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 (auf keinen Fall)	16	14,7	20,8	20,8
	2	12	11,0	15,6	36,4
	3	14	12,8	18,2	54,5
	4	5	4,6	6,5	61,0
	5 (auf jeden Fall)	30	27,5	39,0	100,0
	Gesamt	77	70,6	100,0	
Fehlend	System	32	29,4		
Gesamt		109	100,0		

Tabelle 12: Bereitschaft, sich selbstständig mittels KI zu untersuchen

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	2	4	19,0	23,5	23,5
	3	3	14,3	17,6	41,2
	4	1	4,8	5,9	47,1
	5 (auf jeden Fall)	9	42,9	52,9	100,0
	Gesamt	17	81,0	100,0	
Fehlend	System	4	19,0		
Gesamt		21	100,0		

Tabelle 13: Bereitschaft, sich selbstständig mittels KI zu untersuchen, bei hohem Vorwissen

		Anzahl		Anzahl in %	
Die KI diagnostiziert die Hautveränderung als Krebs:	1 (auf keinen Fall)		0		0,0%
	2		1		1,7%
	3		0		0,0%
	4		1		1,7%
	5 (auf jeden Fall)		58		96,7%
Die KI diagnostiziert die Hautveränderung als gutartig:	1 (auf keinen Fall)		1		1,6%
	2		4		6,6%
	3		7		11,5%
	4		5		8,2%
	5 (auf jeden Fall)		44		72,1%

Tabelle 14: Verlangen, den Arzt aufzusuchen, in Abhängigkeit vom Ergebnis der Untersuchung mittels KI

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 (auf keinen Fall)	15	13,8	14,6	14,6
	2	19	17,4	18,4	33,0
	3	39	35,8	37,9	70,9
	4	12	11,0	11,7	82,5
	5 (auf jeden Fall)	18	16,5	17,5	100,0
	Gesamt	103	94,5	100,0	
Fehlend	System	6	5,5		
Gesamt		109	100,0		

Tabelle 15: Anteil der Patienten, die auf den Einsatz von KI beim Arzt bestehen

			Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Keine Angabe	Fehlend	System	1	100,0		
Stimulus	Gültig	1 (auf keinen Fall)	22	40,0	42,3	42,3
		2	12	21,8	23,1	65,4
		3	13	23,6	25,0	90,4
		4	2	3,6	3,8	94,2
		5 (auf jeden Fall)	3	5,5	5,8	100,0
		Gesamt	52	94,5	100,0	
	Fehlend	System	3	5,5		
Gesamt		55	100,0			
Kontrollgruppe	Gültig	1 (auf keinen Fall)	29	54,7	56,9	56,9
		2	7	13,2	13,7	70,6
		3	12	22,6	23,5	94,1
		4	2	3,8	3,9	98,0
		5 (auf jeden Fall)	1	1,9	2,0	100,0
		Gesamt	51	96,2	100,0	
	Fehlend	System	2	3,8		
Gesamt		53	100,0			

Tabelle 16: Vertrauen in die Diagnose einer KI, ohne vom Arzt untersucht worden zu sein

			Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Keine Angabe	Fehlend	System	1	100,0		
Stimulus	Gültig	1 (dem Arzt vertrauen)	22	40,0	45,8	45,8
		2	12	21,8	25,0	70,8
		3	12	21,8	25,0	95,8
		4	2	3,6	4,2	100,0
		Gesamt	48	87,3	100,0	
	Fehlend	System	7	12,7		
	Gesamt	55	100,0			
Kontrollgruppe	Gültig	1 (dem Arzt vertrauen)	28	52,8	59,6	59,6
		2	9	17,0	19,1	78,7
		3	10	18,9	21,3	100,0
		Gesamt	47	88,7	100,0	
	Fehlend	System	6	11,3		
		Gesamt	53	100,0		

Tabelle 17: Vertrauen der Patienten in die KI oder den Arzt in den Subgruppen Stimulus und Kontrollgruppe

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	3	7	33,3	35,0	35,0
	4	4	19,0	20,0	55,0
	5 (auf jeden Fall)	9	42,9	45,0	100,0
	Gesamt	20	95,2	100,0	
Fehlend	System	1	4,8		
Gesamt		21	100,0		

Tabelle 18: Zustimmung der Patienten für den klinischen Einsatz von KI bei hohem Vorwissen

		Anzahl	Anzahl in %
Diagnosesicherheit	Rangplatz 1	16	42,1%
	Rangplatz 2	12	31,6%
	Rangplatz 3	5	13,2%
	Rangplatz 4	5	13,2%
jederzeit Verfügbarkeit	Rangplatz 1	7	18,4%
	Rangplatz 2	6	15,8%
	Rangplatz 3	15	39,5%
	Rangplatz 4	10	26,3%
anschließend schnelle Terminvermittlung zum Arzt	Rangplatz 1	12	30,0%
	Rangplatz 2	19	47,5%
	Rangplatz 3	8	20,0%
	Rangplatz 4	1	2,5%
Erstattung durch die Krankenkasse	Rangplatz 1	5	13,2%
	Rangplatz 2	1	2,6%
	Rangplatz 3	10	26,3%
	Rangplatz 4	22	57,9%

Tabelle 19: Prioritäten der Patienten für den präklinischen Einsatz von KI

			Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
.	Fehlend	System	1	100,0		
Stimulus	Gültig	1 (auf keinen Fall)	4	7,3	7,5	7,5
		2	5	9,1	9,4	17,0
		3	15	27,3	28,3	45,3
		4	18	32,7	34,0	79,2
		5 (auf jeden Fall)	11	20,0	20,8	100,0
		Gesamt	53	96,4	100,0	
	Fehlend	System	2	3,6		
	Gesamt	55	100,0			
Kontrollgruppe 1	Gültig	1 (auf keinen Fall)	6	11,3	11,8	11,8
		2	6	11,3	11,8	23,5
		3	23	43,4	45,1	68,6
		4	6	11,3	11,8	80,4
		5 (auf jeden Fall)	10	18,9	19,6	100,0
		Gesamt	51	96,2	100,0	
	Fehlend	System	2	3,8		
	Gesamt	53	100,0			

Tabelle 20: Bereitschaft KI Verwandten weiterzuempfehlen in den Subgruppen Stimulus und Kontrollgruppe

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 (auf keinen Fall)	30	27,5	28,6	28,6
	2	19	17,4	18,1	46,7
	3	25	22,9	23,8	70,5
	4	15	13,8	14,3	84,8
	5 (auf jeden Fall)	16	14,7	15,2	100,0
	Gesamt	105	96,3	100,0	
Fehlend	System	4	3,7		
Gesamt		109	100,0		

Tabelle 21: Bedenken der Patienten bezüglich der Datensicherheit beim Einsatz von KI

Anstellung			Fachrichtung		Gesamt
			Dermatologe	Hausarzt (Allgemeinmedi- ziner/Internist)	
Anstellung	ambulant	Anzahl	22	30	52
		% von Fachrichtung spezifisch	42,3%	57,7%	100,0%
		% von Fachrichtung	73,3%	100,0%	86,7%
	stationär	Anzahl	3	0	3
		% von Fachrichtung spezifisch	100,0%	0,0%	100,0%
		% von Fachrichtung	10,0%	0,0%	5,0%
	beides	Anzahl	5	0	5
		% von Fachrichtung spezifisch	100,0%	0,0%	100,0%
		% von Fachrichtung	16,7%	0,0%	8,3%
Gesamt	Anzahl	30	30	60	
	% von Fachrichtung spezifisch	50,0%	50,0%	100,0%	
	% von Fachrichtung	100,0%	100,0%	100,0%	

Tabelle 22: Fachrichtung und Anstellungsverhältnis der Ärzte

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Facharzt	58	96,7	96,7	96,7
	Assistenzarzt	2	3,3	3,3	100,0
	Gesamt	60	100,0	100,0	

Tabelle 23: Ärztliches Ausbildungsniveau

		Anzahl	Anzahl in %
KI generell gehört	ja	59	98,3%
	nein	1	1,7%
Inwieweit würden Sie von Sich selbst behaupten, zu wissen, wie eine KI funktioniert?	1 (keine Ahnung)	5	8,3%
	2	10	16,7%
	3	19	31,7%
	4	18	30,0%
	5 (sehr gutes Verständnis)	8	13,3%

Tabelle 24: Bekanntheit von KI und Wissen um die Funktionsweise von KI

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
KI Nutzung	ja	11	18,3	18,3	18,3
	nein	47	78,3	78,3	96,7
	Ich diagnostiziere keinen Hautkrebs.	2	3,3	3,3	100,0
	Gesamt	60	100,0	100,0	

Tabelle 25: Einsatz von KI im Rahmen der Hautkrebsdiagnostik

		Anzahl	Anzahl in %
Internetsuchmaschinen (z.B. Google)	nicht gewählt	3	5,6%
	ausgewählt	51	94,4%
spezielle Diagnoseprogramme für den Computer	nicht gewählt	45	83,3%
	ausgewählt	9	16,7%
Gesundheits-Apps	nicht gewählt	48	88,9%
	ausgewählt	6	11,1%
Sonstiges	nicht gewählt	34	63,0%
	ausgewählt	20	37,0%

Tabelle 26: Mediennutzung der Ärzte zur Information über Krankheitssymptome

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 (auf keinen Fall)	23	38,3	38,3	38,3
	2	15	25,0	25,0	63,3
	3	11	18,3	18,3	81,7
	4	9	15,0	15,0	96,7
	5 (auf jeden Fall)	2	3,3	3,3	100,0
	Gesamt	60	100,0	100,0	

Tabelle 27: Patienten den selbstständigen Einsatz von KI empfehlen

		Anzahl	Anzahl in %
Die KI diagnostiziert die Hautveränderung als bösartig: Würden Sie dem Patienten dazu raten, Sie aufzusuchen?	1 (auf keinen Fall)	0	0,0%
	2	1	2,7%
	3	1	2,7%
	4	2	5,4%
	5 (auf jeden Fall)	33	89,2%
Die KI diagnostiziert keine bösartige Hauterscheinung: Würden Sie dem Patienten dazu raten, Sie aufzusuchen?	1 (auf keinen Fall)	0	0,0%
	2	2	5,4%
	3	12	32,4%
	4	8	21,6%
	5 (auf jeden Fall)	15	40,5%

Tabelle 28: Empfehlung der Ärzte an die Patienten, einen Arzt aufzusuchen, in unterschiedlichen hypothetischen Situationen

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 (auf keinen Fall)	2	3,3	3,3	3,3
	2	1	1,7	1,7	5,0
	3	17	28,3	28,3	33,3
	4	15	25,0	25,0	58,3
	5 (auf jeden Fall)	25	41,7	41,7	100,0
	Gesamt	60	100,0	100,0	

Tabelle 29: Bereitschaft der Ärzte KI bei ihrer Untersuchung als Hilfsmittel einzusetzen

Fachrichtung		Gültig	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Dermatologe	Gültig	3	4	19,0	19,0	19,0
		4	4	19,0	19,0	38,1
		5 (auf jeden Fall)	13	61,9	61,9	100,0
		Gesamt	21	100,0	100,0	
Hausarzt (Allgemeinmediziner/ Internist)	Gültig	4	2	40,0	40,0	40,0
		5 (auf jeden Fall)	3	60,0	60,0	100,0
		Gesamt	5	100,0	100,0	

Tabelle 30: Bereitschaft der Ärzte bei hohem Vorwissen KI bei ihrer Untersuchung als Hilfsmittel einzusetzen

		Anzahl	Anzahl in %
Wie würden Sie vorgehen, wenn Sie eine therapeutisch relevante Diagnose fällen, die KI diese jedoch verneint?	therapieren	9	15,5%
	nicht therapieren	0	0,0%
Wie würden Sie vorgehen, wenn die KI eine therapeutisch relevante Diagnose fällen, Sie diese jedoch ausschließen?	Biopsie nehmen	38	65,5%
	andere Optionen wählen	11	19,0%
Wie würden Sie vorgehen, wenn Sie eine therapeutisch relevante Diagnose fällen, die KI diese jedoch verneint?	therapieren	2	3,4%
	nicht therapieren	4	6,9%
Wie würden Sie vorgehen, wenn die KI eine therapeutisch relevante Diagnose fällen, Sie diese jedoch ausschließen?	Biopsie nehmen	40	69,0%
	andere Optionen wählen	12	20,7%

Tabelle 31: Vorgehen der Ärzte in verschiedenen hypothetischen Situationen

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 (auf keinen Fall)	5	8,3	8,3	8,3
	2	1	1,7	1,7	10,0
	3	21	35,0	35,0	45,0
	4	19	31,7	31,7	76,7
	5 (auf jeden Fall)	14	23,3	23,3	100,0
	Gesamt	60	100,0	100,0	

Tabelle 32: Bereitschaft der Ärzte dem Patientenwunsch nach dem Einsatz von KI nachzukommen

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 (auf keinen Fall)	2	3,3	3,8	3,8
	2	13	21,7	24,5	28,3
	3	8	13,3	15,1	43,4
	4	9	15,0	17,0	60,4
	5 (auf jeden Fall)	21	35,0	39,6	100,0
	Gesamt	53	88,3	100,0	
Fehlend	nicht beantwortet	7	11,7		
Gesamt		60	100,0		

Tabelle 33: Ansicht der Ärzte, ob das Verständnis um die Funktionsweise von KI als Voraussetzung für deren Einsatz gelten sollte

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 (auf keinen Fall)	2	3,3	3,4	3,4
	2	4	6,7	6,9	10,3
	3	22	36,7	37,9	48,3
	4	16	26,7	27,6	75,9
	5 (auf jeden Fall)	14	23,3	24,1	100,0
	Gesamt	58	96,7	100,0	
Fehlend	nicht beantwortet	2	3,3		
Gesamt		60	100,0		

Tabelle 34: Meinung der Ärzte, ob KI Teil ihrer Facharztausbildung sein sollte

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	verbessert Versorgung	50	83,3	83,3	83,3
	verschlechtert Versorgung	1	1,7	1,7	85,0
	hat keinen Einfluss auf die Versorgung	9	15,0	15,0	100,0
	Gesamt	60	100,0	100,0	

Tabelle 35: Einschätzung der Ärzte in Bezug auf das Potential von KI in der Hautkrebstdiagnostik

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 (auf keinen Fall)	4	6,7	6,7	6,7
	2	3	5,0	5,0	11,7
	3	8	13,3	13,3	25,0
	4	15	25,0	25,0	50,0
	5 (auf jeden Fall)	30	50,0	50,0	100,0
	Gesamt	60	100,0	100,0	

Tabelle 36: Bereitschaft der Ärzte, die während der Diagnostik anfallenden Daten zum Training von KI bereitzustellen

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Diagnosesicherheit (hohe Sensitivität)	49	81,7	81,7	81,7
	Erkennen möglichst vieler verschiedener Diagnosen	7	11,7	11,7	93,3
	Andere als die hier genannten Eigenschaften	4	6,7	6,7	100,0
	Gesamt	60	100,0	100,0	

Tabelle 37: Rating der Ärzte zu wichtigen Eigenschaften von KI

		Anzahl	Anzahl in %
leichte Bedienbarkeit	Rangplatz 1	24	41,4%
	Rangplatz 2	23	39,7%
	Rangplatz 3	9	15,5%
	Rangplatz 4	2	3,4%
Nachvollziehbarkeit des Ergebnisses	Rangplatz 1	27	46,6%
	Rangplatz 2	16	27,6%
	Rangplatz 3	13	22,4%
	Rangplatz 4	2	3,4%
Analysegeschwindigkeit (kein Zeitverlust)	Rangplatz 1	5	8,6%
	Rangplatz 2	16	27,6%
	Rangplatz 3	25	43,1%
	Rangplatz 4	12	20,7%
Delegierbarkeit	Rangplatz 1	2	3,4%
	Rangplatz 2	3	5,2%
	Rangplatz 3	11	19,0%
	Rangplatz 4	42	72,4%

Tabelle 38: Angabe der Ärzte, was ihnen bei der Anwendung einer KI wichtig wäre

9 Danksagung

Mein Dank gilt in erster Linie Dr. Rietz, der mich durch diese Arbeit begleitet und mir die spannende Thematik der KI nähergebracht hat. Er hat mir immer wieder neue Denkanstöße gegeben und mich bei der praktischen Umsetzung der Befragung unterstützt.

Darüber hinaus möchte ich dem Pflegepersonal der Hautklinik der Universitätsmedizin Mainz danken, welches mich bei der Durchführung der Befragung sehr unterstützt hat.

Ich danke auch allen Befragten für ihre Teilnahme, ohne die diese Arbeit nicht zustande gekommen wäre.

Zuletzt danke ich meinen Eltern, meinen Geschwistern und meiner Verlobten für die Unterstützung und den Rückhalt während des gesamten Studiums.

10 Tabellarischer Lebenslauf

Persönliche Daten:

Nachname: Klaedtke
Vorname: Damian
Geburtsdatum: 24.08.1997
Geburtsort: Lahnstein

Ausbildung:

Oktober 2017 bis November 2023 Studium der Humanmedizin
Johannes Gutenberg-Universität Mainz
März 2017 Abitur
Johannes-Gymnasium Lahnstein