

Aus der Praxis für Pränataldiagnostik in der Friedrichstrasse 147 in Berlin

DISSERTATION

Ultraschall in der Pränataldiagnostik-
Standardisierung durch künstliche Intelligenz am Beispiel der
Neurosonographie und der Darstellung des Aortenbogens

Ultrasound in prenatal diagnosis- Standardisation through artificial
intelligence based on the example of neurosonography and
demonstration of aortic arch

zur Erlangung des akademischen Grades Doctor medicinae (Dr. med.)

vorgelegt der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

von

Julia Fabienne Umlauff

Datum der Promotion: 23.03.2024

Inhaltsverzeichnis

Tabellenverzeichnis	4
Abbildungsverzeichnis	5
Abkürzungsverzeichnis	7
1 Abstract auf Deutsch	8
2 Abstract auf Englisch	9
3 Einleitung	10
3.1 Einführung in das Thema	10
3.2 Technik der Sonographie	11
3.3 Anfänge der Ultraschalluntersuchungen allgemein und Entwicklung im Laufe der Jahrzehnte	11
3.4 Ultraschalluntersuchungen in der Geburtshilfe mit Schwerpunkt fetale Biometrie und Aortenbogen	12
3.4.1 Anfänge	12
3.4.2 Fetale Biometrie	13
3.4.3 Der fetale Aortenbogen	14
3.5 Matrixsonde und künstliche Intelligenz sowie deren Anwendung in Bezug auf die fetale Biometrie und die Darstellung des Aortenbogens	16
3.5.1 Elektronische Matrixsonde	16
3.5.2 Automatisierte Messung des fetalen Kopfes mittels künstlicher Intelligenz	17
3.5.3 Biplane Technik	18
3.6 Zielsetzung	20
4 Material und Methoden	22
4.1 Studienaufbau	22
4.2 Ultraschallgerät	23
4.3 Durchführung der Ultraschalluntersuchung	23
4.4 Definition der Messebenen und manuelle Messung	24

4.5	Messung der gesuchten Parameter mit Hilfe der künstlichen Intelligenz . . .	27
4.6	Vergleich beider Messmethoden	28
4.7	Darstellung des fetalen Aortenbogens mit der Biplane Technik	29
4.8	Bildmaterial	29
4.8.1	Fetale Biometrie	30
4.8.2	Fetaler Aortenbogen	32
4.9	Statistische Analyse	33
4.9.1	Begründung der Fallzahl	33
5	Ergebnisse	34
5.1	Fetale Biometrie	34
5.1.1	Deskriptive Analyse	35
5.1.2	Interrater-Reliabilität	38
5.1.3	Darstellung der Messungen von Untersucher 1 und KI mittels Bland- Altman-Plots	39
5.1.4	Varianzanalyse von potentiellen Störfaktoren	45
5.1.5	Multiple Regression für alle Kriterien mit mehr als einem signifikan- ten Prädiktor	51
5.1.6	Fehlende Volumenaufnahmen und Fehlmessungen	51
5.1.7	Fehlerwertanalyse	53
5.1.8	Deskriptive Analyse der Ergebnisse des Untersuchers 2 und der KI	54
5.1.9	Interrater-Reliabilität des Untersuchers 2	57
5.1.10	Darstellung der Messungen von Untersucher 2 und KI mittels Bland- Altman-Plots	58
5.1.11	Vergleich Untersucher 1 und Untersucher 2	63
5.2	Fetaler Aortenbogen	64
5.2.1	Zusammenfassung der Ergebnisse in Bezug auf den maternalen BMI	68
5.2.2	Vergleich der Darstellbarkeit des Aortenbogens in Bezug auf die in- trauterine Lage	69
5.2.3	Vergleich der Darstellbarkeit des Aortenbogens in Bezug auf die dorsoanteriore Stellung	70
6	Diskussion	72

6.1	Schlussfolgerung	72
6.1.1	Diskussion der Ergebnisse am Beispiel des Hinteren Hirnseitenven- trikels (VP)	76
6.1.2	Diskussion der Ergebnisse am Beispiel der Cisterna magna (CM) .	77
6.2	Limitation der Methode	81
6.3	Ausblick	83
	Literatur	85
	Eidesstattliche Versicherung	90
	Lebenslauf	92
	Danksagung	94
	Bescheinigung des akkreditierten Statistikers	95

Tabellenverzeichnis

2	Angaben zur elektronischen Matrixsonde	17
3	Übersicht über das Patientinnenkollektiv	23
4	Tolerable Messwertabweichung	28
5	Übersicht über die Anzahl der erfolgten Untersuchungen in Bezug auf die fetale Biometrie	35
6	Anzahl der nicht von der KI gemessenen Werte	35
7	Deskriptive Datenanalyse der Daten des Untersuchers 1 und der KI	37
8	Intraklassen-Korrelation Untersucher 1	38
9	Fehlerwertanalyse	53
10	Deskriptive Datenanalyse der Daten des Untersuchers 2 und der KI	54
11	Intraklassen-Korrelation Untersucher 2	57
12	BMI-Verteilung über die Gruppen	66
13	Übersicht intrauterine Lagen und ihre Häufigkeitsverteilung	70
14	Übersicht dorsoanteriore Stellung und ihre Häufigkeitsverteilung	71

Abbildungsverzeichnis

1	Aortenbogen Längsschnitt	16
2	Beispiel der Biplane Technik	19
3	Manuelle Messung BPD und FOD	25
4	Manuelle Messung von CM und Cerebellum	26
5	Manuelle Messung des Ventrikels	27
6	Volumenaufnahme des Kopfes mit dem SonoCNS Programm	30
7	Differenz der Messung der KI und des U1 am dargestellten Beispiel	31
8	Biplane Modus mittels Matrixsonde	32
9	Grafische Darstellung der paarweisen Messungen pro Parameter des Untersuchers 1 und der KI	36
10	Bland-Altman-Plot für den Parameter BPD, Untersucher 1	40
11	Bland-Altman-Plot für den Parameter KU, Untersucher 1	41
12	Bland-Altman-Plot für den Parameter VP, Untersucher 1	42
13	Bland-Altman-Plot für den Parameter CM, Untersucher 1	43
14	Bland-Altman-Plot für den Parameter Cerebellum, Untersucher 1	44
15	Einfluss des BMI auf die Messdifferenz	46
16	Einfluss des fetalen Gewichts auf die Messdifferenz	47
17	Einfluss der fetalen Lage auf die Messdifferenz	48
18	Einfluss der dorsoanterioren Stellung auf die Messdifferenz	49
19	Einfluss der Lage der Plazenta auf die Messdifferenz	50
20	Einfluss von Gewicht und Plazentalage auf die Messung des Hirnseitenventrikels	51
21	Fehlerhafte KI Messung mit falschen Messebenen	52
22	Grafische Darstellung der paarweisen Messungen für Untersucher 2 und KI	55
23	Bland-Altman-Plot für den Parameter BPD, Untersucher 2	58
24	Bland-Altman-Plot für den Parameter KU, Untersucher 2	59
25	Bland-Altman-Plot für den Parameter VP, Untersucher 2	60
26	Bland-Altman-Plot für den Parameter CM, Untersucher 2	61
27	Bland-Altman-Plot für den Parameter Cerebellum, Untersucher 2	62
28	Darstellbarkeit des Aortenbogens und der Abgangsgefäße mittels Biplaner Bildgebung	64

29	Grafische Darstellung der Verteilung der Messungen in Bezug auf den maternalen BMI	65
30	Boxplot zur Darstellung der Verteilung der Messungen in Bezug auf den maternalen BMI	67
31	Volumenaufnahme mit fehlender Ventrikelmessung	77
32	Nicht optimale Darstellung der Cisterna magna durch die künstliche Intelligenz	80

Abkürzungsverzeichnis

AU	Abdomenumfang
BMI	Body-Mass-Index
BPD	Biparietaler Durchmesser
Cereb	Cerebellum, auch CB
CM	Cisterna magna
df	Number of degrees of freedom (Freiheitsgrade)
FOD	Frontookzipitaler Durchmesser
GA	Gestationsalter
G-BA	Gemeinsamer Bundesausschuss
ICC	Intra-Class-Correlation
HC	Kopfumfang, synonym zu KU
KI	Künstliche Intelligenz
KU	Kopfumfang
LoA	Limit of Agreement
MV	Mittelwert
n	Anzahl der Untersuchungen
SD	Standardabweichung
SSW	Schwangerschaftswoche
U1	Untersucher 1
U2	Untersucher 2
VP	Hinterer Hirnseitenventrikel

1 Abstract auf Deutsch

Die fetale Biometrie des Kopfes und die Darstellung des Aortenbogens sind wichtige Bestandteile der Pränataldiagnostik im zweiten Trimenon. In der folgenden Arbeit geht es um die Messung der fetalen Biometrie mit Hilfe künstlicher Intelligenz im Vergleich zur herkömmlichen manuellen Messung, hierfür wurden die Daten von 393 Patientinnen eines ersten Untersuchers und 20 Patientinnen eines zweiten Untersuchers ausgewertet. Das Programm SonoCNS von GE wurde verwendet. Die Daten wurden mittels Bland-Altman-Plots verglichen, hier zeigte sich eine Äquivalenz der Messmethoden für die Parameter Biparietaler Durchmesser, hinterer Hirnseitenventrikel und Cerebellum. Keine genaue statistische Aussage kann zum Parameter Kopfumfang getroffen werden. Für die Cisterna magna konnte gezeigt werden, dass die künstliche Intelligenz und der menschliche Untersucher nicht äquivalent messen. Außerdem wurden anhand von Störfaktoren (BMI der Mutter, Gewicht des Feten, Lage des Feten, Stellung des Feten und Plazentalokalisation) mit logistischen Regressionen überprüft, ob diese Faktoren einen Einfluss auf die Messung der künstlichen Intelligenz hatten. Hier wurde eine minimale Auswirkung des fetalen Gewichtes und der Plazentalage auf die Messung des hinteren Hirnseitenventrikels gefunden. Die Studie kommt insgesamt zu dem Schluss, dass die künstliche Intelligenz gute und reproduzierbare Messergebnisse im Bereich der fetalen Biometrie des Kopfes erzielen kann. Im zweiten Teil der Studie geht es um die Nutzbarkeit der Biplane Technik zur Darstellung des fetalen Aortenbogens, hier wurden die Daten von 392 Patientinnen ausgewertet. Mit der Biplane Technik können in Echtzeit Bilder erstellt werden, die orthogonal zueinander stehen. Dabei handelt es sich bei beiden Bildern um originale Ultraschallbilder. In 55,98 % der Fälle gelang die Darstellung des Aortenbogens mit dieser Technik, in allen anderen Fällen konnte kein adäquates Bild generiert werden. In 27,81 % der Fälle war es technisch nicht möglich, überhaupt ein Bild zu generieren. War die Anwendung der biplane Technik aber möglich, gelang die Darstellung in knapp 78 %. Auch hier wurden die Störfaktoren BMI und dorsoanteriore Stellung des Feten auf ihren Einfluss auf die Darstellung überprüft. Sowohl die maternale Adipositas als auch die dorsoanteriore Stellung haben einen negativen Einfluss auf die Darstellung, wobei der Einfluss der Stellung größer ist.

2 Abstract auf Englisch

The fetal biometry of the head and the demonstration of the aortic arch are important parts of the prenatal diagnostics in the second trimester of pregnancy. This paper is about the measurement of the fetal biometry with the aid of artificial intelligence in comparison to the conventional manual measurement. Therefore the data of 393 patients of a first examiner and 20 patients of a second examiner were evaluated. The program SonoCNS from GE was used. The data were compared with Bland Altman- plots. The results show an equivalence for the parameters biparietal diameter, lateral ventricle and cerebellum. No exact statistic statement can be given for the parameter head circumference. The plots show that there is no equivalence between artificial intelligence and the human examiner for the measurement of the cisterna magna. In addition to that, it was tested with logistic regression if confounding factors (maternal BMI, fetal weight, fetal position and placental localisation) had an impact to the artificial intelligence measurement. A minimal influence of the fetal weight and the placental localisation was found for the lateral ventricle. The study concludes overall, that artificial intelligence obtains good and reproducible measurement result for the fetal biometry of the head. The second part of the study is about the usability of the Biplane technique for the demonstration of the aortic arch. The data of 392 patients were analyzed. Biplane technique allows generating ultrasound images which are orthogonal to each other in real time. Both ultrasound images are original images. In 55,98 % of the cases the demonstration of the aortic arch was possible with Biplane technique, in the other cases no adequate image could be obtained. In 27,81 % it was not possible in general to obtain a image at all. If the application of Biplane technique was possible in general, the demonstration of the aortic arch was possible in almost 78 %. The influence of the confounding factors maternal BMI and fetal position were tested. Both factors have a negative influence of the demonstration of the aortic arch, whereby the influence of the fetal position is larger.

3 Einleitung

3.1 Einführung in das Thema

Die Ultraschalldiagnostik nimmt eine zentrale Rolle in der Pränatalmedizin ein und ist auch in den Mutterschaftsrichtlinien verankert [1]. Die Ultraschalluntersuchung des Feten soll nach den Mutterschaftsrichtlinien des gemeinsamen Bundesausschusses (G-BA) [2] in jedem Trimenon zu festgelegten Zeiten durchgeführt werden¹ und dient unter anderem der Feststellung der fetalen Entwicklung anhand von Normkurven. Durch Einstellung der dafür erforderlichen Messebenen ergeben sich zudem Hinweise auf eventuell vorliegende Fehlbildungen. In der vorliegenden Arbeit geht es um das erweiterte Ultraschallscreening im zweiten Trimenon der Schwangerschaft. Auf die weiteren Ultraschalluntersuchungen in der Schwangerschaft soll daher nur oberflächlich eingegangen werden. Das erweiterte Ultraschallscreening oder sogenanntes Organscreening wird bei Risikoschwangerschaften, bereits detektierten Auffälligkeiten im Basisultraschall sowie bei psychischer Belastung der Schwangeren, einer sogenannten Fehlbildungsangst, angewandt. Hier werden im Gegensatz zum in den Mutterschaftsrichtlinien für alle vorgesehenen Basisultraschall nicht nur Biometriedaten, Sitz der Plazenta und einige andere Punkte untersucht, sondern auch gezielt Organsysteme dargestellt. Der Untersucher oder die Untersucherin muss hierfür eine Prüfung bei der Kassenärztlichen Vereinigung absolvieren. Im Rahmen dieser Untersuchung können einige fetale Auffälligkeiten detektiert werden, die eine prä- oder postnatale Behandlung benötigen. In Studien konnte gezeigt werden, dass die Mortalität und Morbidität durch pränatale Erkennung nachweislich gesenkt wird. Ein Beispiel hierfür ist im Bereich des fetalen Herzens die pränatale Diagnose einer Transposition der großen Arterien [3] sowie die Aortenisthmusstenose. Viele fetale Auffälligkeiten sind dank moderner Ultraschallgeräte und hoch qualifizierten Untersuchern und Untersucherinnen bereits im Ersttrimesterscreening detektierbar, andere können erst mit fortschreitender Schwangerschaft festgestellt werden, als Beispiele seien hier die Mikrocephalie und die zystisch adenomatoide Malformation der Lunge genannt. Ultraschalluntersuchungen haben aber auch stets Limitationen - die Patientin muss im Vorfeld der Untersuchung darüber aufgeklärt werden, dass nicht alle Entwicklungsstörungen erkannt werden können und nicht alle auffälligen Befunde behandelt werden können. Insbesondere stellen maternale adipöse

¹9.-12. SSW, 19.-22. SSW und 29.-32. SSW

Bauchdecken und eine ungünstige fetale Lage intrauterin ultrasonografische Hindernisse dar. Die pränatale Diagnostik profitiert von stetig verbesserten hochauflösenden Ultraschallgeräten. In der vorliegenden Arbeit soll es vor allem um die Nutzbarkeit der elektronischen Matrixsonde und der künstlichen Intelligenz gehen sowie um die Möglichkeit, damit fetale Strukturen besser darstellen zu können und somit etablierte Untersuchungen zu vereinfachen und gegebenenfalls die pränatale Detektionsrate von Fehlbildungen zu erhöhen.

3.2 Technik der Sonographie

Die konventionelle Sonographie ist ein nicht-invasives Verfahren, bei dem mithilfe von Ultraschallwellen Dichteunterschiede von Gewebe sichtbar gemacht werden. So können innere Organe oder Strukturen sowie Feten im Mutterleib dargestellt werden. Diese Methode ist unschädlich, es entsteht keine Strahlenbelastung. Ultraschallwellen sind hochfrequente Schallwellen, welche im Ultraschallkopf durch spezielle Kristalle erzeugt und wieder empfangen werden können [4]. Die Kristalle werden durch elektrische Spannung verformt, dies wird piezoelektrischer Effekt genannt. Hierbei entstehen elektrische Impulse, welche vom Ultraschallgerät als Bildpunkte dargestellt werden. Die Besonderheit der in der Studie genutzten elektronischen Matrixsonde liegt in der hohen Anzahl an eingebauten Kristallen, dies wird in Kapitel 3.5.3 genauer erläutert. Die Schallfrequenz ist die Anzahl der Wellen pro Sekunde und wird in Hertz ($\text{Hz} = 1 \text{ Schwingung/s}$) angegeben [5]. Je höher die Frequenz, desto besser die Bildauflösung, aber desto geringer die Eindringtiefe [5]. Mittlerweile ist die Ultraschalluntersuchung ein in fast allen medizinischen Fächern routinemäßig eingesetztes Verfahren und kann zur Detektion von zahlreichen Erkrankungen genutzt werden.

3.3 Anfänge der Ultraschalluntersuchungen allgemein und Entwicklung im Laufe der Jahrzehnte

Die ersten diagnostischen Ultraschalluntersuchungen wurden 1942 beschrieben [6] und im Laufe der Jahre stetig weiterentwickelt. Zunächst gab es nur das eindimensionale A-Bild. Hierbei werden Ultraschallechos als Amplituden auf einer Zeitachse dargestellt. Die nächste Entwicklung in den 1950er Jahren war das B-Bild, ein zweidimensionales Grau-

stufenbild, bei dem die Echos als Helligkeit (B=Brightness) dargestellt wurden [7]. Nachteil war allerdings die lange Zeitdauer der Bildentstehung und somit der Untersuchungsdauer, wobei viele Bewegungsartefakte entstanden [8]. Die Notwendigkeit einer schnelleren Abtastung führte zur Entwicklung des „Vidoson-Systems“, eines Ultraschallsystems, bei dem Bilder in Echtzeit angezeigt werden konnten. Somit konnten bewegte Bilder erzeugt werden [9]. Die Eindringtiefe des Vidoson-Systems wurde mit 15-16cm angegeben. Über diese Entwicklung erhielt der Ultraschall auch Einzug in die Pränataldiagnostik-der Fet konnte nun in seiner Eigenschaft als sich bewegende „Struktur“ auf dem Monitor dargestellt werden. Als nächster Meilenstein konnte durch die Entwicklung der Dopplersonographie ein Nachweis über die fetale Herzaktion im Ultraschall 1964 erbracht werden [10]. Durch die Grauwertabstufung gelang es, Strukturen besser zu differenzieren [11], die Bilder wurden stetig klarer, die Geräte handlicher. Durch die Weiterentwicklung der Computer konnten Ultraschalldaten zunehmend digital verarbeitet werden. Schließlich gelang Mitte der 1980er Jahre die 3-dimensionale Darstellung im Ultraschall [12]. Seit Beginn des 21. Jahrhunderts ist auch eine 4D-Darstellung möglich, also ein 3-dimensionales bewegtes Bild. Durch verbesserte Technik wurden auch die Ultraschallsonden handlicher. Zunächst wurden mechanische Schallköpfe verwendet, in denen sich wenige Kristalle befanden. Diese wurden dann durch die Array-Technik abgelöst, hierbei wurden mehr Kristalle ringförmig angeordnet um so eine bessere Fokussierung und Steuerung des Schallbündels zu erreichen. Die elektronische Matrixsonde kann durch die Anordnung der Kristalle die zu untersuchende Region in mehrere Richtungen gleichzeitig abtasten. Neue Technologien wie die Nutzung der künstlichen Intelligenz haben das Potential, die Ultraschalluntersuchungen weiter zu revolutionieren.

3.4 Ultraschalluntersuchungen in der Geburtshilfe mit Schwerpunkt fetale Biometrie und Aortenbogen

3.4.1 Anfänge

Die Ultraschalldiagnostik erhielt schon früh Einzug in die Gynäkologie und Geburtshilfe. 1958 gelang I.Donald und T.Brown in Großbritannien erstmals die sonografische Darstellung eines Feten [13, Seite 20]. Die Darstellung erfolgte als eindimensionales A-Bild [14]. In Deutschland beschäftigten sich Manfred Hansmann in Bonn und Hans-Jürgen Hollän-

der in Münster in den 1960er Jahren mit Ultraschalluntersuchungen in der Gynäkologie und Geburtshilfe und legten so die Grundsteine für die moderne Pränatalmedizin. Die Anwendung des Ultraschalls führte dazu, dass die mit einer Strahlenbelastung verbundenen Röntgenuntersuchungen in der Schwangerschaft Anfang der 1970er Jahre deutlich zurückgingen. Zunächst lag der Schwerpunkt dieser Ultraschalluntersuchungen weniger auf der Entdeckung von Fehlbildungen als vielmehr in der Erkennung von Mehrlingsgraviditäten und in der korrekten Datierung des Schwangerschaftsalters [13, Seite 7]. Als erstes Land weltweit führte Deutschland bereits 1979 das Ultraschallscreening in die Mutterschaftsrichtlinien ein. Die frühesten Publikationen zum Thema fetale Biometrie sind aus den Jahren 1964, 1977 und 1978, hierauf wird genauer im nächsten Kapitel eingegangen. In den vergangenen Jahrzehnten hat sich die Pränataldiagnostik stetig weiterentwickelt – ein großer Zugewinn ist natürlich die Weiterentwicklung der Technik. Durch verbesserte Geräte und Erfahrung der Untersucherinnen und Untersucher lassen sich heute eine Vielzahl von Entwicklungsstörungen und Auffälligkeiten bereits präpartal erkennen. So können Risiken für Mutter und Kind wie zum Beispiel die Plazenta praevia oder die Präeklampsie anhand von Ultraschallparametern rechtzeitig erkannt und einer adäquaten Behandlung zugeführt werden.

3.4.2 Fetale Biometrie

Mit Beginn der Sonographie in der Geburtshilfe wurde versucht, den Feten auszumessen und somit ein fetales Schätzwert zu errechnen. Dies dient der Kontrolle des fetalen Wachstums und kann zur Diagnose einer fetalen Wachstumsrestriktion oder Makrosomie beitragen sowie Hinweise auf Fehlbildungen durch Einstellungen der geforderten Messebenen geben. Das Aufsuchen und die Einstellung der korrekten Messebenen werden im Methodenteil genauer erläutert. Zu den Biometriemaßen existieren Normkurven, anhand derer sich das fetale Wachstum in Bezug auf das Schwangerschaftsalter und Voruntersuchungen beurteilen lässt. Die Referenzbereiche werden in Perzentilen angegeben. Zunächst wurde vor allem der Biparietale Durchmesser (BPD) zur fetalen Gewichtsschätzung herangezogen [15]. Mit der Messung des BPD wurden auch die ersten Feten mit Anenzephalus pränatal diagnostiziert – in diesen Fällen ist es nicht möglich, den BPD einzustellen, da der Hirnschädel fehlt. Der BPD wurde auch zur Bestimmung des Schwangerschaftsalters herangezogen [16]. Durch lineare Regressionsberechnungen wurden For-

meln zur fetalen Gewichtsschätzung entwickelt, zunächst ausgehend vom BPD. Die Ergebnisse dieser Formeln zeigten allerdings keine signifikante Verbesserung der Gewichtsschätzung im Vergleich zur damals etablierten Methode der Inspektion und Palpation des Abdomens der Schwangeren [17]. Es wurde nun versucht, mit der Hinzunahme der Messung des Brust- und Abdomenumfangs des Feten eine erhöhte Genauigkeit bezüglich der fetalen Gewichtsschätzung zu erlangen [18]. Außerdem wurde zusätzlich zu den Biometriedaten das Gestationsalter zur Berechnung herangezogen [19]. Es konnte gezeigt werden, dass die Bestimmung des fetalen Gewichts unter optimalen Schallbedingungen so auf 10% genau angegeben werden kann [20]. Schließlich wurde noch die Femurlänge in die fetale Gewichtsschätzung mit einbezogen. Hier konnte gezeigt werden, dass die Kombination aus Kopfumfang, biparietalem Durchmesser, Abdomenumfang und Femurlänge signifikant bessere Ergebnisse erzielte als einzelne Parameter [21]. Diese Formel nach Hadlock ist die am häufigsten in Deutschland verwendete Formel seit nunmehr über 30 Jahren. Um die Messungengenauigkeit weiter zu reduzieren wurden weitere Formeln entwickelt in denen Parameter wie das Geschlecht des Kindes, die ethnische Herkunft der Eltern, das Gewicht und die Körpergröße der werdenden Mutter berücksichtigt wurden. Außerdem wurden spezielle Formeln für kleine Feten [22] sowie für makrosome Feten [23] entwickelt, da sich in diesen Fällen die gängigen Formeln als unzuverlässiger erwiesen haben. Die Genauigkeit der fetalen Gewichtsschätzung hängt außerdem von der Erfahrung der untersuchenden Personen ab [24]. Trotz aller Bemühungen und Arbeiten zu dem Thema fetale Gewichtsschätzung bleibt eine Messungengenauigkeit von +/- 10% bestehen oder ± 500 Gramm [25]. In der Studie von Kehl et al. (2012) wird postuliert, dass die Genauigkeit der fetalen Gewichtsschätzung damit ihre Limitationen erreicht hat und weitere Verbesserung nur durch neue Ultraschallansätze erreicht werden kann [25].

3.4.3 Der fetale Aortenbogen

Die Untersuchung des fetalen Aortenbogens ist Bestandteil der Feindiagnostik im zweiten Trimenon. Die fetale Aorta beginnt nach der Aortenklappe des linken Ventrikels und verläuft als Aorta ascendens normalerweise zunächst nach kranial, um dann bogenförmig nach kaudal als Aorta descendens weiterzulaufen. Die bogenförmige Aorta liegt im Mediastinum außerhalb des Perikards und wird als Aortenbogen bezeichnet. Am Aortenbogen zweigen drei Gefäße ab, zunächst der Truncus brachiocephalicus, dann die Arteria carotis

communis sinistra und die Arteria subclavia sinistra. Die abgehenden Gefäße versorgen Kopf und Arme. In Abbildung 1 wird die ultrasonografische Darstellung dieser Region aufgezeigt.

Es gibt 5 Aortenbogenanomalien, welche im Rahmen der Untersuchung detektiert werden können:

1. Aortenisthmusstenose
2. Unterbrechung des Aortenbogens
3. Rechter Aortenbogen
4. Doppelter Aortenbogen
5. Abberante rechte Arteria subclavia

Die Detektion dieser Anomalien ist wichtig, um die Entbindung im Perinatalzentrum zu veranlassen und somit die optimale postpartale Versorgung sowie gegebenenfalls eine Operation des Neugeborenen zu ermöglichen. Hierdurch hat sich das fetale Outcome deutlich verbessert [26]. Des Weiteren zeigen nicht wenige Feten mit diesen Auffälligkeiten Chromosomenstörungen. Daher sollte die Darstellung einer Anomalie zur genetischen Beratung der Eltern führen. Generell liegt die Prävalenz von angeborenen Herzfehlern in Deutschland bei 1,08 Prozent laut der Pan-Studie [27]. Die Häufigkeit der Aortenisthmusstenose bei diesen Kindern wird mit circa 4 Prozent angegeben [27]. Es stellt den viert-häufigsten Herzfehler dar. Hierbei handelt es sich um eine Verengung der Aorta im Bereich des Übergangs von Aortenbogen zur Aorta descendens. Durch die Stenose kann es postnatal nach Verschluss des Ductus Botalli unbehandelt zur Zyanose und zur Herzinsuffizienz kommen. Kausale Therapie ist die Operation mit Beseitigung der Stenose. Die Aortenisthmusstenose kommt gehäuft beim Turnersyndrom und der Trisomie 13 vor [13, Seite 212]. Die Unterbrechung des Aortenbogens ist ein sehr seltener Herzfehler, welcher postnatal gut operiert werden kann, aber auch gehäuft bei einer Mikrodeletion 22q11 vorkommt. Der doppelte Aortenbogen ist ebenfalls ein seltener Herzfehler, der bei Neugeborenen durch Einengung von Trachea und Ösophagus zu Atem- und Schluckstörungen führen kann und daher ebenfalls in den ersten Lebensmonaten operiert werden sollte. Die abberante rechte Arteria subclavia ist die häufigste der oben genannten Anomalien und kann auch als Normvariante angesehen werden [13, Seite 217]. Sie ist allerdings auch ein

Softmarker für Chromosomenstörungen wie die Trisomie 21 und die Mikrodeletion 22q11 [28], daher sollte eventuell weiterführende Diagnostik in Betracht gezogen werden.

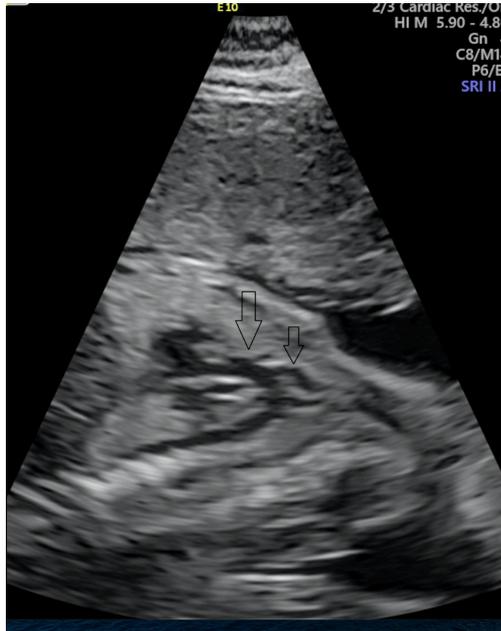


Abbildung 1: Darstellung des Aortenbogens mit den drei abgehenden Gefäßen im Längsschnitt. Der größere Pfeil zeigt auf die Aorta, der kleinere Pfeil auf den Truncus brachiocephalicus, dieser geht als erstes aus dem Aortenbogen ab.

3.5 Matrixsonde und künstliche Intelligenz sowie deren Anwendung in Bezug auf die fetale Biometrie und die Darstellung des Aortenbogens

3.5.1 Elektronische Matrixsonde

Die Firma GE ² entwickelte eine 3D/4D-Matrixsonde. Diese kann mit dem Ultraschallgerät Voluson E10 genutzt werden und wird in der Praxis für Pränataldiagnostik in der Friedrichstraße routinemäßig eingesetzt. Die Sonde wird eM6c genannt. Sie verfügt über eine sehr gute Eindringtiefe in das Gewebe und kann daher auch bei adipösen Patientinnen eingesetzt werden. Des Weiteren können Biplane Aufnahmen in Echtzeit erfolgen. Die Biplane Technik wird gesondert erläutert. Einige Informationen des Herstellers über die Sonde sind in Tabelle 2 zusammengefasst [29]. Der Matrixschallkopf ermöglicht als

²General Electric Company

3D/4D-Sonde Volumenaufnahmen.

Eindringtiefe	26 cm
Frequenzbereich	1-6 MHz
Anzahl der Elemente	8192

Tabelle 2: Angaben zur elektronischen Matrixsonde

3.5.2 Automatisierte Messung des fetalen Kopfes mittels künstlicher Intelligenz

Durch künstliche Intelligenz ist es mittlerweile möglich, Biometriedaten des Feten vom Computer ausmessen zu lassen. Das benötigte Computerprogramm von GE heißt „SonoCNS“, zur Nutzung wird eine 3D-fähige Ultraschallsonde benötigt. Eine genaue Beschreibung der Technik erfolgt im Abschnitt Material und Methoden. Durch die Messung der fetalen Parameter mit der künstlichen Intelligenz kann Zeit eingespart werden, da deutlich weniger Tasten bedient werden müssen als beim konventionellen Verfahren. Ein Vorteil des SonoCNS-Programms ist außerdem, dass auch nachträglich aus bereits aufgenommenen Volumenaufnahmen Messungen erfolgen können. Dies ermöglicht also auch Messungen, wenn die Patientin schon die Praxis verlassen hat. Das SonoCNS-Programm ist nur eine Form der künstlichen Intelligenz. Ein prominentes Beispiel für künstliche Intelligenz im pränatalen Ultraschall ist die Dopplersonographie. Zunächst musste nachdem der Blutfluss in einem fetalen Gefäß dargestellt war eine Hüllkurve manuell über das Flussmuster gelegt werden. Mittlerweile legt der Computer nach Betätigung entsprechender Tasten eigenständig die Hüllkurve darüber und berechnet die zugehörigen Werte anhand eingespeicherter Formeln. Des Weiteren wird die künstliche Intelligenz auch im Rahmen des Ersttrimesterscreenings zur Messung der Nackentransparenz genutzt. Die Nackentransparenz ist ein Parameter zur Detektion von fetalen Anomalien. Es wird die Größe eines nuchalen Ödems in Millimetern gemessen. Diese Messung kann sowohl manuell als auch automatisiert erfolgen. Bei der automatisierten Messung wird ein Fenster über die zu untersuchende Region gelegt und der Computer generiert nun automatisch Messdaten für die fetale Nackentransparenz, für die gezeigt werden konnte, dass die Intra- und Interobserver-Variabilität reproduzierbar ist [30]. Weitere Studien zeigten, dass durch die automatische Messung der Nackentransparenz eine Standardisierung erreicht werden

kann [31]. Somit hat auch die Nutzung der künstlichen Intelligenz im Bereich des fetalen Kopfes das Potential, Untersuchungen zu verbessern.

3.5.3 Biplane Technik

Diese Technik ist von der elektronischen Matrixsonde abhängig. Hier wird die Anordnung der Kristalle in der Sonde ausgenutzt. In einer herkömmlichen 3D-Sonde befinden sich circa 120-190 Kristalle in einer einzelnen Reihe, in die Matrixsonde werden 64 Reihen mit je 128 Kristallen eingebaut [32]. So können Ultraschallwellen in allen Reihen in Echtzeit gesendet werden und es kann ein orthogonal zum einen Bild stehendes zweites Original-Ultraschallbild erstellt werden [32]. Es handelt sich also nicht um ein errechnetes Bild. Die Technik kann sowohl in 2D als auch mit Farbdoppler angewandt werden. Auf dem Bildschirm werden zwei Bilder angezeigt, im linken Bild ist jeweils das vom Untersucher eingestellte Bild meist im Querschnitt dargestellt mit der über die interessierende Region gelegten Linie, im rechten Bild ist das orthogonal zur Linie im linken Bild stehende Bild im Längsschnitt dargestellt. Dies erlaubt eine genauere Darstellung von Strukturen, da durch die sonst notwendige Bewegung des Schallkopfes Ungenauigkeiten entstehen können und eventuell die gewünschte Ebene verloren wird. Die Biplane Technik bietet also die Möglichkeit, Strukturen standardisiert und reproduzierbarer zu messen. Diese Technik kann vielfältig eingesetzt werden, und möglicherweise auch fetalchirurgische Eingriffen wie die intrauterine Pleurapunktion oder Chordozentese sicherer machen [33]. Die Biplane Darstellung kann auch zur Untersuchung anderer fetaler Strukturen herangezogen werden und nützlich sein, ein Beispiel ist die sichere Darstellung der Augen. Erstellt man einen Sagittalschnitt mittig durch das fetalen Gesicht, eine sogenannte Profilaufnahme, kann der Biplane Modus oberhalb der Nase im Bereich der Augen gesetzt werden. Nun wird ein orthogonales Bild in Echtzeit aufgenommen, und es werden bei regelrechter Anatomie zwei Augenhöhlen mit zugehöriger Linse dargestellt. So können Fehlbildungen in diesem Bereich einfach und standardisiert ausgeschlossen werden. Ein Beispiel ist in Abbildung 2 dargestellt. Eine weitere Möglichkeit ebenfalls ausgehend vom fetalen Profil, ist die Biplane Darstellung des Mundbereiches um eine Lippen-Kiefer-Gaumenspalte zu detektieren oder auszuschließen. Ein weiteres Beispiel ist die Darstellung der Wirbelsäule, hier kann eine Biplane Aufnahme den Höhenstand einer etwaigen Spina bifida aufzeigen. In der vorliegenden Studie wird die Bedeutung der Biplanen Darstellung in Bezug auf den

Aortenbogen untersucht. Zunächst wird die Ebene des sogenannten Drei-Gefäß-Trachea-Blickes aufgesucht, eine Ebene welche zur Standardeinstellung in der Feindiagnostik im zweiten Trimenon zählt. Hierbei wird eine transversale, leicht schräge Ebene kranial des Herzens aufgesucht. So werden die Pulmonalarterie, die Aorta und die obere Hohlvene dargestellt sowie die Trachea. Bei unauffälligem Drei-Gefäß-Blick ist davon auszugehen, dass keine schwerwiegende Fehlbildung wie zum Beispiel die Transposition der großen Arterien vorliegt, welche direkt nach Geburt eine Behandlung und Operation im Herzzentrum unabdingbar macht. Wenn man nun auf die fetale Aorta im 3-Gefäß-Blick den Biplane Modus legt, zeigt das rechte Ultraschallbild die Struktur im rechten Winkel dazu, in diesem Fall ist das der fetale Aortenbogen im Längsschnitt und gelegentlich auch die drei aus diesem Bogen abgehenden Gefäße. Der Biplane Modus kann so die Untersuchung des fetalen Herzens und der Aorta erleichtern, auch bei etwaigen Herzvitien könnten diese Aufnahmen hilfreich sein.

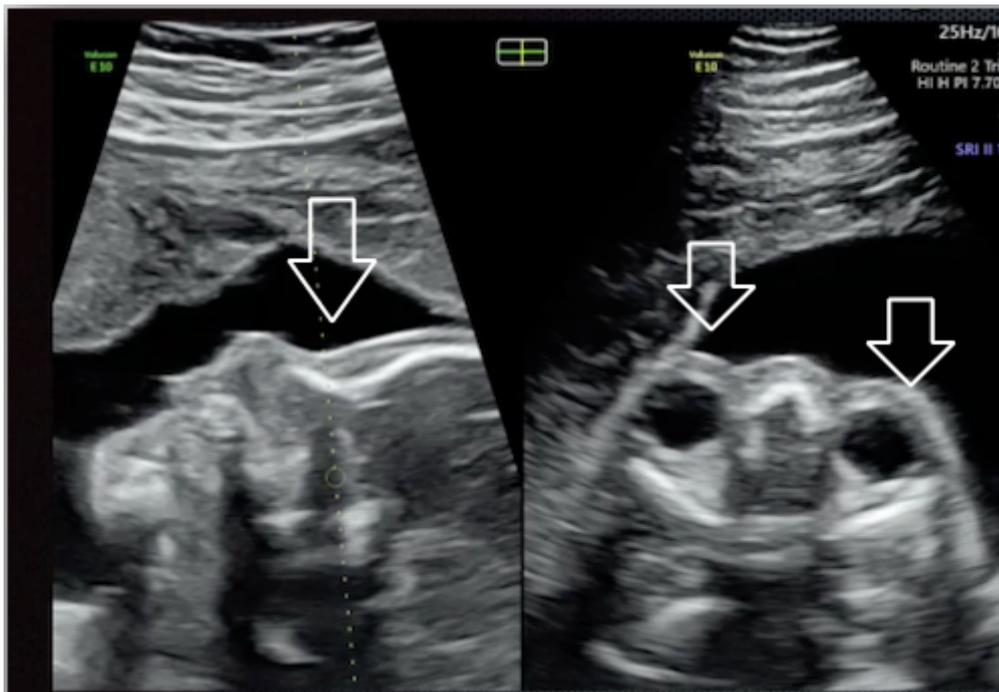


Abbildung 2: Darstellung von zwei vorhandenen Augen im rechten Bild (markiert mit den weißen Pfeilen) ausgehend von einer Profilaufnahme mittels Biplaner Darstellung im linken Bild. Der Pfeil im linken Bild zeigt auf das fetale Nasenbein.

3.6 Zielsetzung

Künstliche Intelligenz ist Teil unseres Alltags geworden. Im Alltag nutzen wir sie zum Beispiel bei der Smartphone-Gesichtserkennung oder der Google-Suchmaschine. Allerdings sind noch viele Schwachstellen vorhanden, zum Beispiel die Frage nach rechtlicher Haftung bei Fehlern, die sich aus der Nutzung ergeben [34]. Auch in der Medizin werden stetig neue Programme und Technologien entwickelt, die das Potential haben, Arbeitsabläufe zu erleichtern und eine bessere Behandlung zu ermöglichen. Ziel der Arbeit ist der Vergleich der computergestützten Messungen mithilfe der künstlichen Intelligenz mit der herkömmlichen manuellen Messung des menschlichen Untersuchers. Die fetale Biometrie ist zwar etabliert und gehört zum Goldstandard der fetalen Beurteilung, bisher wurden aber automatisierte Messungen und manuelle Messungen noch keinem direkten Vergleich unterzogen. Bis dato ist es in den Krankenhäusern und Praxen Standard, für die fünf in dieser Studie untersuchten Parameter des fetalen Kopfes insgesamt drei Bildaufnahmen des Ungeborenen in unterschiedlichen Ebenen zu erstellen und manuell auszumessen. Durch die künstliche Intelligenz reicht die Darstellung einer Ebene, die Volumenaufnahme des Kopfes und die Messungen werden dann automatisch durchgeführt. Falls die Messungen ähnlich genaue Messdaten zeigen sollten, wäre die Darstellung mittels künstlicher Intelligenz reproduzierbarer für weitere Ärztinnen und Ärzte, hilfreich für etwaige Nachkontrollen, sowie retrospektiv beurteilbar, da anhand eines einmal aufgenommen Volumens auch Nachuntersuchungen ohne Anwesenheit der Patientin erfolgen können. Messungen könnten standardisierter erfolgen und die bisherige Messungengenauigkeit der menschlichen Untersucher eventuell verringert werden. Ultraschalluntersuchungen hängen auch stets von der Erfahrung und vom Können der Anwender und Anwenderinnen ab. Sowohl die künstliche Intelligenz als auch die Biplane Darstellung haben das Potential, Untersuchungen zu vereinfachen und die Intra- und Interobserver-Variabilität zu senken [35]. So könnten in Zukunft auch weniger erfahrene Gynäkologinnen und Gynäkologen mit höherer Wahrscheinlichkeit auch komplexe Pathologien erkennen und Mutter und Kind einer adäquaten Weiterbetreuung zuführen. Eventuell könnte so die kindliche Morbidität noch weiter gesenkt werden sowie die Rate an pränatal diagnostizierten Anomalien erhöht werden. Des Weiteren könnte die Nutzung der künstlichen Intelligenz mit einer Reduktion der

Untersuchungsdauer verbunden sein. Die Organisation EUROCAT³ ist ein europaweites Register für kongenitale Fehlbildungen und gibt in regelmäßigen Berichten auch deren pränatale Detektionsrate an. Für Sachsen-Anhalt wird in den Jahren 2015-2019 nur eine Detektionsrate von 36,41% für alle Anomalien angegeben [36]. Es ist mehr als wünschenswert, diese Rate zu erhöhen, wobei Programme der künstlichen Intelligenz hilfreich sein können. Die in der Studie genutzte Matrixsonde verfügt außerdem über eine sehr gute Eindringtiefe ins Gewebe, dies könnte nützlich sein in Anbetracht der Tatsache, dass es eine steigende Anzahl an adipösen Patienten gibt.

³European network of population-based registries for the epidemiological surveillance of congenital anomalies

4 Material und Methoden

4.1 Studienaufbau

Es handelt sich um eine retrospektive Kohortenstudie. Das Patientenkollektiv umfasst alle Schwangeren, die im Zeitraum von Februar 2020 bis Juni 2020 zur Feindiagnostik in der Praxis Friedrichstraße bei PD Dr. Heling (Untersucher 1) im zweiten Trimenon vorstellig wurden, sowie eine Vergleichsgruppe der Patientinnen von Prof. Chaoui (Untersucher 2) im Februar 2021. Die früheste Feindiagnostik wurde in 20+0 SSW durchgeführt, die späteste in 26+2 SSW. Es handelte sich um ein Standardkollektiv mit gesunden Schwangeren. Beide Untersucher verfügen über langjährige Ultraschallerfahrung und sind DEGUM III zertifiziert ⁴[37]. Ausgeschlossen wurden Feten mit Fehlbildungen, Mehrlinge sowie minderjährige Patientinnen. Die Altersspanne der Frauen reicht von 20 Jahre bis 45 Jahre (Mittelwert 32,76 Jahre, Median 33 Jahre) der BMI der Frauen lag zwischen 17,2 und 59,9 (10,19% waren übergewichtig mit einem BMI ab 30). Im Median wurde das Gewicht der untersuchten Feten auf 525 Gramm geschätzt. Außerdem wurde die intrauterine Lage der Feten (Schädellage, Beckenendlage, Querlage), das Vorliegen einer dorsoanterioren Stellung, Lage der Plazenta, Fruchtwassermenge, allgemeine Schwangerschaftsrisiken sowie die Gravidität und Parität erfasst. Es erfolgte für die Anwendung der künstlichen Intelligenz eine generelle Einteilung in technisch möglich sowie technisch nicht möglich. Für die Anwendung der Biplane Technik wurden drei Untergruppen gebildet (Bilder vorhanden und darstellbar, Bilder vorhanden und nicht darstellbar, keine Bilder). Insgesamt wurden 393 Ultraschalluntersuchungen des Untersuchers 1 und 20 Untersuchungen des Untersuchers 2 ausgewertet in Bezug auf die fetale Biometrie und 392 Bilder des Untersuchers 1 in Bezug auf den fetalen Aortenbogen. Der Untersucher 2 reichte keine Bilder des fetalen Aortenbogens ein. Die Bilder und Messwerte wurden wie in der Praxis zur Dokumentation üblich im Viewpoint-Programm gespeichert, einer Software speziell zur Archivierung von Ultraschalluntersuchungen. Dieses Programm ist mit dem Ultraschallgerät verknüpft und stellt die erhobenen Daten grafisch dar. Die Daten der eingeschlossenen Patientinnen wurden über das Viewpoint-Programm in der Praxis abgerufen und anonymisiert in Excel-Tabellen erfasst. Alle Messdaten werden im Rahmen dieses Screenings ohnehin erhoben, es entstanden keine zusätzlichen Untersuchungen für die Schwangere

⁴Höchstqualifizierte Sonographie in entsprechendem Fachgebiet

oder den Feten. Alle teilnehmenden Frauen stimmten im Vorfeld der anonymisierten Datennutzung im Rahmen der medizinischen Forschung zu. Die Datenschutzvorgaben und die Vorgaben der Charité zur guten wissenschaftlichen Praxis wurden beachtet. Es liegt ein positives Votum der Ethikkommission der Charité vor.

Parameter	Range
Schwangerschaftsalter	20+0 SSW-26+2 SSW
Alter der Frauen	20-45 Jahre, Median 33 Jahre
BMI	17,2-59,9, Median 23,7, Mittelwert 24,5
Gewicht des Kindes	Median 525g
Fruchtwassermenge	bei allen unauffällig
Intrauterine Lage	SL, BEL, QL
Stellung Dorsoanterior	Ja in 56 Fällen, Nein in 327 Fällen
Nulliparae	210
Multiparae	169 , 5 Angaben fehlten

Tabelle 3: Übersicht über das Patientinnenkollektiv

4.2 Ultraschallgerät

Verwendet wurde das Ultraschallgerät Voluson E10 von GE mit der eM6C-Sonde für 3D/4D Bilder. Die Ultraschallsonde wird Matrixsonde genannt. Es handelt sich um einen Schallkopf in konvexer Bauweise. Die Matrixsonde verfügt über den neuen Biplane Modus. Eine weitere Computersoftware namens SonoCNS kann die automatisierten Messungen durchführen. Laut Hersteller ist dieses Ultraschallgerät speziell für die Feindiagnostik in Geburtshilfe und Gynäkologie entwickelt worden [38].

4.3 Durchführung der Ultraschalluntersuchung

Zunächst muss der Name der Patientin und das Schwangerschaftsalter am Computer eingegeben und mit dem Ultraschallgerät verknüpft werden, dies erfolgt in der Praxis Friedrichstraße mit dem oben erwähnten Programm Viewpoint. Dann wird die Ultraschallsonde am Gerät angewählt, in diesem Fall die Matrixsonde. Der Ablauf einer Ultraschalluntersuchung im Allgemeinen kann zwischen Untersucherinnen und Untersuchern leicht

divergieren. Generell startet die Untersuchung mit der Kontrolle der fetalen Herzaktion und der intrauterinen Lage. Im Anschluss wird nun bei der Feindiagnostik die komplette Sonoanatomie des Feten dargestellt und untersucht sowie routinemäßig circa 90-130 Bilder zur Dokumentation gespeichert. Bei der Ultraschalluntersuchung haben erfahrene Ärztinnen und Ärzte ein festgelegtes Untersuchungsschema um die zu untersuchenden Regionen nacheinander abzuarbeiten. Ausnahme bildet die Untersuchung des fetalen Herzens, die gegebenenfalls im Rahmen der Untersuchung vorgezogen oder nach hinten verschoben wird da sie stark von der fetalen Lage abhängt. Liegt der Fet zum Beispiel dorsoanterior, also mit dem Rücken zur Uterusvorderwand, kann das Herz schwer dargestellt werden durch den Schallschatten der knöchernen Strukturen. Es kommt häufig zur spontanen Drehung des Feten in eine günstigere Lage, sodass zunächst andere Strukturen untersucht werden können. Die Feindiagnostik dauert im Schnitt in der Praxis Friedrichstraße circa 20-30 Minuten, bei fetalen Auffälligkeiten oder eingeschränkten Untersuchungsbedingungen, zum Beispiel bei adipösen Frauen, auch länger. Im Rahmen der vorliegenden Studie erfolgt zusätzlich zur manuellen standardmäßig durchgeführten Biometrie-Messung des fetalen Kopfes auch eine Volumenaufnahme des fetalen Kopfes in 3D um die automatisierte Messung mittels künstlicher Intelligenz zu ermöglichen. Außerdem erfolgte eine zusätzliche Bilddokumentation der fetalen Aorta mit der Biplanen Technik. Zum Ende der Untersuchung werden eventuelle Auffälligkeiten mit der Patientin besprochen sowie weitere notwendige Folgetermine vereinbart. Durch die Anwendung der Biplane Technik und der automatisierten Messung ergaben sich im vorliegenden Patientinnenkollektiv keine Indikationen zur weiterführenden Diagnostik oder Untersuchungen.

4.4 Definition der Messebenen und manuelle Messung

Die unten genannten Messebenen sind Standardeinstellungen, welche zur fetalen Gewichtsschätzung benötigt werden. Das Aufsuchen und die korrekte Einstellung dieser Messebenen werden angehenden Gynäkologinnen und Gynäkologen stets beigebracht, die genaue Beschreibung der Messebenen wurden aus dem Buch Ultraschalldiagnostik in Geburtshilfe und Gynäkologie von Gembruch, Hecher und Steiner übernommen[13, Seite 18-20]. Für die Messung der fetale Biometrie ist unter anderem das Aufsuchen folgender Messebenen notwendig:

Transthalamische Ebene Es erfolgt ein transversaler Schnitt durch den fetalen Kopf mit Darstellung der Thalamuskern und des Cavum septum pellucidum für die Messung des biparietalen Durchmessers (BPD) und des Kopfumfangs (KU). Die Falx cerebri darf im frontalen Abschnitt nicht durchgehend erkennbar sein [13]. Der BPD wird an der breitesten Stelle, orthogonal zur Mittellinie vom äußeren Rand des proximalen knöchernen Schädels bis zum äußeren Rand des distalen knöchernen Schädels gemessen [13]. Für den Kopfumfang wird um den knöchernen Schädel eine Ellipsenform gelegt. Die Ellipsenform soll nur den knöchernen Teil des Schädels umfassen und nicht die Kopfhaut [13]. Alternativ kann der Computer den Kopfumfang über eine Formel unter Einbeziehung des biparietalen Durchmessers und des frontookzipitalen Durchmessers (FOD) errechnen. Im Beispielbild wurde der FOD gemessen, die Künstliche Intelligenz hingegen benutzt den Kopfumfang. Der FOD ist ein Längsdurchmesser von der Stirn bis zum Hinterhaupt. Die transthalamische Ebene und eine beispielhafte Messung sind in Abbildung 3 dargestellt.

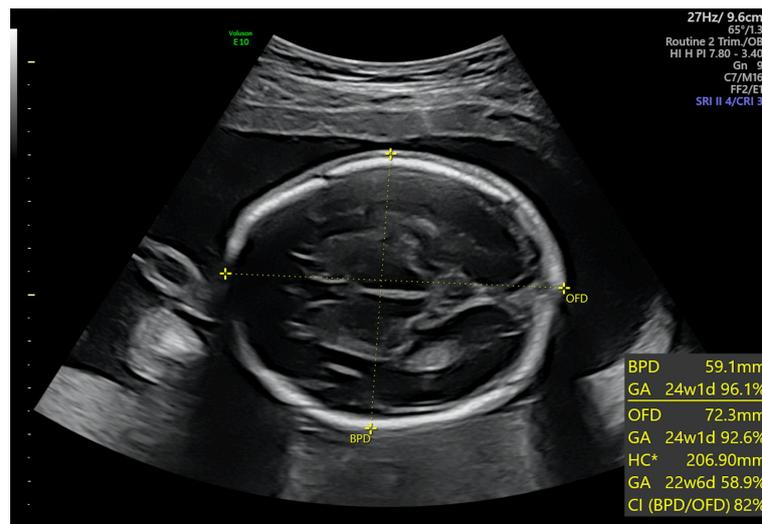


Abbildung 3: Manuelle Messung des biparietalen Durchmessers und des frontookzipitalen Durchmessers, im rechten unteren Bildrand hat der Computer mit beiden Messungen den Kopfumfang (HC) errechnet.

Transzerebelläre Ebene Ein transversaler Schnitt durch das Kleinhirn mit Cisterna magna. Das Kleinhirn wird in der größten Ausdehnung gemessen, die Cisterna magna ebenfalls [13]. Ein Beispiel ist in Abbildung 4 dargestellt.

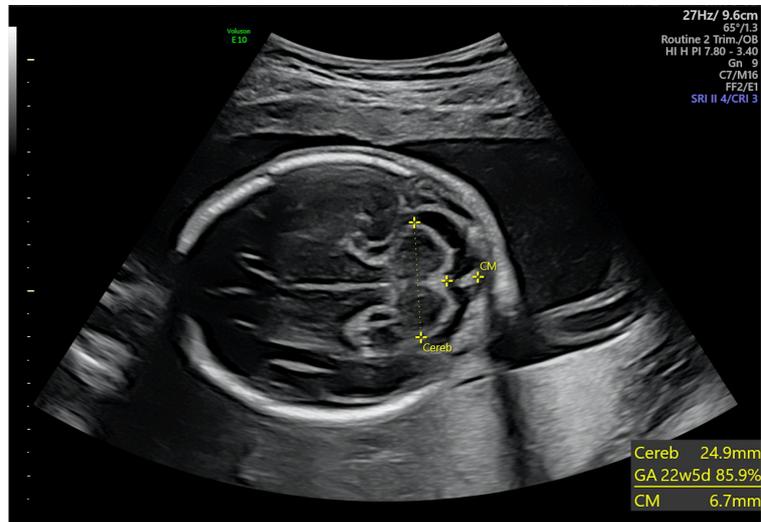


Abbildung 4: Manuelle Messung von Cisterna magna und Cerebellum in der transzerebellären Ebene.

Transventrikuläre Ebene Ein transversaler Schnitt durch die Ventrikel, kaudal und parallel zur transthalamischen Ebene [13]. Abbildung 5 zeigt wiederum ein Beispiel.



Abbildung 5: Manuelle Messung eines Ventrikels, die Ebene befindet sich kaudal und parallel zur transthalamischen Ebene.

Auf die Beschreibung der anderen für die Gewichtsschätzung notwendigen Ebenen wird hier verzichtet, da sie für diese Studie keine Bedeutung hatten.

4.5 Messung der gesuchten Parameter mit Hilfe der künstlichen Intelligenz

Die Messung wird ebenfalls mit der elektronischen Matrixsonde gemacht, ist allerdings nicht von dieser abhängig, sondern kann theoretisch mit jeder 3D-fähigen Sonde aufgenommen werden. Zur Messung wird eine 3D-Volumenaufnahme des fetalen Kopfes benötigt. Hierzu wird ebenfalls wie zur manuellen Messung von BPD und KU die transthalamische Ebene dargestellt. Nun wird die 3D-Taste betätigt und am Bildschirm das vorinstallierte Programm „SonoCNS“ ausgewählt. Nun nimmt das Programm eine Volumenaufnahme des fetalen Kopfes auf, dargestellt in Abbildung 6. So werden automatisch vier Bilder angezeigt, zunächst im linken oberen Bildrand ein Sagittalschnitt des Kopfes. Auf diesem Bild werden die drei Messebenen, die der Computer anhand der Volumenaufnahme erstellt, mit farbigen Linien dargestellt. Die diagonale, blaue Linie „TC“ steht für transzerebelläre Ebene, diese wird im linken unteren Bildrand gezeigt. Die gelbe, waa-

gerechte Linie wird mit „TV“ für transventrikulär bezeichnet und wird im rechten oberen Bild dargestellt. Die untere pinke waagerechte Linie wird als „TT“ bezeichnet und steht für die transthalämische Ebene, welche im rechten unteren Bildrand dargestellt wird. Nun erkennt das SonoCNS-Programm die gesuchten fetalen Strukturen und misst diese automatisch im nächsten Schritt aus. Die Ergebnisse werden in tabellarischer Form im rechten unteren Bildrand angezeigt, siehe Abbildung 7a. So entstehen aus der Darstellung einer einzigen Ebene, der transthalämischen, drei multiplanare Bilder. Die manuell benötigte Einzeldarstellung der drei Messebenen entfällt.

4.6 Vergleich beider Messmethoden

Um die Ähnlichkeit der manuell gemessenen Werte mit den Werten der künstlichen Intelligenz beweisen oder widerlegen zu können, wurde im Rahmen der Vorbereitung zur Studie eine Variationsbreite der Messwerte festgelegt. Innerhalb dieser Variationsbreite erscheint die Abweichung der Werte zwischen manueller Messung und künstlicher Intelligenz tolerabel. Die festgelegte maximale Differenz für die jeweiligen Parameter zeigt Tabelle 4. Diese Abweichung wurde gewählt, da auch zwei erfahrene menschliche Untersucher bei der Untersuchung desselben Feten geringe Messdifferenzen aufweisen können. Dies zeigt auch eine Studie von Sarris et al., die die Interobserver-Variabilität für den Parameter Kopfumfang mit ± 9.8 mm angibt [39]. Die tolerable Abweichung wurde von einem erfahrenen Pränataldiagnostiker festgelegt. Die Werte sollten ähnlich sein, müssen aber zum Beweis der These, dass der Computer äquivalent zum erfahrenen Untersucher misst, nicht gleich sein.

Parameter	Tolerable Abweichung
BPD	$\leq 4,5$ mm
KU	≤ 10 mm
VP	≤ 2 mm
CM	≤ 2 mm
Cereb	$\leq 3,5$ mm

Tabelle 4: Darstellung der tolerierten Abweichung der Messergebnisse für den Untersucher und die künstliche Intelligenz in Millimetern pro Parameter.

4.7 Darstellung des fetalen Aortenbogens mit der Biplane Technik

Für den zweiten Teil der Studie wird die Biplane Technik der eM6C-Sonde genutzt. Hierzu wird ein transversaler Schnitt durch den fetalen Thorax eingestellt, es handelt sich hier um die Ebene des sogenannten 3-Gefäß-Thymus-Blickes. Nun wird am Display „Biplan“ ausgewählt, und man muss eine vertikale Linie in die zu untersuchende Struktur, die sogenannte „Region of interest“ legen, in dem Fall der angeschnittenen Aortenbogen. Der Computer gibt nun in Echtzeit ein Bild orthogonal zu dieser Region of interest an. Orthogonal zum Aortenbogen im Querschnitt zeigt sich der Aortenbogen im Längsschnitt. Diese Dokumentation ist also an die Echtzeit gebunden und kann nicht retrospektiv erfolgen.

Die Beurteilung der so erhobenen Bildern erfolgt visuell nach folgenden Kriterien:

1. Darstellbarkeit des Aortenbogens
2. Darstellbarkeit der Abgangsgefäße (Truncus brachiocephalicus, linke A. carotis communis, linke A. subclavia)
3. Messungen möglich (bei Darstellbarkeit immer ja)
4. Technisch nicht möglich (keine Bilder vorhanden)

Ein mögliches Ergebnis ist, dass Bilder vorhanden sind und der Aortenbogen sowie gegebenenfalls, aber nicht zwingend, die Abgangsgefäße dargestellt werden können. Ein zweites mögliches Ergebnis lautet, dass zwar Bilder mit der Biplane Technik erstellt wurden, aber die gesuchten Strukturen nicht dargestellt werden können. Ein weiteres Ergebnis ist, dass überhaupt keine Bilder erstellt werden können, es also technisch nicht möglich ist bei diesem Feten die Biplane Technik im Bereich der Aorta anzuwenden.

4.8 Bildmaterial

Alle in der vorliegenden Arbeit verwendeten Ultraschallbilder wurden in der Praxis Friedrichstraße erstellt und wurden zur anonymisierten Veröffentlichung im Rahmen dieser Dissertation von dem zuständigen Untersucher freigegeben.

4.8.1 Fetale Biometrie

In Abbildung 6 ist die Volumenaufnahme des Kopfes dargestellt. Die Aufnahme erfolgte wie oben beschrieben.

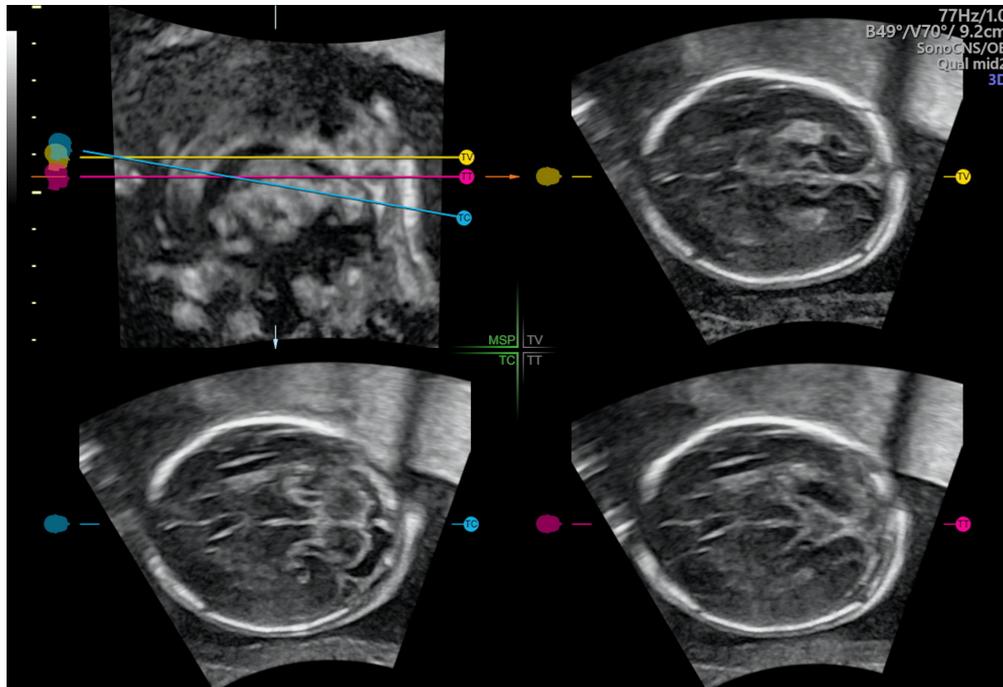
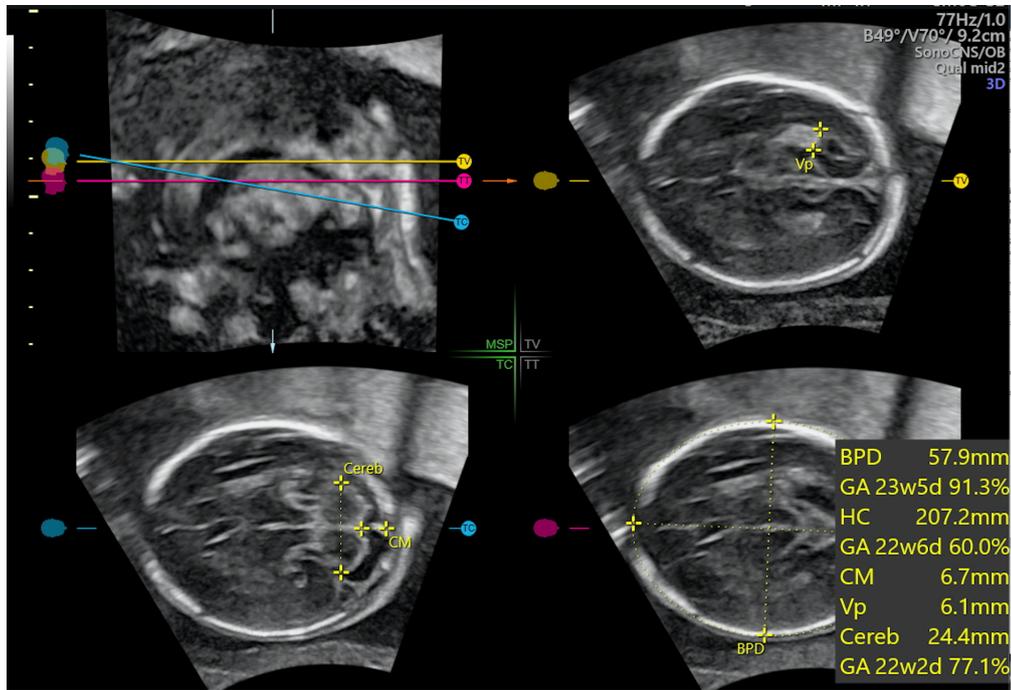


Abbildung 6: Volumenaufnahme des Kopfes zur Darstellung aller Messebenen, noch ohne Messwerte, Aufnahme in der 22+2 SSW. Im linken oberen Bild ist ein Sagittalschnitt des Kopfes dargestellt, die farbigen Linien zeigen die Messebenen an. Rechts oben ist die transventrikuläre Ebene „TV“, rechts unten die transthalämische Ebene „TT“, links unten die transzerebelläre Ebene „TC“.

Abbildung 7a zeigt Abbildung 6 mit den von der künstlichen Intelligenz gemessenen Messwerten.

Die Werte dieses Feten wurden wie bei allen anderen zusätzlich zur automatischen Messung vom Untersucher auch händisch ausgemessen, hier zeigt sich eine sehr geringe Differenz, welche im in Tabelle 4 vorgegebenen Rahmen liegt und in Abbildung 7b beispielhaft dargestellt ist. Der BMI dieser Patientin betrug 20,7, das Kind lag in Beckenendlage und die Plazenta befand sich an der Hinterwand.



(a) Es handelt sich um Abbildung 6 aber mit den von der künstlichen Intelligenz gemessenen Werten, welche im rechten unteren Bildrand tabellarisch aufgelistet werden. In den Bildern selbst zeigt sich die Calipersetzung der künstlichen Intelligenz.

Messebenen	Differenz in mm
BPD	1,2
KU	-0,3
VP	-0,9
CM	0
Cereb	0,7

(b) Differenz der Messung der künstlichen Intelligenz und des menschlichen Untersuchers des in Abbildung 7a dargestellten Beispiels. Es zeigen sich für alle Parameter nur geringe Messunterschiede.

Abbildung 7: Differenz der Messung der KI und des U1 am dargestellten Beispiel

4.8.2 Fetaler Aortenbogen

In Abbildung 8 wird ein Beispielbild der Aufnahme der fetalen Aorta mittels Biplane Modus dargestellt. Der Untersucher stellt hierfür wie oben beschrieben einen Querschnitt durch den fetalen Brustkorb dar und sucht die Ebene kranial des fetalen Herzens auf. Dort befindet sich die Aorta. Nun wird mit der vertikalen Linie durch die Aorta der Biplane Modus angewandt, dies ist im linken Bild dargestellt. Das rechte Bild zeigt das in Echtzeit aufgenommene orthogonale Bild dazu, die Aorta im Längsschnitt. Hier ist der Aortenbogen mit den drei Abgangsgefäßen dargestellt.

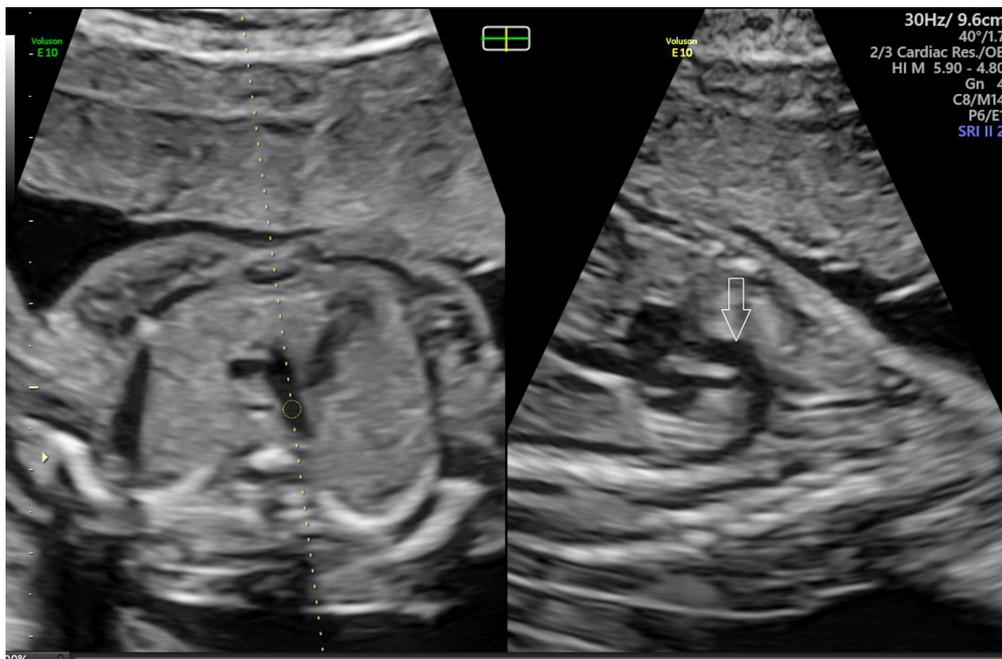


Abbildung 8: Biplane Modus mittels Matrixsonde. Das linke Bild wurde vom Untersucher eingestellt sowie eine Linie über die Region of Interest gelegt. Das rechte Bild ist das orthogonal dazu stehende originale Ultraschallbild, welches mit dem Biplane Modus aufgenommen wurde und den Aortenbogen im Längsschnitt zeigt. Der Pfeil im rechten Bild zeigt auf die Stelle des Aortenbogens, an der sich im linken Bild der Kreis befindet.

4.9 Statistische Analyse

Die statistische Auswertung der Daten erfolgte mittels R und Python (SciPy.stats Paket). Es wurden Bland-Altman-Plots für den Vergleich der manuellen Messung mit der automatischen Messung erstellt. Univariate lineare Regressionen und multiple Regressionen wurden zur Analyse von potentiellen Störfaktoren genutzt. Mittels Kerndichteschätzer wurde der BMI in Relation zur Darstellungsmöglichkeit der Aorta gesetzt. Der Chi-Quadrat-Unabhängigkeitstest erfolgte zur Testung, ob die fetale Lage Einfluss auf die Anwendbarkeit des Biplane Modus hat. Generell gilt die statistische Signifikanz bei einem p-Wert <0.05 .

4.9.1 Begründung der Fallzahl

Wie bereits in Abschnitt Kapitel 4.1 erwähnt, beträgt die Fallzahl des Untersuchers 1 393, während die des Untersuchers 2 20 beträgt. Laut des Papers „Sample Size for Assessing Agreement between Two Methods of Measurement by Bland-Altman Method“[40] wird, bei einem wahren Unterschied von 0 Standardabweichungen (also perfekter Übereinstimmung der beiden Methoden), 20% Typ II Fehlerwahrscheinlichkeit (80% Power), 5% Typ I Fehlerwahrscheinlichkeit und Agreement Limit von 2.3 Standardabweichungen, eine Fallzahl von 271 benötigt. Die Fallzahl von U1 kann unter diesen Voraussetzungen also als ausreichend angesehen werden. Die Fallzahl von U2 ist zu klein um daraus statistisch relevante Schlüsse zu ziehen. Sie dient lediglich als Anhaltspunkt, dass die KI auch durch weitere menschliche Untersucher sinnvoll genutzt werden kann.

5 Ergebnisse

5.1 Fetale Biometrie

Insgesamt wurden die Daten und Bilder von 393 Patientinnen des Untersuchers 1 ausgewertet und beurteilt sowie die Daten von 20 Patientinnen des Untersuchers 2. Von den genannten 393 Patientinnen des Untersuchers 1 wurden 9 Frauen aus der statistischen Analyse ausgeschlossen bei denen die Künstliche Intelligenz offensichtlich die falschen Messebenen ausgemessen hat, vergleiche hierzu Abbildung 21 und Abschnitt 5.1.6. Diese falschen Messebenen sind auch für weniger geübte Untersucherinnen und Untersucher sofort erkennbar. Die Patientinnen wurden in der Excel Tabelle als fehlerhaft markiert. Es muss davon ausgegangen werden, dass die Hinzunahme dieser Daten das Ergebnis verfälscht hätte, da als primärer Studienendpunkt der Vergleich von korrekt erfolgten Messungen gewählt wurde und nicht die technische Möglichkeit der Durchführung. Man könnte diese Patientinnen auch theoretisch zur Untergruppe „technisch nicht möglich“ zählen, zu dieser Gruppe wurden allerdings nur die 15 Patientinnen gezählt, bei denen das Computerprogramm überhaupt keine Volumenaufnahme aufgenommen hat, eine Übersicht zeigt Tabelle 5. Insgesamt wurden somit die Daten von 384 Frauen analysiert und auf Ähnlichkeit und Abweichungen überprüft. Die ausgeschlossenen Patientinnen fließen in die Fehlerwertanalyse mit ein. Die von der künstlichen Intelligenz gemessenen Werte liegen etwas unter der Anzahl von 384, da auch wenn ein Volumen aufgenommen wurde, der Computer in einigen Fällen nicht für jeden Parameter Messungen angezeigt hat. Eine Übersicht zeigt hier Tabelle 6. Auffällig ist, dass es der künstlichen Intelligenz trotz korrekt aufgenommenem Volumen und Messung der anderen Parameter in 18 Fällen nicht gelingt den hinteren Hirnseitenventrikel (VP) auszumessen. In 4 dieser 18 Fälle gilt die Patientin mit einem BMI von über 30 als adipös. 8 der Feten lagen in Schädellage, 9 in BEL und einer in QL, kein Fet war in dorsoanteriorer Stellung. Weitere Ausführungen hierzu finden sich in Abschnitt 6.1.1.

Untersucher	Untersuchungsanzahl	Technisch möglich	Technisch nicht möglich	Ausgeschlossene Patientinnen
Untersucher 1	393	96% (378)	4% (15)	2% (9)
Untersucher 2	20	100% (20)	0% (0)	0%(0)

Tabelle 5: Übersicht über die Anzahl der erfolgten Untersuchungen in Bezug auf die fetale Biometrie

Parameter	Fehlende Messungen
BPD	2
KU	2
VP	18
CM	0
Cereb	0

Tabelle 6: Anzahl der nicht von der KI gemessenen Werte trotz aufgenommenem Volumen und Klassifikation als „Technisch möglich“, aufgeteilt nach einzelnen Parametern. Die meisten fehlenden Messungen waren beim Parameter „VP“, also dem hinteren Hirnseitenventrikel zu verzeichnen, weitere Erläuterungen hierzu in Abschnitt 6.1.1.

5.1.1 Deskriptive Analyse

Tabelle 7 fasst die gemessenen Werte in den Untergruppen „Biparietaler Durchmesser (BPD)“, „Kopfumfang (KU)“, „Hinterer Hirnseitenventrikel (VP)“, „Cisterna magna (CM)“ und „Cerebellum (Cereb)“ von Untersucher 1 sowie die Werte der Künstlichen Intelligenz nach Mittelwerten und Standardabweichungen zusammen. Für jeden Parameter wurde im ersten Schritt ein einfacher t-Test durchgeführt um nach signifikanten Mittelwertsunterschieden zwischen Untersucher 1 und KI zu suchen. Die Anzahl der Freiheitsgrade wird mit df (number of degrees of freedom) angegeben (=n-1), die Effektstärke mit Cohen d. Eine grafische Darstellung der einzelnen Messungen pro Parameter für den Untersucher 1 und die künstliche Intelligenz ist in Abbildung 9 dargestellt.

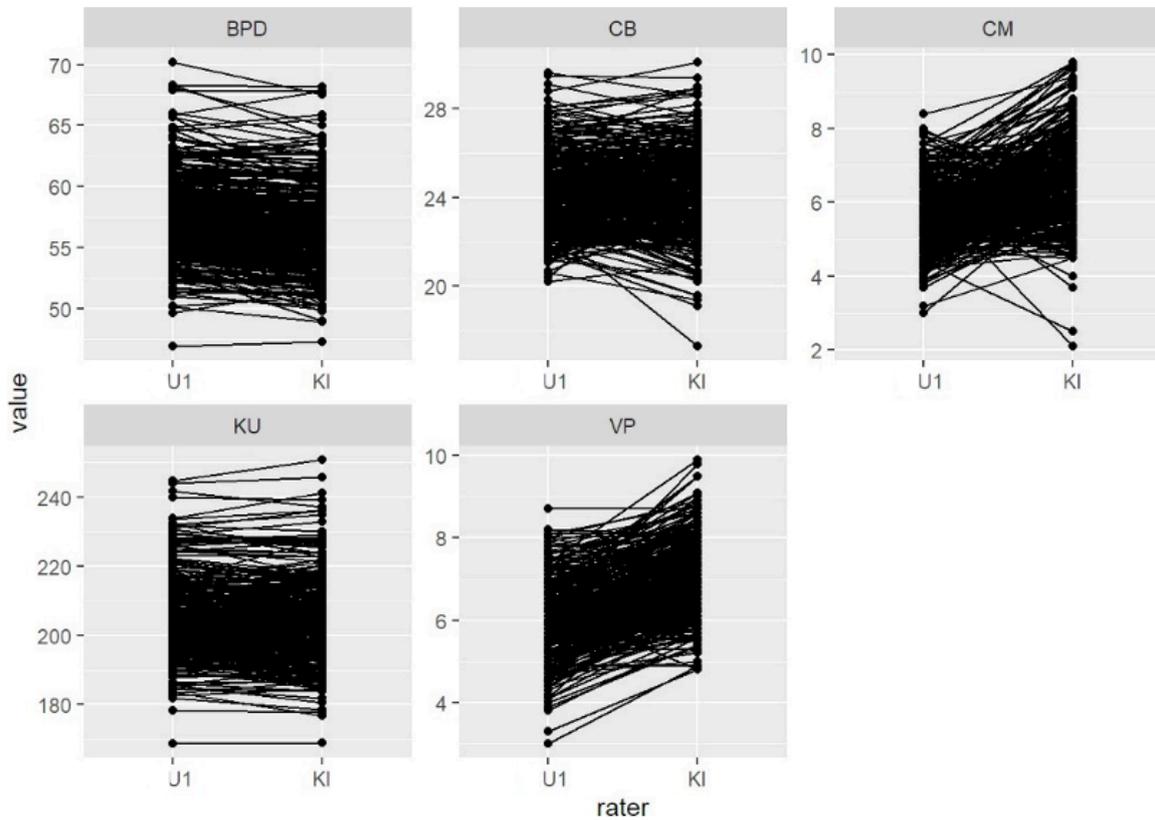


Abbildung 9: Grafische Darstellung der paarweisen Messungen pro Parameter BPD (Biparietaler Durchmesser), CB (Cerebellum), CM (Cisterna magna), KU (Kopfumfang), VP (Hinterer Hirnseitenventrikel). Für jeden Parameter zeigt sich links die Anzahl der gemessenen Millimeter, der Punkt auf der linken Seite steht für die Messung des Untersuchers 1 (U1), dieser wird mit der jeweiligen Messung desselben Feten durch die künstliche Intelligenz (KI) auf der rechten Seite per durchgehender Linie verbunden. Wurden größere Abweichungen der Messung zwischen Untersucher 1 und KI gefunden, ist die Linie nicht waagrecht oder annähernd waagrecht sondern steil nach oben oder unten abgewichen. Dies zeigt, dass die Werte in diesen Fällen nicht als äquivalent gelten. Bedingt durch die große Anzahl der Untersuchungen sind allerdings die Linien meist nicht klar zu trennen.

Parameter	MW U1 in mm	SD U1	n U1	MW KI in mm	SD KI	n KI	p-Wert	df	Cohen d
BPD	57.58	3.34	384	56.31	3.40	369	<0.001	368	0.38
KU	205.23	11.00	384	203.67	11.94	369	<0.001	368	0.14
VP	6.00	0.90	384	7.06	0.93	353	<0.001	352	-1.16
CM	5.65	0.92	384	6.52	1.17	372	<0.001	371	-0.83
Cereb	24.15	1.68	384	24.24	1.88	372	0.214	371	-0.05

Tabelle 7: Deskriptive Datenanalyse der Daten des Untersuchers 1 und der künstlichen Intelligenz, aufgeteilt in die 5 untersuchten Parameter, es wird jeweils die Anzahl der Untersuchungen n, der Mittelwert, die Standardabweichung, der p-Wert, die Freiheitsgrade und die Effektstärke angegeben. Bei allen Parametern bis auf das Cerebellum wurden signifikante Unterschiede in den Messungen gefunden ($p < 0,001$), nähere Erläuterungen hierzu in Abschnitt 5.1.1.

Biparietaler Durchmesser und Kopfumfang Für diese beiden Parameter liegt der p-Wert $< 0,001$. Der Untersucher hat im Vergleich zur künstlichen Intelligenz im Schnitt signifikant mehr Millimeter für diese Strukturen gemessen. Eine weitere Analyse zeigt Abschnitt 5.1.3. Die Effektstärke ist für beide Parameter klein, dies deutet darauf hin, dass die beiden Rater möglicherweise trotz signifikantem Unterschied als äquivalent anzusehen sind.

Hinterer Hirnseitenventrikel und Cisterna magna Für diese beiden Parameter liegt der p-Wert ebenfalls $< 0,001$. Hier hat der Untersucher im Vergleich zur künstlichen Intelligenz signifikant weniger Millimeter im Schnitt ausgemessen. Im Schnitt hat Untersucher 1 den Ventrikel um -1,06 mm kleiner gemessen und die Cisterna magna -0,87 mm kleiner. Hier liegt eine große Effektstärke für beide Parameter vor, dies deutet darauf hin, dass die Messungen der beiden Rater möglicherweise nicht äquivalent sind. In Abschnitt 5.1.3 folgen weitere Ausführungen.

Cerebellum Hier liegt der p-Wert bei 0,214. Für diesen Parameter lässt sich kein signifikanter Messunterschied zwischen der Messung des menschlichen Untersuchers und der automatischen Messung der künstlichen Intelligenz feststellen. Für diesen Parameter wird somit statistisch festgestellt, dass beide Messverfahren äquivalent sind. Untersucher

1 hat das Cerebellum im Schnitt nur -0,09 mm kleiner gemessen als die künstliche Intelligenz. Da kein signifikanter Messunterschied festgestellt werden kann, ist die Effektstärke nicht aussagekräftig.

5.1.2 Interrater-Reliabilität

Zur Überprüfung der Übereinstimmung der Ergebnisse zwischen der manuellen Messung und der automatischen Messung erfolgte eine Quantifizierung der Übereinstimmung mittels Intraklassen-Korrelation (ICC). Eine ICC > 0,7 bedeutet eine hohe Übereinstimmung. Auswertet nach der Interrater-Reliabilität zeigt sich in Tabelle 8, dass insbesondere der BPD und der KU eine sehr hohe ICC aufweisen, während die VP und das Cerebellum einen Wert im Mittelfeld anzeigen. Die CM dagegen hat eine sehr geringe ICC, somit stimmen hier die Ratings des Untersuchers und der KI nur sehr schwach überein.

Parameter	ICC
BPD	0.900
KU	0.901
VP	0.559
CM	0.152
Cereb	0.676

Tabelle 8: Intraklassen-Korrelation (ICC). Hiermit wird die Übereinstimmung der Messungen von Untersucher 1 und künstlicher Intelligenz quantifiziert, es erfolgt eine Quantifizierung für jeden einzelnen Parameter. Werte >0,7 stehen für eine hohe Übereinstimmung. Somit zeigen die Messungen der Cisterna magna (CM) eine sehr geringe Übereinstimmung, während die Messungen des biparietalen Durchmessers (BPD) und des Kopfumfanges (KU) eine hohe Übereinstimmung aufweisen.

5.1.3 Darstellung der Messungen von Untersucher 1 und KI mittels Bland-Altman-Plots

Bland-Altman-Plots dienen dem Vergleich zweier Messmethoden. Hier werden die Messdifferenzen zwischen Untersucher 1 und der KI verglichen. Für jeden Parameter wurden im Vorfeld kritische maximale Abweichungen festgelegt, welche in Tabelle 4 aufgezeigt werden. Bei Bland-Altman-Plots wird das kritische Intervall ausgehend von den Mittelwerten der Differenzen festgelegt. Dies ist sinnvoll, da angenommen werden kann, dass auch die manuelle Messung nicht exakt genau ist. Daher wird der Mittelwert aller Differenzen aus manueller und automatischer Messung verwendet, da angenommen wird, dass dies der wahrscheinlich wahre Wert ist. Ist der Mittelwert ins positive verschoben, bedeutet dies, dass die manuellen Messungen im Mittel größer ausfallen und umgekehrt. Wenn die Grenzen des Limits of Agreement (LoA) innerhalb der kritischen Grenzen liegen, werden die Messmethoden als äquivalent betrachtet. Das LoA muss kleiner der erlaubten Abweichung sein.

Biparietaler Durchmesser Wie oben erwähnt wurde als kritische Abweichung der Messungen zwischen manueller und automatischer Messung beim biparietalen Durchmesser $\leq 4,5\text{mm}$ festgelegt. Ausgehend vom Mittelwert der Differenzen (1,22 mm) wird das Intervall der erlaubten Abweichung berechnet ($1,22 \pm 4,5 \text{ mm} = -3,28 - 5,72 \text{ mm}$). Der Mittelwert der Differenzen $\pm 1,96$ Standardabweichungen bildet das obere und untere Limit of Agreement (LoA). Wie in Abbildung 10 ersichtlich, liegen die Grenzen des LoA in diesem Fall bei jeweils -1,73 und 4,16 mm. Auch wenn man die 95 %-Konfidenzintervalle zu den Grenzen des LoA (der graue Bereich) mit einbezieht, bewegen sich die Werte innerhalb der erlaubten Abweichung. Da sich diese Werte deutlich innerhalb der erlaubten Abweichung befinden, wird festgestellt, dass die KI-Messung des Biparietalen Durchmessers äquivalent zur menschlichen Messung für diesen Parameter ist.

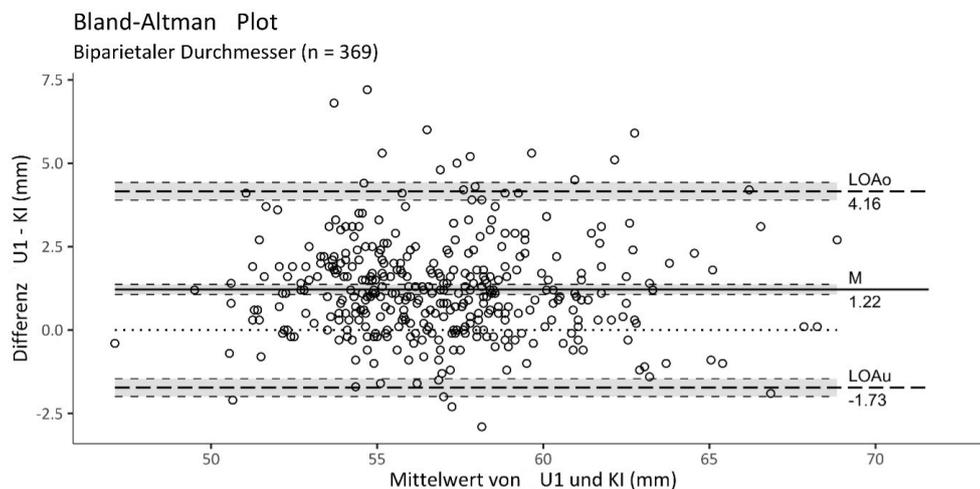


Abbildung 10: Bland-Altman-Plot für den Parameter BPD, Messungen des Untersuchers 1. Der Mittelwert der Differenzen ist $M=1,22$. Dieser Wert $\pm 1,96$ Standardabweichungen bildet das obere und untere Limit of Agreement (LoA). Die tolerable Abweichung für den Parameter BPD wurde bei $\leq 4,5 \text{ mm}$ festgelegt. Hieraus lässt sich das Intervall der erlaubten Abweichung berechnen ($1,22 \pm 4,5 \text{ mm} = -3,28 - 5,72 \text{ mm}$). Aus der Abbildung wird ersichtlich, dass die Grenzen der LoAs innerhalb dieser kritischen Grenze liegen. Somit gilt, dass KI und Untersucher 1 für den Parameter BPD äquivalent messen.

Kopfumfang Für den Parameter Kopfumfang wurde ≤ 10 mm als kritische Differenz festgelegt. Vom Mittelwert 1,44 mm ausgehend ist das Intervall der erlaubten Abweichung - 8,56 mm - 11,44 mm ($1,44 \text{ mm} \pm 10,00 \text{ mm} = -8,56 \text{ mm} - 11,44 \text{ mm}$). Das Limit of Agreement liegt bei 8,54 – 11,41, wie in Abbildung 11 dargestellt. Damit liegen 95 % der Messwertdifferenzen noch ganz knapp unterhalb der erlaubten Differenzen. Allerdings wird unter Berücksichtigung des 95%-Konfidenzintervalls der LOAs dieser Bereich schon leicht überschritten. Für den Parameter Kopfumfang ist die Abweichung zwischen manueller Messung und KI somit schon tendenziell ungenau, mit einer Tendenz zu höheren Messungen des menschlichen Untersuchers. Somit kann keine eindeutige statistische Aussage zur Äquivalenz der Messungen getroffen werden.

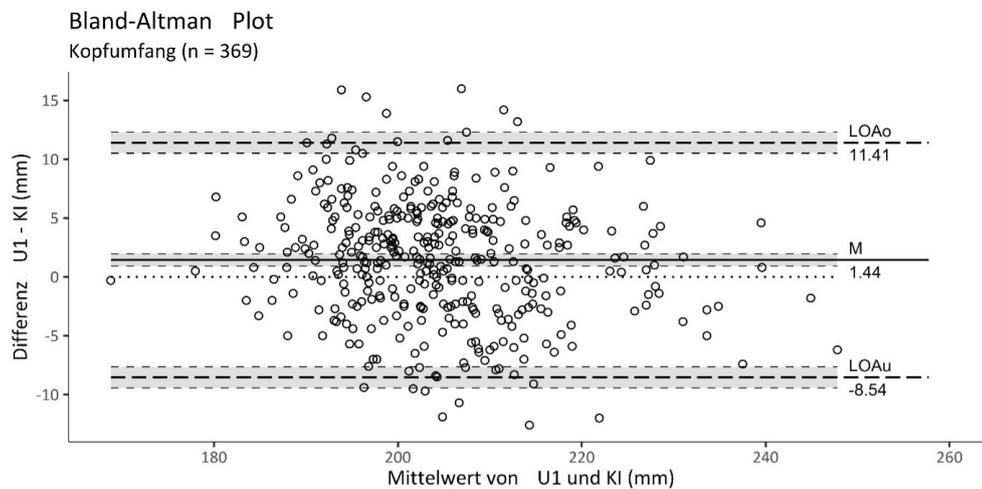


Abbildung 11: Bland-Altman-Plot für den Parameter KU, Messungen des Untersuchers 1. Der Mittelwert der Differenzen ist $M=1,44$. Dieser Wert $\pm 1,96$ Standardabweichungen bildet das obere und untere Limit of Agreement (LoA). Die tolerable Abweichung für den Parameter KU wurde bei ≤ 10 mm festgelegt. Hieraus lässt sich das Intervall der erlaubten Abweichung berechnen ($1,44 \pm 10,00 \text{ mm} = -8,56 - 11,44 \text{ mm}$). Aus der Abbildung wird ersichtlich, dass die Grenzen der LoAs unter Berücksichtigung des 95 %-Konfidenzintervalls (der graue Bereich) diesen Bereich schon leicht überschreiten. Die Messungen sind somit schon tendenziell ungenau, es kann keine eindeutige statistische Aussage zur Äquivalenz der Messung beim Parameter KU getroffen werden.

Hinterer Hirnseitenventrikel Die kritische Differenz wurde hier bei ≤ 2 mm festgelegt. Der Mittelwert der Differenzen beträgt $-1,06$ mm. Somit ergibt sich das Intervall der erlaubten Abweichung: $-1,06 \pm 2,00$ mm = $-3,06 - 0,94$ mm, das Limit of Agreement ist $-2,74 - 0,62$, siehe Abbildung 12. Somit liegen die LoAs für den Parameter VP noch unterhalb des Intervalls der kritischen Differenzen, auch wenn die 95%-Konfidenzintervalle berücksichtigt werden. Die manuelle Messung ist hier äquivalent zur künstlichen Intelligenz mit einer Tendenz zu geringeren Messwerten.

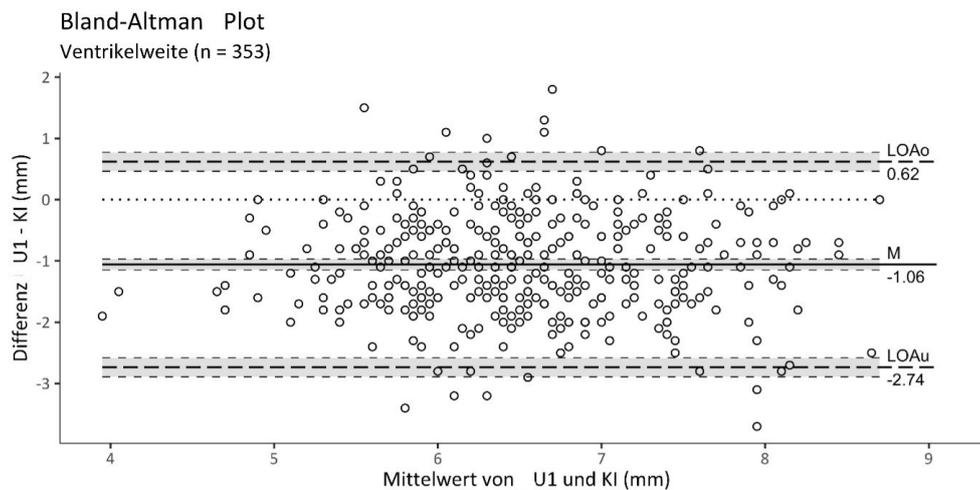


Abbildung 12: Bland-Altman-Plot für den Parameter VP, Messungen des Untersuchers 1. Der Mittelwert der Differenzen ist $M = -1,06$. Dieser Wert $\pm 1,96$ Standardabweichungen bildet das obere und untere Limit of Agreement (LoA). Die tolerable Abweichung für den Parameter VP wurde bei ≤ 2 mm festgelegt. Hieraus lässt sich das Intervall der erlaubten Abweichung berechnen ($-1,06 \pm 2,00$ mm = $-3,06 - 0,94$ mm). Aus der Abbildung wird ersichtlich, dass die Grenzen der LoAs innerhalb dieser kritischen Grenze liegen. Somit gilt, dass KI und Untersucher 1 für den Parameter VP äquivalent messen.

Cisterna magna Die kritische Differenz ist hier ebenfalls ≤ 2 mm. Der Mittelwert der Differenzen beträgt $-0,87$ mm. Somit ergibt sich das Intervall der erlaubten Abweichung: $-0,87 \pm 2,00$ mm = $-2,87 - 1,13$ mm. Im Vergleich mit den LoAs ($-3,56 - 1,82$) in Abbildung 13 zeigt sich, dass die gemessenen Differenzen das Intervall der kritischen Differenz deutlich überschreiten. Die Abweichung ist also zu hoch, die Messmethoden werden somit als statistisch nicht äquivalent betrachtet. Der menschliche Untersucher misst tendenziell weniger als die künstliche Intelligenz. Dies deckt sich mit den Vermutungen aus der deskriptiven Analyse, bei der dieser Parameter eine große Effektstärke aufweist.

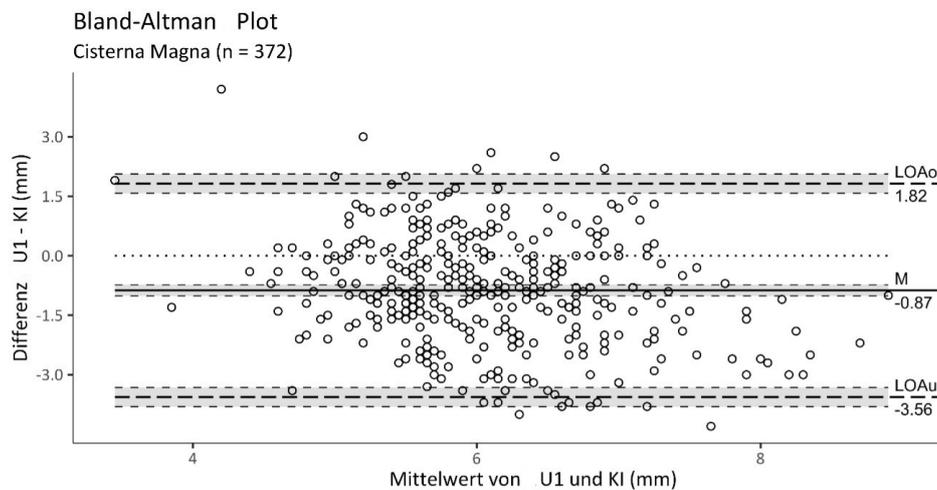


Abbildung 13: Bland-Altman-Plot für den Parameter CM, Messungen des Untersuchers 1. Der Mittelwert der Differenzen ist $M = -0,87$. Dieser Wert $\pm 1,96$ Standardabweichungen bildet das obere und untere Limit of Agreement (LoA). Die tolerable Abweichung für den Parameter CM wurde bei ≤ 2 mm festgelegt. Hieraus lässt sich das Intervall der erlaubten Abweichung berechnen ($-0,87 \pm 2,00$ mm = $-2,87 - 1,13$ mm). Somit liegen die LoAs ($-3,56-1,82$) außerhalb der erlaubten Abweichung, die Messwerte gelten somit als statistisch nicht äquivalent.

Cerebellum Hier wurde die kritische Differenz bei $\leq 3,5$ mm festgelegt. Der Mittelwert der Differenzen beträgt $-0,09$ mm. Somit ergibt sich das Intervall der erlaubten Abweichung: $-0,87 \pm 3,5$ mm = $-3,59 - 3,41$ mm. Es ergibt sich aus Abbildung 14, dass für den Parameter Cerebellum die LoAs ($-2,9 - 2,72$) der gemessenen Abweichungen noch innerhalb des Intervalls der kritischen Differenzen liegen, selbst wenn die 95% Konfidenzintervalle dazu gerechnet werden. Hier gilt wieder die Äquivalenz beider Messmethoden. Dies passt zu den Ergebnissen der deskriptiven Analyse, die schon aufgezeigt hat, dass KI und Untersucher 1 für diesen Parameter keinen signifikanten Unterschied im Mittelwert gemessen haben. Untersucher 1 und KI haben im Mittel sehr ähnlich gemessen, sodass keine klare Tendenz ins positive oder negative zu erkennen ist.

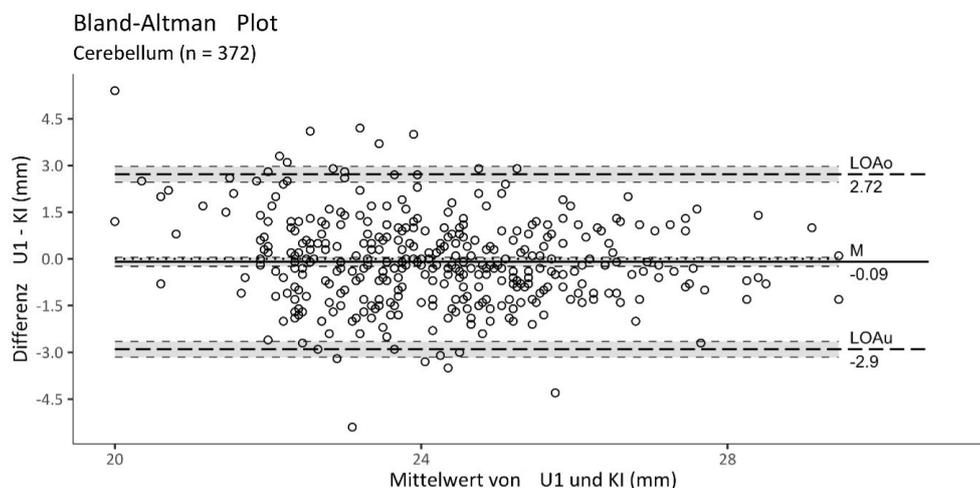


Abbildung 14: Bland-Altman-Plot für den Parameter Cerebellum, Messungen des Untersuchers 1. Der Mittelwert der Differenzen ist $M = -0,09$. Dieser Wert $\pm 1,96$ Standardabweichungen bildet das obere und untere Limit of Agreement (LoA). Die tolerable Abweichung für den Parameter Cerebellum wurde bei $\leq 3,5$ mm festgelegt. Hieraus lässt sich das Intervall der erlaubten Abweichung berechnen ($-0,87 \pm 3,5$ mm = $-3,59 - 3,41$ mm). Aus der Abbildung wird ersichtlich, dass die Grenzen der LoAs innerhalb dieser kritischen Grenze liegen. Somit gilt, dass KI und Untersucher 1 für den Parameter Cerebellum äquivalent messen.

5.1.4 Varianzanalyse von potentiellen Störfaktoren

In diesem Abschnitt wird untersucht, ob es Risikofaktoren für größere Abweichungen zwischen der künstlichen Intelligenz und dem Untersucher innerhalb der einzelnen gemessenen Parameter gibt. Im Rahmen der Vorbereitung zur Studie wurden verschiedene mögliche Störfaktoren besprochen, welche nun aufgezählt werden:

1. BMI

Der BMI der Mutter lag im Mittel bei $24.51 \pm 4,81$ SD

2. Gewicht des Kindes

Das Gewicht des Kindes lag im Mittel bei $540 \text{ g} \pm 90,33 \text{ g}$ SD

3. Fetale Lage

Unterteilt in Schädellage (52,0 % oder 199 Feten), Beckenendlage (37,1 % oder 142 Feten) und Querlage (10,9 % oder 42 Feten)

4. Dorsoanteriore Stellung

Ja in 14,62 % (56 Feten)

Nein in 85,37 % (327 Feten)

5. Lage der Plazenta

Vorderwand 48,3 % (184)

Hinterwand 50,1 % (191)

Seitenwand 0,8 % (3)

Plazenta praevia 0,8 % (3)

Im nächsten Schritt wurden die einzelnen Risikofaktoren bezüglich eines möglichen Einflusses auf die Differenz beider Untersucher einzeln für jeden gemessenen Parameter untersucht mittels univariater linearer Regression. Wenn man den Risikofaktor (zum Beispiel den BMI) auf der x-Achse und die Differenz der Messungen zwischen Untersucher und KI auf der y-Achse plottet, wird mithilfe der univariaten linearen Regression die lineare Funktion gefunden, die 'am besten zu den Daten passt'. Eine sehr große absolute Steigung (entspricht in den Abbildungen Beta) deutet auf einen starken Zusammenhang an. Allerdings nur, wenn die statistische Signifikanz gegeben ist (bei einem p-Wert <

0.05). Ergaben sich pro analysiertem Parameter (BPD, KU, VP, CM, Cereb) mehr als ein signifikanter Prädiktor, wurden diese in im Rahmen einer multiplen linearen Regression noch einmal gemeinsam getestet. Dies ist wichtig, da die Störfaktoren voneinander abhängig sein könnten. Solche Abhängigkeiten und Zusammenhänge können möglicherweise durch die multiple Regression aufgezeigt werden. Der Übersichtlichkeit halber wurden alle in der univariaten linearen Regression analysierten Parameter (BPD, KU, VP, CM, Cereb) pro Störfaktor gruppiert dargestellt. Die Tabellen stellen trotzdem 5 verschiedene univariate Regressionen dar, die inhaltlich nicht zusammenhängen.

BMI Der BMI der Mutter hatte bei keinem der gemessenen Parameter einen signifikanten Einfluss auf die Messdifferenz zwischen Untersucher 1 und der KI, der p-Wert ist nie kleiner als 0,05 wie in Abbildung 15 dargestellt ist. Die niedrige Bedeutung des BMI erkennt man auch daran, dass das R², also die erklärte Varianz im Rahmen des Modells, nahe 0 lag beziehungsweise 0 ist. Es wird also so gut wie keine Varianz durch den BMI erklärt ⁵.

BMI

Predictors	Differenz BPD			Differenz KU			Differenz VP			Differenz CM			Differenz Cereb		
	Beta	CI	p	Beta	CI	p	Beta	CI	p	Beta	CI	p	Beta	CI	p
BMI	0.01	-0.02 – 0.04	0.522	-0.02	-0.13 – 0.09	0.750	-0.00	-0.02 – 0.02	0.968	-0.01	-0.04 – 0.02	0.564	-0.02	-0.05 – 0.01	0.315
Observations	369			369			353			372			372		
R ²	0.001			0.000			0.000			0.001			0.003		

Abbildung 15: Einfluss des BMI auf die Messdifferenz. Alle 5 Parameter wurden in Bezug auf den BMI separat betrachtet. Der p-Wert ist bei keinem Parameter < 0,05, das bedeutet, dass der BMI zu keiner signifikanten Messdifferenz zwischen Untersucher 1 und künstlicher Intelligenz geführt hat. Bei fehlender Signifikanz sind die Beta-Werte nicht aussagekräftig. Den niedrigen Einfluss des BMI zeigen auch die R²-Werte, die 0 oder nahe 0 liegen.

Gewicht des Kindes Das Gewicht des Kindes hat einen signifikanten Einfluss auf die Messdifferenz der Ventrikelbreite zwischen Untersucher 1 und der künstlichen Intelligenz.

⁵R² = 1 würde bedeuten, dass 100 % der Varianz durch den Prädiktor erklärt werden

Pro steigendem Gramm steigt die Messdifferenz um 0,002mm (beta) innerhalb der Ventrikelbreite (p = 0.003). Das Modell erklärt dabei 2,4 % der Gesamtvarianz der Messdifferenz (R2 ist 0,024). Das heißt, je schwerer das Kind, desto größer wird die Differenz zwischen beiden Untersuchern, allerdings sind die Effekte klein. Bei allen anderen Parametern (BPD, KU, CM und Cerebellum) hat das fetale Gewicht keinen Einfluss auf die Messdifferenz, da die p-Werte >0,05 sind, wie in Abbildung 16 aufgezeigt wird.

Gewicht des Kindes

Predictors	Differenz BPD			Differenz KU			Differenz VP			Differenz CM			Differenz Cereb		
	Beta	CI	p	Beta	CI	p	Beta	CI	p	Beta	CI	p	Beta	CI	p
Gewicht des Kindes	0.001	-0.001	0.255	-0.001	-0.007	0.758	0.002	0.001	0.003	-0.001	-0.002	0.375	-0.000	-0.002	0.861
		0.003			0.005			0.003			0.001			0.001	
Observations	367			367			351			370			370		
R ²	0.004			0.000			0.024			0.002 / -0.001			0.000 / -0.003		

Abbildung 16: Einfluss des fetalen Gewichtes auf die Messdifferenz. Hier zeigt sich beim Parameter VP ein signifikanter Einfluss des fetalen Gewichtes auf die Messdifferenz (p-Wert 0,003). Beta= 0,002 bedeutet, dass pro steigendem Gramm die Messdifferenz des Ventrikels um 0,002 mm ansteigt. Da R2 mit 0,024 niedrig ist, ist der Effekt insgesamt aber klein.

Fetale Lage Die Lage des Kindes zeigt eine signifikante Risikosteigerung einer um 0,37 mm höheren Differenz bei Beckenendlage im Vergleich zur Schädellage bezüglich des biparietalen Durchmessers ($p = 0.028$). Das Modell erklärt dabei 1,4 % der Gesamtvarianz. Somit erschwert die Beckenendlage im Vergleich zur Schädellage die Übereinstimmung beider Rater in geringem Maße. Bei der Querlage wird für keinen Parameter eine signifikante Messdifferenz gefunden im Vergleich zur Schädellage.

Lage des Kindes

Predictors	Differenz BPD			Differenz KU			Differenz VP			Differenz CM			Differenz Cereb		
	beta	CI	p	beta	CI	p	beta	CI	p	beta	CI	p	beta	CI	p
BEL	0.37	0.04 – 0.70	0.028	-0.09	-1.21 – 1.03	0.878	0.03	-0.16 – 0.22	0.765	0.02	-0.28 – 0.32	0.908	0.06	-0.26 – 0.37	0.719
QL	0.01	-0.50 – 0.52	0.963	-0.67	-2.40 – 1.06	0.445	0.20	-0.09 – 0.49	0.181	0.23	-0.23 – 0.70	0.326	0.45	-0.03 – 0.94	0.067
Observations	368			368			352			371			371		
R ²	0.014			0.002			0.005			0.003			0.009		

Abbildung 17: Einfluss der fetalen Lage auf die Messdifferenz. Hier zeigt sich für den Parameter BPD eine signifikante ($p=0,028$) Risikosteigerung einer höheren Messdifferenz bei Beckenendlage von 0,37 mm (beta) im Vergleich zur Schädellage. Somit erschwert die Beckenendlage die Übereinstimmung zwischen Untersucher und künstlicher Intelligenz, aber wieder nur in geringem Maße ($R^2=0,014$). Alle anderen Parameter bleiben davon unberührt.

Dorsoanteriore Stellung Eine dorsoanteriore Stellung des Feten hatte im Vergleich zum nicht vorne liegenden Rücken keinen Einfluss auf die Messdifferenz beider Untersucher, egal welcher Parameter gemessen wurde ($p \geq 0.199$). R2 ist für jeden Parameter 0 oder nahe 0.

Dorsoanteriore Stellung

Predictors	Differenz BPD			Differenz KU			Differenz VP			Differenz CM			Differenz Cereb		
	Beta	CI	p	Beta	CI	p	Beta	CI	p	Beta	CI	p	Beta	CI	p
Ja,DA	-0.08	-0.51 – 0.36	0.729	-0.96	-2.43 – 0.51	0.199	0.12	-0.13 – 0.36	0.357	0.18	-0.21 – 0.58	0.363	-0.03	-0.44 – 0.38	0.886
Observations	368			368			352			371			371		
R ²	0.000			0.005			0.002			0.002			0.000		

Abbildung 18: Einfluss der dorsoanterioren Stellung auf die Messdifferenz. Bei keinem der untersuchten Parameter hatte die dorsoanteriore Stellung einen Einfluss auf die Messdifferenz zwischen Untersucher 1 und KI (p -Werte nie $< 0,05$). R2 ist wiederum 0 oder nahe 0. Somit sind die beta-Werte zu vernachlässigen.

Lage der Plazenta Die Lage der Plazenta hat bei der Messung der Ventrikelweite einen signifikanten Einfluss auf die Messdifferenz: Die Seitenwandlage reduziert die Messdifferenz um 1,07 mm im Vergleich zur Vorderwandlage ($p = 0.030$). Das Modell erklärt 2,1 % der Gesamtvarianz, somit ist der Effekt auch hier klein. Für alle anderen Parameter sowie Plazentalokalisationen zeigt sich kein signifikanter Einfluss auf die Messdifferenz, p-Werte sind stets > 0.05 .

Lage der Plazenta

Predictors	Differenz BPD			Differenz KU			Differenz VP			Differenz CM			Differenz Cereb		
	Beta	CI	p	Beta	CI	p	Beta	CI	p	Beta	CI	p	Beta	CI	p
[HW]	0.08	-0.23 – 0.39	0.609	0.57	-0.49 – 1.62	0.292	-0.13	-0.31 – 0.05	0.155	0.14	-0.15 – 0.42	0.343	0.06	-0.24 – 0.36	0.688
[SW]	0.94	-0.78 – 2.66	0.282	2.70	-3.14 – 8.54	0.363	-1.07	-2.03 – -0.10	0.030	0.85	-0.73 – 2.42	0.290	0.63	-1.02 – 2.28	0.450
[Praevia]	0.08	-1.64 – 1.80	0.932	-4.77	-10.60 – 1.07	0.109	0.47	-0.50 – 1.43	0.342	0.68	-0.89 – 2.26	0.395	0.63	-1.02 – 2.28	0.450
Observations	366			366			350			369			369		
R ²	0.004			0.013			0.021			0.007			0.003		

Abbildung 19: Einfluss der Lage der Plazenta auf die Messdifferenz. Hier zeigt sich, dass die Seitenwandplazenta (SW) einen signifikanten Einfluss auf die Messdifferenz zwischen Untersucher 1 und KI für den Parameter VP hatte ($p=0,030$). Im Vergleich zur Vorderwandplazenta reduziert die Seitenwandplazenta die Messdifferenz um 1,07 mm ($\text{beta}=-1,07$). Wiederum ist der Effekt gering mit einer Gesamtvarianz von 2,1 % ($R^2=0,021$).

5.1.5 Multiple Regression für alle Kriterien mit mehr als einem signifikanten Prädiktor

Da bei der Ventrikelmessung zwei Prädiktoren einen signifikanten Einfluss aufweisen, wurden beide noch einmal im Rahmen einer multiplen linearen Regression gemeinsam getestet. Hierbei zeigt sich, dass bei der gemeinsamen Testung der Signifikanz beide Prädiktoren ihre Signifikanzen verlieren, wobei das Gewicht zumindest einen leichten Trend zur Signifikanz behält (p-Wert 0,086), dies ist in Abbildung 20 dargestellt.

<i>Predictors</i>	Differenz VP		
	<i>beta</i>	<i>CI</i>	<i>p</i>
Plazenta_fac [HW]	-0.069	-1.229 – 1.090	0.906
Plazenta_fac [SW]	0.784	-18.587 – 20.156	0.937
Plazenta_fac [Praevia]	5.837	-6.365 – 18.039	0.347
Gewicht des Kindes	0.001	<-0.001 – 0.003	0.086
Plazenta_fac [HW] * Gewicht des Kindes	<-0.001	-0.002 – 0.002	0.921
Plazenta_fac [SW] * Gewicht des Kindes	-0.003	-0.038 – 0.031	0.850
Plazenta_fac [Praevia] * Gewicht des Kindes	-0.010	-0.031 – 0.012	0.384
Observations	349		
R ²	0.038		

Abbildung 20: Einfluss von Gewicht und Plazentalage auf die Messung des Hirnseitenventrikels. Da bei der Ventrikelmessung zwei Prädiktoren identifiziert wurden, welche einen Einfluss auf die Messdifferenz hatten, wurde zusätzlich eine multiple lineare Regression durchgeführt. Hier zeigt sich, dass beide Prädiktoren ihre Signifikanz verlieren und somit keinen Einfluss auf die Messdifferenz hatten, alle p-Werte sind >0,05.

5.1.6 Fehlende Volumenaufnahmen und Fehlmessungen

In 4 % der Fälle konnte kein Volumen des fetalen Kopfes aufgenommen werden und daher keine Messung mit Hilfe der künstlichen Intelligenz erfolgen, in weiteren 2 % gelang es dem Computer nicht die richtigen Messebenen zu identifizieren, eine Übersicht zeigt

Tabelle 5. In diesen Fällen muss der Untersucher weiterhin alle Messebenen getrennt aufsuchen und manuelle Messungen anfertigen. Die 2 % der Messungen, bei denen es dem Computer nicht gelingt, die richtigen Messebenen darzustellen erkennt man zum einen am stark divergierenden Ergebnis der Messung sowie zum anderen beim Blick auf das entstandene Ultraschallbild. Ein Beispiel ist in Abbildung 21 dargestellt. Hier hat die Volumenaufnahme keine adäquaten Bilder aufgezeichnet, die oben genannten notwendigen Ebenen sind nicht korrekt dargestellt. Vor allem die transthalämische Ebene ist nicht dargestellt worden, dennoch misst der Computer einen Kopfumfang. Dies geschieht letztlich aus unklaren Gründen, möglich erscheint ein technisches Versagen. Es zeigt allerdings die Wichtigkeit der Überprüfung der künstlichen Intelligenz durch den Menschen.

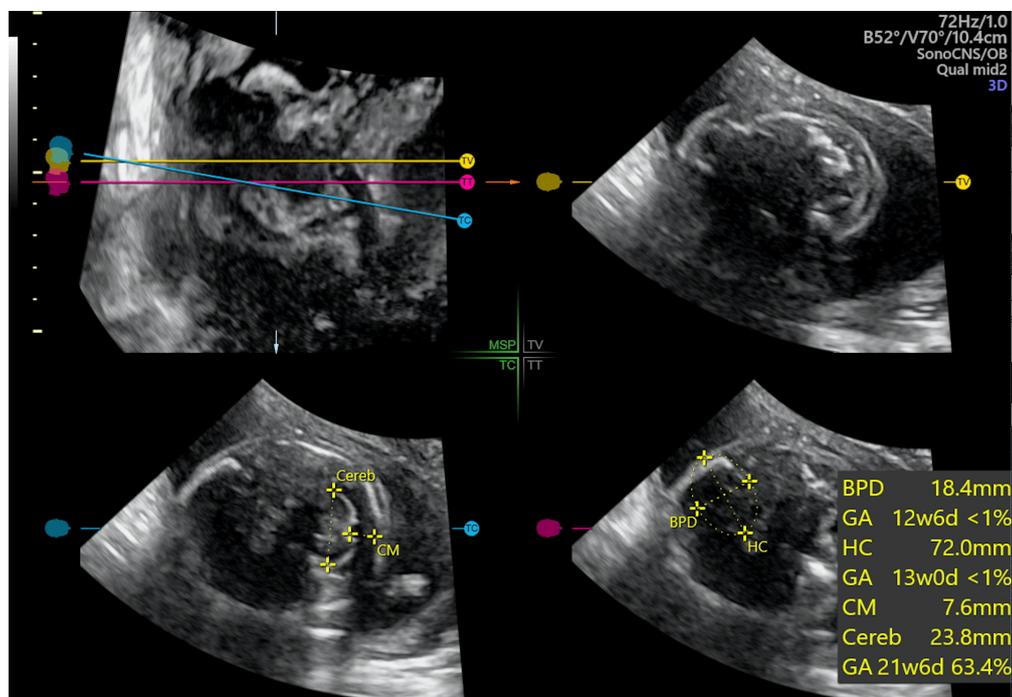


Abbildung 21: Eindeutig falsche Messebenen. Die transthalämische Ebene, in der Abbildung rechts unten, ist nicht korrekt eingestellt, die Hirnstrukturen sind nicht klar zu erkennen. Dennoch misst die künstliche Intelligenz Werte für den Kopfumfang und den biparietalen Durchmesser aus. Diese entsprechen nicht annähernd den wahren Werten. Zum einen ist dies optisch zu erkennen, zum anderen zeigt der Computer in der tabellarischen Auflistung rechts unten, dass die gemessenen Werte einem Gestationsalter (GA) von 12+6 bzw. 13+0 SSW entsprechen. Letztlich bleibt unklar, weshalb die Messebenen in diesem Fall nicht erkannt wurden und die Fehlmessungen entstanden sind.

5.1.7 Fehlerwertanalyse

Tabelle 4 zeigt eine Übersicht über die tolerable Abweichung in Millimeter pro Parameter. Abweichungen über diese Werte hinaus werden als Fehler betrachtet. Auch die zuvor aus der deskriptiven Analyse ausgeschlossenen Patientinnen fließen in diese Analyse mit ein. Die Auswertungen der Anzahl der Fehlmessungen der künstlichen Intelligenz pro Zielparame-ter zeigte, dass die meisten und größten Fehlmessungen bei der Cisterna magna und bei der Ventrikelmessung gemacht wurden. Eine Übersicht zeigt Tabelle 9. Um die Risikofaktoren einer Fehlmessung zu bestimmen, wurden logistische Regressionen mit dem Kriterium „Fehlmessung innerhalb irgendeines der 5 Parameters“ und den Prädikto-ren BMI, Gewicht, dorsoantere Stellung, fetale Lage und Lage der Plazenta berechnet. Auch hier wurden initial alle logistischen Regressionen univariat, also einzeln für jeden Prädiktor berechnet. Dabei fanden sich keine signifikanten Prädiktoren für Fehlmessun-gen innerhalb der einzelnen Parameter.

Parameter	n = Fehler	Gesamtzahl der Messun- gen	Prozent	Mittelwert der ab- soluten Differenz
BPD	13	393	3.31	8.08
KU	28	393	7.12	18.05
VP	49	393	12.47	2.39
CM	93	393	23.66	2.73
Cereb	8	393	2.04	4.33

Tabelle 9: Fehlerwertanalyse nach Parametern getrennt. Alle Messdifferenzen, die die maximale tolerable Abweichung für den jeweiligen Parameter überschritten, wurden als Fehler gewertet. Die meisten Fehler fanden sich für die Messung der Cisterna magna (n=93), die wenigstens für das Cerebellum (n=8). Auf die hohe Anzahl der Fehlmessungen der Cisterna magna wird nochmals in Kapitel 6.1.2 eingegangen.

5.1.8 Deskriptive Analyse der Ergebnisse des Untersuchers 2 und der KI

Tabelle 10 zeigt die Ergebnisse der Untersuchung von einer kleinen Stichprobe mit 20 Patientinnen, welche von einem zweiten Untersucher (U2) mit dem gleichen Ultraschallprogramm „SonoCNS“ durchgeführt wurden. Aufgelistet sind jeweils die Mittelwerte und die Standardabweichungen der beiden Rater, sowie die Anzahl der erfolgreich gemessenen Patientinnen pro Parameter. Es wurde im ersten Schritt ein einfacher t-Test für abhängige Stichproben pro Parameter gerechnet, um nach signifikanten Mittelwertsunterschieden zu suchen. In Abbildung 22 sind die Messwerte beider Untersucher noch einmal grafisch dargestellt für jeden Parameter. In der letzten Spalte der Tabelle 10 ist die Effektstärke für den jeweiligen Parameter aufgeführt. Die Effektstärke ist unabhängig von der Anzahl der Untersuchungen. Die Stichprobengröße ist relativ klein, daher kann keine generelle Aussage über die gesamte Population getroffen werden, die Daten sind somit nur Anhaltspunkte. Die Daten dienen primär dem generellen Vergleich der beiden menschlichen Untersucher mit der künstlichen Intelligenz.

Parameter	MW U2 in mm	SD U2	n U2	MW KI in mm	SD KI	n KI	p-Wert	df	Cohen d
BPD	55.23	1.89	20	53.70	2.07	20	<0.001	19	0.77
KU	199.52	5.99	20	199.14	8.85	20	0.731	19	0.05
VP	5.43	0.66	20	6.22	0.66	20	<0.001	19	-1.18
CM	5.58	0.88	20	6.03	1.01	20	0.147	19	-0.47
Cereb	23.86	0.85	20	24.09	1.41	20	0.300	19	-0.19

Tabelle 10: Deskriptive Datenanalyse von Untersucher 2 und der künstlichen Intelligenz, aufgeteilt in die 5 untersuchten Parameter. Es wird jeweils die Anzahl der Untersuchungen n (wesentlich kleiner als bei Untersucher 1), der Mittelwert, die Standardabweichung, der p-Wert, die Freiheitsgrade und die Effektstärke angegeben. Für die Parameter BPD und VP werden signifikante Messwertunterschiede festgestellt ($p < 0,05$). Dies deckt sich mit den Ergebnissen von Untersucher 1, ebenso wie die fehlende Signifikanz bei der Messung des Cerebellum. Bei Untersucher 2 werden aber auch für den Kopfumfang und die Cisterna magna keine signifikanten Messunterschiede festgestellt.

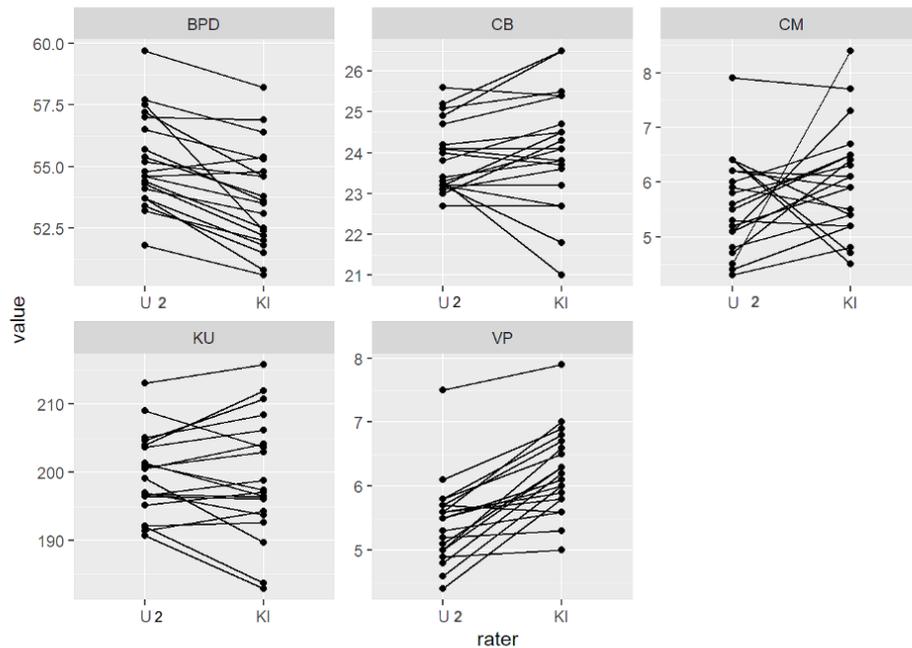


Abbildung 22: Grafische Darstellung der paarweisen Messungen pro Parameter (Biparietaler Durchmesser, Cerebellum, Cisterna magna, Kopfumfang, Hinterer Hirnseitenventrikel) bei Untersucher 2. Für jeden Parameter zeigt sich links die Anzahl der gemessenen Millimeter, der Punkt auf der linken Seite steht für die Messung des Untersuchers 2 (U2), dieser wird mit der jeweiligen Messung desselben Feten durch die künstliche Intelligenz (KI) auf der rechten Seite per durchgehender Linie verbunden. Wurden größere Abweichungen der Messung zwischen Untersucher 2 und KI gefunden, ist die Linie nicht waagrecht oder annähernd waagrecht sondern steil nach oben oder unten abgewichen. Dies zeigt, dass die Werte in diesen Fällen nicht als äquivalent gelten können. Die nach unten abgewichenen Linien beim Parameter BPD (biparietaler Durchmesser) zeigt zum Beispiel, dass der Untersucher 2 hier wie Untersucher 1 mehr Millimeter als die KI ausgemessen hat, wohingegen beim Ventrikel (VP) signifikant mehr Millimeter von der KI ausgemessen wurden.

Biparietaler Durchmesser Hier hat der Untersucher 2 im Schnitt signifikant mehr Millimeter als die KI gemessen (p-Wert <0.001). Auch Untersucher 1 hat signifikant mehr Millimeter gemessen. Die Effektstärke ist mit 0.77 im mittleren Bereich, anderes als bei Untersucher 1. Hieraus lässt sich noch kein Schluss auf die Äquivalenz der beiden Messungen ziehen.

Hinterer Hirnseitenventrikel Hier hat der Untersucher 2 im Schnitt signifikant weniger Millimeter gemessen als die KI (p-Wert <0.001). Im Schnitt hat der Untersucher 2 -0,79 mm weniger gemessen. Die Effektstärke ist hier hoch mit -1.18. Diese Ergebnisse decken sich mit denen von Untersucher 1.

Kopfumfang, Cisterna magna und Cerebellum Hier finden sich keine signifikanten Messunterschiede zwischen Untersucher 2 und der KI, die p-Werte sind in Tabelle 10 dargestellt. Die Effektstärke ist hier aufgrund der fehlenden Signifikanz nicht aussagekräftig. Auch bei Untersucher 1 gibt es keine signifikanten Messunterschiede für den Parameter Cerebellum. Kopfumfang und Cisterna magna zeigen jedoch bei Untersucher 1 signifikante Unterschiede. Die fehlende Signifikanz bei Untersucher 2 könnte der geringen Stichprobengröße geschuldet sein.

5.1.9 Interrater-Reliabilität des Untersuchers 2

Auswertet nach der Interrater-Reliabilität zeigt sich, dass insbesondere der BPD und der Kopfumfang eine hohe ICC aufweisen, während die VP und das Cerebellum einen Wert im Mittelfeld anzeigen. Die CM dagegen hat eine sehr geringe ICC, somit stimmen hier die Ratings des Untersuchers und der KI nur sehr schwach überein, eine Übersicht bietet Tabelle 11. Dies deckt sich mit der Interrater-Reliabilität des Untersuchers 1.

Parameter	ICC
BPD	0.811
KU	0.792
VP	0.668
CM	0.054
Cereb	0.671

Tabelle 11: Intraklassen-Korrelation Untersucher 2. Hiermit wird die Übereinstimmung der Messungen von Untersucher 2 und der künstlichen Intelligenz quantifiziert, es erfolgt eine Quantifizierung für jeden einzelnen Parameter. Werte $>0,7$ stehen für eine hohe Übereinstimmung. Somit zeigen die Messungen der Cisterna magna (CM) eine sehr geringe Übereinstimmung, während die Messungen des biparietalen Durchmessers (BPD) und des Kopfumfangs (KU) eine hohe Übereinstimmung aufweisen, Ventrikel und Cerebellum liegen im Mittelfeld. Dies ist äquivalent zur Vergleichsgruppe des Untersuchers 1.

5.1.10 Darstellung der Messungen von Untersucher 2 und KI mittels Bland-Altman-Plots

Die tolerablen Abweichungen aus Tabelle 4 bleiben auch hier gültig.

Biparietaler Durchmesser Der Mittelwert der Differenz ist 1,53 mm. Das Intervall der erlaubten Abweichung ist somit: $1,53 \pm 4,5 \text{ mm} = -2,97 - 6,03 \text{ mm}$. In Abbildung 23 wird ersichtlich, dass die LoAs der Messwertdifferenzen (-0,86 - 3,91) noch innerhalb des Intervalls der kritischen Differenzen liegen auch unter Berücksichtigung des 95 %-Konfidenzintervalls. Diese sind aufgrund der geringen Stichprobe hier deutlich größer. Der Untersucher 2 hat tendenziell mehr Millimeter als die KI gemessen, die Messungen werden aber als äquivalent betrachtet, genau wie bei Untersucher 1.

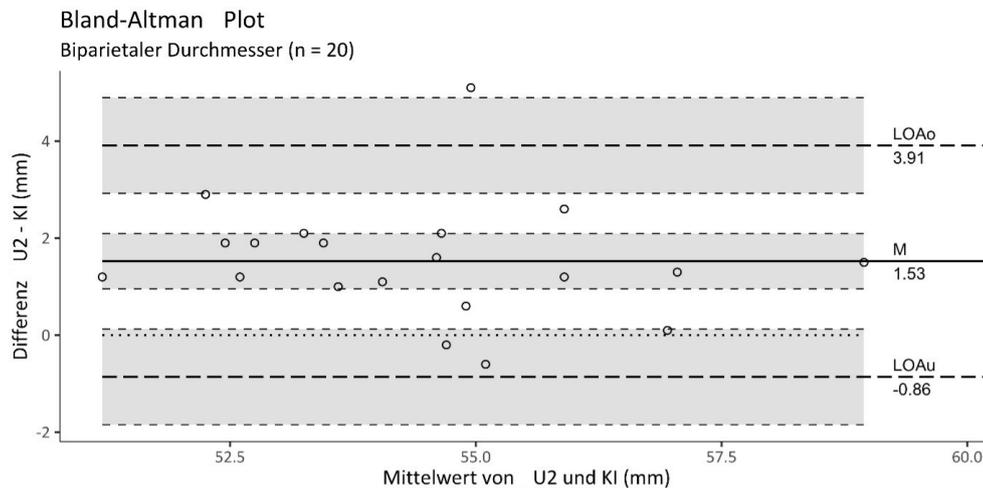


Abbildung 23: Bland-Altman-Plot für den Parameter BPD, Messungen des Untersuchers 2. Der Mittelwert der Differenzen ist $M=1,53$. Dieser Wert $\pm 1,96$ Standardabweichungen bildet das obere und untere Limit of Agreement (LoA). Die tolerable Abweichung für den Parameter BPD wurde bei $\leq 4,5 \text{ mm}$ festgelegt. Hieraus lässt sich das Intervall der erlaubten Abweichung berechnen ($1,53 \pm 4,5 \text{ mm} = -2,97 - 6,03 \text{ mm}$). Aus der Abbildung wird ersichtlich, dass die Grenzen der LoAs innerhalb dieser kritischen Grenze liegen. Somit gilt, dass KI und Untersucher 2 für den Parameter BPD äquivalent messen, dies deckt sich mit den Ergebnissen von Untersucher 1.

Kopfumfang Der Mittelwert der Differenz ist 0,38 mm. Hier ist keine klare Tendenz bezüglich der Abweichungsrichtung zu erkennen. Das Intervall der erlaubten Abweichung ist somit $0,38 \pm 10,00 \text{ mm} = -9,62 - 10,38 \text{ mm}$. Somit liegen wie in Abbildung 24 dargestellt die LoAs (-9,18– 9,94) noch unterhalb des Intervalls der kritischen Differenzen. Dies ändert sich allerdings wenn man die 95 % Konfidenzintervalle berücksichtigt, somit können die Messungen nicht als äquivalent gelten. In diesem Fall bräuchte man mehr Probanden, um eine bessere Interpretation des Ergebnisses zu ermöglichen. Dennoch ähnelt der Bland-Altman-Plot für den Parameter Kopfumfang dem des Untersuchers 1.

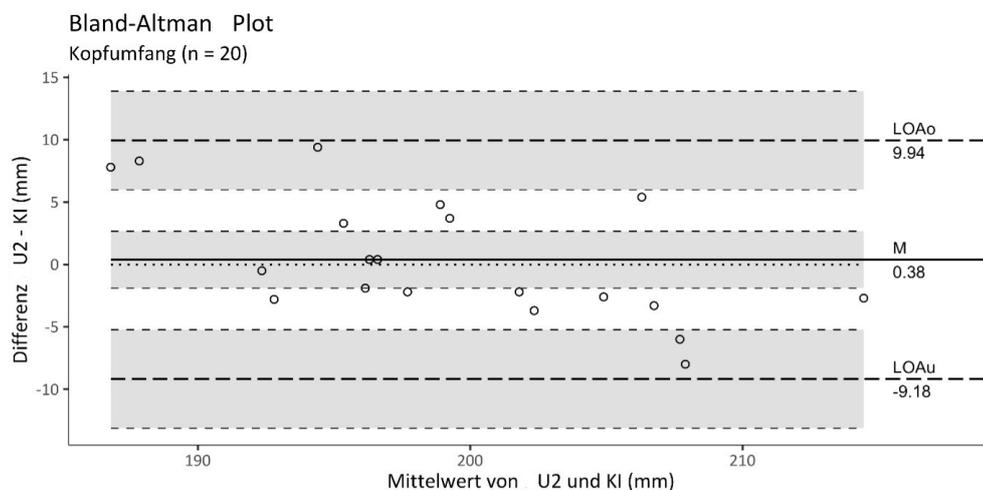


Abbildung 24: Bland-Altman-Plot für den Parameter KU, Messungen des Untersuchers 2. Der Mittelwert der Differenzen ist $M=0,38$. Dieser Wert $\pm 1,96$ Standardabweichungen bildet das obere und untere Limit of Agreement (LoA). Die tolerable Abweichung für den Parameter KU wurde bei $\leq 10 \text{ mm}$ festgelegt. Hieraus lässt sich das Intervall der erlaubten Abweichung berechnen ($0,38 \pm 10,00 \text{ mm} = -9,62 - 10,38 \text{ mm}$). Aus der Abbildung wird ersichtlich, dass die Grenzen der LoAs noch innerhalb dieser kritischen Grenze liegen. Dies ändert sich jedoch, wenn man die 95 % Konfidenzintervalle (der graue Bereich) berücksichtigt. Somit gilt, dass KI und Untersucher 2 für den Parameter KU nicht äquivalent messen.

Hinterer Hirnseitenventrikel Der Mittelwert der Differenzen ist $-0,78$ mm. Das Intervall der erlaubten Abweichung ist hier $-0,78 \pm 2,00$ mm $-2,78$ – $1,22$ mm. Vergleicht man nun Abbildung 25, liegen die LoAs ($-1,84$ – $0,28$) noch unterhalb des Intervalls der kritischen Differenzen, auch wenn die 95 % Konfidenzintervalle berücksichtigt werden. Die Messungen werden somit als äquivalent betrachtet. Untersucher 2 misst tendenziell etwas geringere Werte, dieses Ergebnis stimmt mit Untersucher 1 überein.

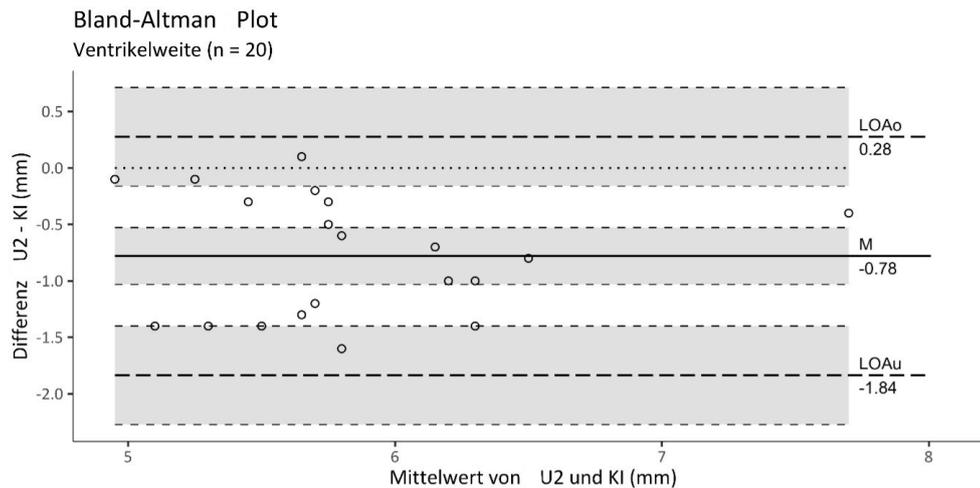


Abbildung 25: Bland-Altman-Plot für den Parameter VP, Messungen des Untersuchers 2. Der Mittelwert der Differenzen ist $M=-0,78$. Dieser Wert $\pm 1,96$ Standardabweichungen bildet das obere und untere Limit of Agreement (LoA). Die tolerable Abweichung für den Parameter VP wurde bei ≤ 2 mm festgelegt. Hieraus lässt sich das Intervall der erlaubten Abweichung berechnen ($-0,78 \pm 2,00$ mm = $-2,78$ – $1,22$ mm). Aus der Abbildung wird ersichtlich, dass die Grenzen der LoAs innerhalb dieser kritischen Grenze liegen. Somit gilt, dass KI und Untersucher 2 für den Parameter VP äquivalent messen, dies stimmt wieder mit den Ergebnissen von Untersucher 1 überein.

Cisterna magna Der Mittelwert der Differenzen ist -0,44 mm. Das Intervall der erlaubten Abweichung ist somit $0,44 \pm 2,00 \text{ mm} = -2,44 - 1,56 \text{ mm}$. Betrachtet man nun die Abbildung 26 und die LoAs (-1,84 – 0,28), wird ersichtlich, dass die gemessenen Differenzen das Intervall der kritischen Differenzen deutlich überschreiten und somit eine zu hohe Abweichung haben. Die Messung der Cisterna magna kann somit nicht als äquivalent angesehen werden. Dies deckt sich wieder mit dem Ergebnis von Untersucher 1.

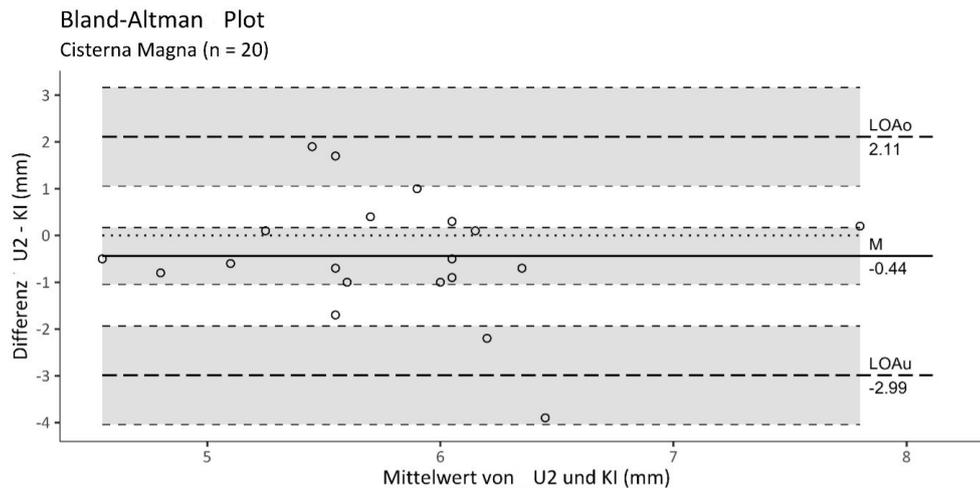


Abbildung 26: Bland-Altman-Plot für den Parameter CM, Messungen des Untersuchers 2. Der Mittelwert der Differenzen ist $M=-0,44$. Dieser Wert $\pm 1,96$ Standardabweichungen bildet das obere und untere Limit of Agreement (LoA). Die tolerable Abweichung für den Parameter CM wurde bei $\leq 2 \text{ mm}$ festgelegt. Hieraus lässt sich das Intervall der erlaubten Abweichung berechnen ($0,44 \pm 2,00 \text{ mm} = -2,44 - 1,56 \text{ mm}$). Aus der Abbildung wird ersichtlich, dass die gemessenen Differenzen dieses Intervall deutlich überschreiten. Somit ist die Messung der Cisterna magna von Untersucher 2 und KI nicht äquivalent, genau wie bei Untersucher 1.

Cerebellum Hier ist der Mittelwert der Differenzen $-0,22$. Das Intervall der erlaubten Abweichung ist $-0,22 \pm 3,5 \text{ mm} = -3,72 - 3,28 \text{ mm}$. Abbildung 27 zeigt somit, dass die LoAs $(-2,08 - 1,63)$ der gemessenen Abweichungen noch innerhalb des Intervalls der kritischen Differenzen liegen, auch mit den 95 %-Konfidenzintervallen. Die Messung ist also gleichwertig und die Messungen sind so ähnlich, dass keine klare Tendenz ins positive oder negative zu erkennen ist. Dies ist wiederum äquivalent zu den Ergebnissen des Untersuchers 1.

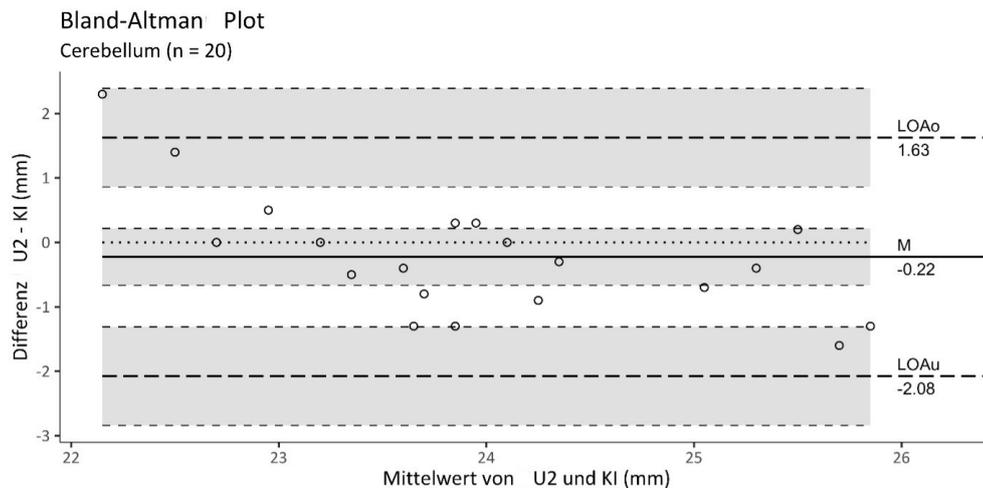


Abbildung 27: Bland-Altman-Plot für den Parameter Cerebellum, Messungen des Untersuchers 2. Der Mittelwert der Differenzen ist $M = -0,22$. Dieser Wert $\pm 1,96$ Standardabweichungen bildet das obere und untere Limit of Agreement (LoA). Die tolerable Abweichung für den Parameter Cerebellum wurde bei $\leq 3,5 \text{ mm}$ festgelegt. Hieraus lässt sich das Intervall der erlaubten Abweichung berechnen ($-0,22 \pm 3,5 \text{ mm} = -3,72 - 3,28 \text{ mm}$). Aus der Abbildung wird ersichtlich, dass die Grenzen der LoAs innerhalb dieser kritischen Grenze liegen. Somit gilt, dass KI und Untersucher 2 für den Parameter Cerebellum äquivalent messen, dies stimmt wiederum mit den Ergebnissen von Untersucher 1 überein.

5.1.11 Vergleich Untersucher 1 und Untersucher 2

Generell stimmen die Ergebnisse, die mittels deskriptiver Datenanalyse erhoben wurden, in drei von fünf Punkten überein. Für die Parameter BPD und VP zeigen sich signifikante Messunterschiede der KI sowohl bei Untersucher 1 als auch 2. Für den Parameter Cerebellum wurden bei beiden menschlichen Untersuchern keine signifikanten Messwertunterschiede im Vergleich mit der KI festgestellt. Bei Untersucher 2 werden aber im Gegensatz zu Untersucher 1 auch für den KU und die Cisterna magna keine signifikanten Messunterschiede festgestellt. Es muss aber bedacht werden, dass die Patientinnen zwar aus dem gleichen Kollektiv stammen, es sich jedoch nicht um dieselben Patientinnen handelt. Auch die Ergebnisse der Bland-Altman-Plots stimmen im Wesentlichen überein. Für die Werte BPD, VP und Cerebellum konnte bei beiden Untersuchern eine Äquivalenz zu den KI-Messungen festgestellt werden. Die Messungen der Cisterna magna stimmen hingegen bei beiden Untersuchern nicht mit denen der künstlichen Intelligenz überein. Bei den Bland-Altman-Plots für den Kopfumfang sind die Werte bei beiden Untersuchern grenzwertig, hier kann keine endgültige Aussage getroffen werden, man kann allerdings beim Vergleich der beiden Plots zu dem Schluss kommen, dass diese sich sehr ähnlich sind. Dies ist ein Indikator dafür, dass die künstliche Intelligenz auch durch andere Untersucher anwendbar ist.

5.2 Fetaler Aortenbogen

Für die Auswertung wurden die Daten und Bilder von 392 Patientinnen herangezogen. Abbildung 28 zeigt, dass es generell in 55,98 Prozent der Fälle technisch möglich war, ein Bild des fetalen Aortenbogens mit der Biplane Technik zu erzeugen.

	Aortenbogen darstellbar	Aortenbogen nicht darstellbar	Abgangsgefäße darstellbar	Abgangsgefäße nicht darstellbar
Mehr als ein Bild	77,37 % (188)	22,63 % (55)	43,62% (106)	56,38% (137)
Ein Bild	78,05 % (32)	21,95 % (9)	31,71% (13)	68,29% (28)
Keine Bilder	0 % (0)	100% (109)	0% (0)	100% (109)
Insgesamt (inklusive 0 Bilder)	55,98 % (220)	44,02 % (173)	30,28% (119)	69,72% (274)

Abbildung 28: Darstellbarkeit des Aortenbogens und der Abgangsgefäße mittels Biplaner Bildgebung. Wenn Bilder aufgenommen werden konnten, gelang die Darstellung des Aortenbogens in knapp 78 % der Fälle. Die Abgangsgefäße konnten deutlich seltener dargestellt werden (43,62 % wenn mehrere Bilder vorlagen, bei nur einem Bild in 31,71 % der Fälle). In 109 Fällen konnte überhaupt kein Bild mit dem Biplane Modus aufgenommen werden.

In 109 Fällen (27,81 %) konnte kein Bild mit der Biplane Technik generiert werden. Von den Fällen, bei denen die Technik angewandt werden konnte und ein oder mehrere Ultraschallbilder erzeugt wurden, liegt die Erfolgsrate der Darstellbarkeit des Aortenbogens bei 78,05 Prozent im Falle eines einzelnen aufgenommen Bildes und 77,37 Prozent wenn mehrere Bilder aufgenommen wurden. Die drei abgehenden Gefäße konnten deutlich seltener dargestellt werden, auf den vorhandenen Bildern in 43,62 Prozent der Fälle. Im Rahmen der Vorbereitung zu dieser Studie wurde angenommen, dass bei adipösen Patientinnen oder einer ungünstigen fetalen Lage intrauterin (hiermit sind die Querlage und die dorsoanteriore Stellung des Feten gemeint) die Wahrscheinlichkeit sinkt, dass die Biplane Technik angewandt werden kann, da diese Kriterien generell als schwierige Schallbedingungen angesehen werden.

Um die Bedeutung des BMI in Bezug auf die Anwendbarkeit der Biplane Technik darzustellen wurden drei Gruppen gebildet:

1. Bilder vorhanden und darstellbar
2. Keine Bilder
3. Bilder vorhanden aber gesuchte Strukturen (der fetale Aortenbogen und gegebenenfalls zusätzlich die Abgangsgefäße) nicht darstellbar

Eine grafische Darstellung der Verteilung zeigt Abbildung 29. Die Wahrscheinlichkeitsdichte der Zufallsvariablen „BMI“ für die drei verschiedenen Gruppen ist auf der y-Achse angegeben und wurde mithilfe eines Kerndichteschätzers mit Epanechnikov-Kern (Bandbreite $h=2$) anhand der vorliegenden Stichproben geschätzt, die x-Achse zeigt den BMI. Man kann die Fläche unter den drei Kurven als kumulierte Wahrscheinlichkeit für die Zufallsvariable (hier der BMI) sehen, die in Summe 1 ergeben muss. Die Wahrscheinlichkeit, dass jemand aus der Gruppe 'Bilder vorhanden und darstellbar' einen BMI über 30 hat, ist also sehr gering, weil die Fläche unter der orangenen Linie für $BMI \leq 30$ deutlich größer ist, als für $BMI > 30$.

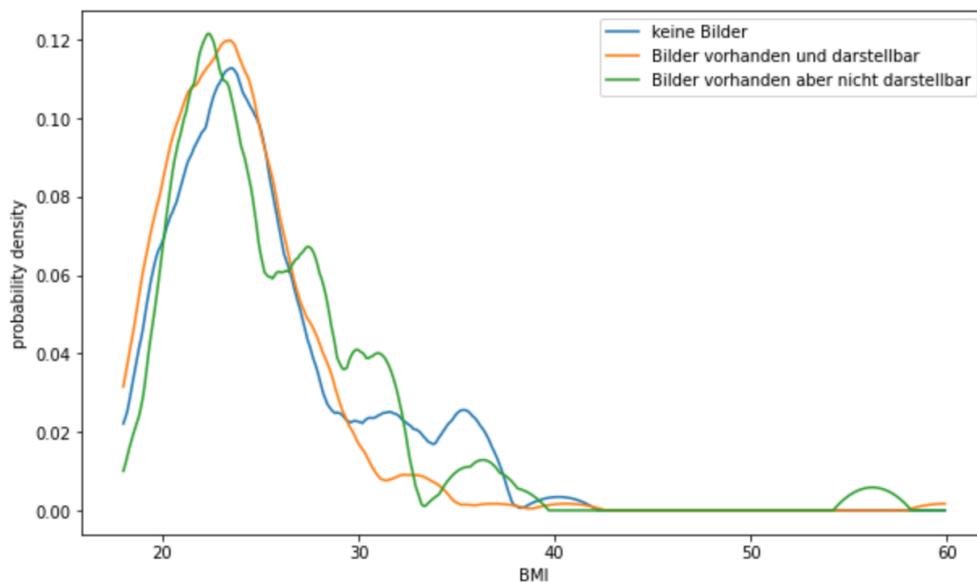


Abbildung 29: Grafische Darstellung der Verteilung der Messungen in Bezug auf den maternalen BMI. Die Abbildung zeigt, dass die Wahrscheinlichkeit bei adipösen Patienten ($BMI > 30$) erhöht ist, überhaupt kein Bild generieren zu können (blaue Linie). Ebenfalls kommt es zu einer Häufung von nicht darstellbaren Strukturen auf vorhandenen Bildern im oberen BMI-Bereich (grüne Linie). Die Abbildung zeigt aber auch, dass bei normalgewichtigen Frauen nicht immer Bilder generiert werden konnten.

Gruppe	BMI Mittelwert	BMI SD
Keine Bilder	24,58	4,56
Bilder darstellbar	23,85	4,34
Bilder nicht darstellbar	25,57	5,73

Tabelle 12: BMI-Verteilung über die drei Gruppen. Bei den darstellbaren Bildern ist er am niedrigsten. Am höchsten ist er bei den Bildern, bei denen es zwar technisch möglich war, ein Bild zu generieren, aber die Strukturen nicht erkennbar waren.

Die meisten Daten liegen im Bereich des medianen BMI von 23,7 vor. Die Abbildung zeigt, dass es bei adipösen Patientinnen zu einer Häufung von nicht darstellbaren Bildern kommt, also das generell kein orthogonales Bild mit Hilfe der Matrixsonde erzeugt werden konnte (blaue Linie). Wenn bei adipösen Patientinnen Bilder aufgenommen werden konnten, kommt es ebenfalls gehäuft vor, dass die gesuchten Strukturen nicht erkennbar waren (grüne Linie). Abbildung 30 zeigt ebenfalls die Datenverteilung in Bezug auf den maternalen BMI und stellt die Ausreißer im sehr adipösen Bereich dar. Zwei Patientinnen weisen einen sehr hohen BMI im Sinne einer Adipositas permagna auf. Bei beiden Patientinnen wurden mehrere Bilder mittels Biplaner Bildgebung aufgenommen, die Darstellung des fetalen Aortenbogens war in einem der beiden Fälle sogar möglich. Eventuell ist dies durch eine in diesem Fall vorliegende günstige fetale Lage zu erklären, der Fet befand sich zum Zeitpunkt der Untersuchung in dorsoposteriorer Schädellage. Im anderen Fall lag eine Querlage vor, diese Lage gilt allgemein auch bei normgewichtigen Frauen als schwierig zu untersuchen. Insgesamt sind zwei extrem adipöse Patientinnen aber zu wenig, um eine statistisch signifikante Aussage treffen zu können. Die oben genannten drei Gruppen werden nun mit Hilfe des t-Testes verglichen. Da der t-Test auf der Annahme gleicher Varianzen beruht, wird zunächst der Levene-Test durchgeführt. Hier wurde ein p-Wert von 0,101 berechnet, womit man die Nullhypothese der Varianzgleichheit unter einem Signifikanzniveau von 5 % nicht verwerfen kann. Damit wird die Grundvoraussetzung der Varianzgleichheit des t-Tests erfüllt. Auf die drei Gruppen wird paarweise jeweils der Zweistichproben-t-Test für unabhängige Stichproben zum Vergleich des Mittelwerts in Bezug auf den BMI angewandt. Als H0 wird angenommen das die Mittelwerte der Grundgesamtheiten gleich sind ($\mu_1 = \mu_2$). Die H1 ist entsprechend, dass sich die Mittelwerte der Grundgesamtheiten unterscheiden ($\mu_1 \neq \mu_2$).

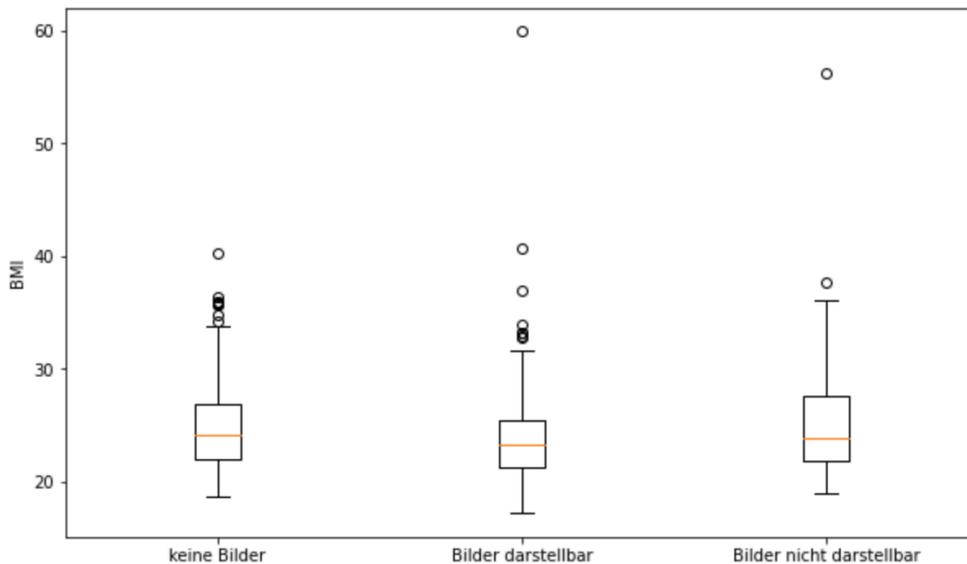


Abbildung 30: Boxplot zur Darstellung der Verteilung der Messungen in Bezug auf den maternalen BMI, der Mittelwert ist farbig abgehoben. Bei den darstellbaren Bildern ist der Mittelwert des BMI nur minimal kleiner. Die Unterschiede in den Mittelwerten sind generell sehr gering, was dafür spricht, dass der BMI zwar einen signifikanten, aber dennoch geringen Effekt auf die Darstellung des Aortenbogens mittels Biplane Technik hat.

Vergleich der Gruppen 'Bilder darstellbar, Bilder nicht darstellbar'

$$statistic = -2.5918993697705535 \quad pvalue = 0.010041983418379247$$

Bei diesen Gruppen kann die Nullhypothese abgelehnt werden, da der p-Wert von 0.010 deutlich kleiner als 0.05 ist. Man kann deswegen davon ausgehen, dass die Mittelwerte der Grundgesamtheiten dieser Gruppen verschieden sind. Dies zeigt, dass bei adipösen Patientinnen die Wahrscheinlichkeit höher ist, wenn Bilder mittels Matrixsonde dargestellt werden können, die gesuchten Strukturen auf den vorliegenden Bildern nicht erkennen zu können.

Vergleich der Gruppen 'Bilder darstellbar, keine Bilder'

$$statistic = -2.516411768691102 \quad pvalue = 0.012333894841807979$$

Bei diesen Gruppen, kann die Nullhypothese ebenfalls abgelehnt werden, da der p-Wert von 0.012 deutlich kleiner als 0.05 ist. Man kann daher ebenfalls davon ausgehen, dass die Mittelwerte der Grundgesamtheiten dieser Gruppen verschieden sind. Dies zeigt, dass

bei adipösen Patientinnen die Wahrscheinlichkeit signifikant höher ist, überhaupt keine Bilder des fetalen Aortenbogens mit Hilfe der Biplane Technik generieren zu können.

Vergleich der Gruppen 'Bilder nicht darstellbar, keine Bilder'

$$statistic = 0.49355708787912966 \quad pvalue = 0.6222523169834753$$

Hier kann die Nullhypothese nicht abgelehnt werden. Der p-Wert liegt mit 0.622 weit über 0.05. Dies bedeutet, dass nicht ausgeschlossen werden kann, dass die Mittelwerte der Grundgesamtheiten in Bezug auf die Variable BMI gleich sind. Bei den Frauen, bei denen zwar Bilder vorlagen, aber der Aortenbogen nicht darstellbar war, und bei den Frauen, bei denen überhaupt keine Bilder gemacht werden konnten, lässt sich kein Unterschied in Bezug auf die Verteilung des BMI feststellen. Dies ist grafisch auch im Boxplot dargestellt, da die Mittelwerte für diese beiden Gruppen auf einer Linie liegen.

5.2.1 Zusammenfassung der Ergebnisse in Bezug auf den maternalen BMI

Generell kann postuliert werden, dass ein Zusammenhang zwischen der erfolgreichen Auswertung und dem BMI besteht. Ein höherer BMI erhöht demnach die Wahrscheinlichkeit, dass die Methode nicht angewendet werden kann. Somit muss davon ausgegangen werden, dass die Matrixsonde und die Biplane Technik auch Limitationen bei adipösen Frauen hat. Allerdings ist der Effekt trotz signifikanter Auswirkung insgesamt klein wenn man die Unterschiede in den Mittelwerten der Gruppen im Box-Plot in Abbildung 30 vergleicht. Hier ist der Mittelwert bei den darstellbaren Bildern zwar geringer als in den anderen zwei Gruppen, allerdings nicht wesentlich. Dies zeigt auch Tabelle 12. Die Box der nicht darstellbaren Bildern ist aber im oberen BMI Bereich höher als bei den anderen beiden Gruppen. Der BMI kann aber nicht als die einzige Variable angesehen werden, da auch im unteren BMI-Bereich Bilder nicht darstellbar waren. Zur Ursachenfindung wird daher auch nach einem Zusammenhang zwischen fetaler Lage und Darstellbarkeit gesucht.

5.2.2 Vergleich der Darstellbarkeit des Aortenbogens in Bezug auf die intrauterine Lage

Der Chi-Quadrat-Test wird angewandt, um herauszufinden, ob die intrauterine Lage des Feten einen Einfluss auf die Darstellbarkeit des Aortenbogens mittels Matrixsonde hat. Der Chi-Quadrat Test ist ein Hypothesentest, der verwendet wird um Zusammenhänge zwischen zwei kategorischen Variablen festzustellen. Da sowohl die Lage des Feten als auch die Darstellbarkeit des Aortenbogens kategorische Variablen sind, ist der Chi-Quadrat Test geeignet. Die Lage wird zunächst in die drei Grundlagen Schädellage, Beckenendlage und Querlage eingeteilt. Hier werden wieder die oben genannten drei Gruppen a) Bilder vorhanden und darstellbar, b) Bilder vorhanden und nicht darstellbar und c) Keine Bilder benutzt. Tabelle 13 gibt einen Überblick über die Häufigkeit der verschiedenen Lagen in Bezug auf die Darstellbarkeit der Bilder. Der Chi-Quadrat-Test testet, ob die Zufallsvariablen unabhängig voneinander sind. Als H₀-Hypothese wird formuliert, dass die fetale Lage und die Darstellbarkeit des Aortenbogens unabhängig voneinander sind.

$$X^2 \text{Koeffizient}_{df=4,1-\alpha=0,95} = 9,49$$

$$\text{Teststatistik} = 5,57$$

Die Teststatistik wurde mit der Chi-Quadratverteilung (4 Freiheitsgrade) verglichen. Da $5,57 < 9,49$ ist, kann die Nullhypothese nicht ausgeschlossen werden. Die Daten deuten also auf keinen systematischen Zusammenhang zwischen fetaler Lage und der Abbildbarkeit des Aortenbogens hin.

Lage des Feten	Bild(er) vorhanden und darstellbar	Bild(er) vorhanden und nicht darstellbar	Keine Bilder	Gesamt	Relative Häufigkeit in Prozent
BEL	84	21	39	144	36,73
QL	22	13	11	46	11,73
SL	114	30	58	202	51,53
Gesamt	220	64	108	392	100
Relative Häufigkeit	56,12	16,33	27,55	100	

Tabelle 13: Übersicht über die intrauterinen Lagen und ihre Häufigkeitsverteilung in Bezug auf die Darstellbarkeit des Aortenbogens mittels Biplane. Trotz vorliegender Querlage war die Darstellung in 22 von 46 Fällen (knapp 48 %) möglich.

5.2.3 Vergleich der Darstellbarkeit des Aortenbogens in Bezug auf die dorsoanteriore Stellung

Als separater Parameter wurde betrachtet, ob der Rücken des Kindes vorne lag, eine sogenannte dorsoanteriore Stellung. Hier wurde unterschieden in a) ja oder b) nein. Eine Übersicht bietet Tabelle 14. Die niedrigere Gesamtzahl von 303 untersuchten Feten erklärt sich durch die Tatsache, dass nicht bei allen Untersuchungen weitere Angaben bezüglich der Stellung gemacht wurden. Bei einigen wurde nach der Differenzierung in Schädellage, Beckenendlage oder Querlage nicht weiter unterteilt, diese Fälle finden daher in dieser Betrachtung keine Beachtung. Es wird wieder der Chi-Quadrat-Unabhängigkeitstest angewandt.

$$X^2 \text{Koeffizient}_{df=2, 1-\alpha=0,95} = 5,99$$

$$\text{Teststatistik} = 72,23$$

Als H₀-Hypothese wird formuliert, dass die dorsoanteriore Stellung und die Darstellbarkeit des Aortenbogens unabhängig voneinander sind. Die Teststatistik (Chi-Quadrat-Koeffizienz von 72,23) wurde mit der Chi-Quadratverteilung (2 Freiheitsgrade, 0.95) verglichen. Die Chi-Quadratverteilung liegt bei 5,99. Damit muss die Nullhypothese abgelehnt

werden, die Teststatistik ist deutlich größer. Das bedeutet, dass die Abbildbarkeit des Aortenbogens deutlich davon abhängt, ob eine dorsoanteriore Stellung vorliegt oder nicht. Bei dorsoanteriorer Stellung des Kindes ist die Wahrscheinlichkeit, den Aortenbogen darzustellen, somit deutlich geringer.

Dorsoanterior	Bild(er) vorhanden und darstellbar	Bild(er) vorhanden und nicht darstellbar	Keine Bil- der	Gesamt	Relative Häufigkeit in Prozent
Ja	44	27	27	98	32
Nein	168	37	0	205	68
Gesamt	212	64	27	303	100
Relative Häufigkeit	70	21	9	100	

Tabelle 14: Übersicht über die dorsoanteriore Stellung und ihre Häufigkeitsverteilung. Im Falle einer dorsoanterioren Stellung ist die Wahrscheinlichkeit, den Aortenbogen darstellen zu können, deutlich geringer.

6 Diskussion

6.1 Schlussfolgerung

Künstliche Intelligenz und fetale Biometrie Generell kann man anhand der vorliegenden Daten zu dem Schluss kommen, dass die künstliche Intelligenz im Bereich der fetalen Anatomie des Kopfes sehr gute Ergebnisse erzielt, wenn auch noch nicht für alle Parameter. Äquivalenz konnte für die Parameter Biparietaler Durchmesser, Hinterer Hirnseitenventrikel und Cerebellum aufgezeigt werden. Für den Parameter Kopfumfang muss festgestellt werden, dass keine eindeutige Aussage über die Äquivalenz der Messungen gemacht werden kann. Die Differenzen liegen noch sehr knapp im Intervall der erlaubten Abweichung, dies ändert sich jedoch wenn man das 95 %-Konfidenzintervall der LoAs miteinbezieht. Die Ergebnisse deuten aber darauf hin, dass bei einer größeren Stichprobe möglicherweise eine höhere Übereinstimmung gefunden werden könnte. Bei beiden Untersuchern kann keine statistisch ausreichende Äquivalenz für die Messung der Cisterna magna postuliert werden. Dies zeigen sowohl die Bland-Altman-Plots für beide Untersucher als auch die Interklassen-Korrelation. Die Ergebnisse dieser Messungen werden in 6.1.2 getrennt betrachtet. Ebenfalls wird der Parameter VP getrennt betrachtet in 6.1.1. Betrachtet man alle Werte zusammen, gelang der KI nur in 6 % der Fälle keine Darstellung oder es wurden stark divergierende Messungen gemacht. Diese Ergebnisse decken sich mit den Ergebnissen einer Studie von Nagayasu et al., die an einem kleineren Kollektiv von 50 Frauen die manuelle Messung mit der automatischen Messung des SonoCNS-Programmes verglichen haben, dort lag die Erfolgsrate mit 85 % aber unter der hier erzielten [41]. Auch andere Arbeiten kommen zu dem Ergebnis, dass das Programm SonoCNS gute Ergebnisse hinsichtlich der Genauigkeit der Biometriemessungen liefert [42]. Eine Arbeit von Salim et al. hat beim Vergleich der automatischen Messung mittels Algorithmus am unbewegten 2D Bild und der manuellen Messung eine mittlere Abweichung von -2,1 mm für den Kopfumfang feststellen können [43], mit der Volumenmessung in dieser Studie liegt die Abweichung generell in der deskriptiven Analyse mit -1,56 mm darunter. Dies könnte für eine verbesserte Technik durch 3D/4D Volumenaufnahmen sprechen. Allerdings ist der Mittelwert der deskriptiven Analyse nur bedingt vergleichbar, da es bei einigen Messungen zu größeren Abweichungen kommt wie im Bland-Altman-Plot grafisch dargestellt ist. Von Beginn an war nicht zu erwarten, dass die künstliche

Intelligenz exakt identische Messwerte im Vergleich zum menschlichen Untersucher erhebt, da auch bei zwei menschlichen Untersuchern desselben Feten kaum eine komplette Übereinstimmung erreicht wird. Dies zeigt auch die Studie von Sarris et al., die Autoren gaben die Interobserver-Variabilität des Kopfumfangs mit ± 9.8 mm an [39]. Die fehlende komplette Übereinstimmung zwischen zwei Untersuchern spricht auch dafür, dass die künstliche Intelligenz trotz einer Abweichung über 2 mm bei CM und VP gute Ergebnisse erzielen kann. Möglicherweise ist die Grenze zu konservativ gesetzt. Betrachtet man die in Tabelle 9 aufgeführten 49 Fehler für VP, verringert sich die Fehlerquote schon um 15 Fälle, wenn man die Grenze bei $\leq 2,1$ mm setzt. Während der Vorbereitungen zu dieser Studie wurde vermutet, dass bestimmte Konstellationen wie zum Beispiel eine adipöse maternale Bauchdecke oder eine ungünstige fetale Lage die Messungen der KI negativ beeinflussen. Dies konnte nicht bestätigt werden. Eine Vermutung war, dass die künstliche Intelligenz die Parameter bei Beckenendlage am besten messen kann, da der fetale Kopf gut mit dem Schallkopf erreicht werden kann. Dies konnte ebenfalls nicht bestätigt werden, bei vorliegender Beckenendlage konnte sogar eine statistisch signifikante höhere Differenz für den Parameter BPD festgestellt werden, wenn auch mit kleinem Effekt. Es konnten nur geringe Abweichungen zwischen Untersucher 1 und KI in Bezug auf die fetale Lage allgemein, die Lage der Plazenta sowie das Gewicht des Kindes gefunden werden. Eine Tendenz zur Steigerung der Messdifferenz bei steigendem Gewicht wurde für die Messung des Ventrikels gefunden. Dies deckt sich mit den Ergebnissen der Studie von Sarris et al., die eine Zunahme der Intra- und Interobserver-Variabilität bei zunehmendem Gestationsalter, und somit im Umkehrschluss bei Zunahme des Gewichtes, finden konnte [39]. Andererseits bleibt unklar, wieso die steigende Messdifferenz ausgerechnet für den Ventrikel nachgewiesen werden konnte, da dieser im Gegensatz zu den Parametern BPD, KU und Cerebellum im Laufe einer normalen Schwangerschaft nicht zunimmt sondern konstant ≤ 10 mm bleibt. Bei genauerer Betrachtung des Ergebnisses fällt auf, dass der Wert zwar statistisch signifikant ist, aber der Effekt klein ist, da das Modell nur 2,4 % der Gesamtvarianz der Messdifferenz erklärt. Es könnte sich somit auch um ein zufälliges Ergebnis handeln, wie in 6.1.1 aufgezeigt wird. Der fehlende Einfluss der maternalen Adipositas auf die Genauigkeit der Messung mittels künstlicher Intelligenz ist sehr interessant und widerspricht der klinischen Erfahrung. Andere neue Studien konnten allerdings ebenfalls zeigen, dass Übergewicht keinen negativen Einfluss auf die Messung der fetalen

Biometrie hat, wenn erfahrene Untersucherinnen und Untersucher den Ultraschall durchführten [44]. Eine weitere Theorie unabhängig von der Erfahrung des Untersuchers ist die gute Qualität der neuen Ultraschallgeräte und die verbesserte Eindringtiefe von Schallköpfen wie der hier benutzten elektronischen Matrixsonde. Somit kann postuliert werden, dass die künstliche Intelligenz ohne wesentliche Einschränkungen bei allen Patientinnen aus unserem Kollektiv angewandt werden kann. Da das Programm „SonoCNS“ aber nicht in allen Fällen gute Ergebnisse erzielt hat, ist es wichtig zu erwähnen, dass der Untersucher oder die Untersucherin sich keinesfalls nur auf die künstliche Intelligenz verlassen darf, sondern visuell die Bilder kontrollieren muss um etwaige Abweichungen korrigieren zu können. Eine kritische Überprüfung der Bilder der KI sowie der manuell erstellten Messungen ist stets sinnvoll, da Fehlmessungen klinisch relevant sind und zu überflüssigen Interventionen führen können. Schon im Jahr 2002 postulierten Dudley et. al., dass Training und Qualitätsmanagement für die Messung der fetalen Biometrie wichtig sind [45]. Dies ist auch auf die Messung mit Hilfe der KI übertragbar. Generell kann die Ähnlichkeit der Datenanalyse und der Bland-Altman-Plots von beiden menschlichen Untersuchern im Vergleich mit der KI als Anhaltspunkt dienen, dass „SonoCNS“ auch von anderen Ärztinnen und Ärzten sinnvoll angewandt werden kann. Der Hersteller vermarktet das Produkt SonoCNS in einem Video mit den Worten, dass die Untersuchung des fetalen Gehirns einfacher, schneller und reproduzierbarer gemacht wird und der Tastenanschlag um 82 % verringert wird [46]. Limitationen gibt die Firma nicht an, es wird lediglich erwähnt, dass der beste Untersuchungszeitraum zwischen 18 und 26 SSW ist und eine gute Einstellung der transthalämischen Ebene wichtig ist, je besser diese Einstellung ist, desto besser sind die Ergebnisse [46].

Matrixsonde und fetaler Aortenbogen Die Biplane Technik konnte in dieser Studie nur in etwas mehr als der Hälfte der Fälle angewandt werden. Auch wenn optimale Untersuchungsbedingungen vorlagen gelang die Biplane Darstellung des Aortenbogens nicht zwangsläufig. Positiv zu sehen ist allerdings, dass es in fast 80 % der Fälle in denen ein Bild mit dieser Technik aufgenommen werden konnte, auch zur Darstellung der gesuchten Struktur kam. Die Darstellung der drei am Aortenbogen abgehenden Gefäße gelang insgesamt nur in etwa 30 % der Fälle. Allerdings wurde die Darstellbarkeit der Abgangsfäße nur mit „Ja“ beantwortet wenn alle 3 Gefäße auf dem Ultraschallbild zu sehen

waren, nicht wenn zum Beispiel nur ein oder zwei Gefäße sichtbar waren. Eine mögliche Ursache der schlechten Darstellbarkeit aller drei Gefäße ist, dass nicht alle im selben Winkel abgehen. Das bedeutet, dass auf dem mittels Biplane Technik aufgenommenen orthogonalen Bild eventuell nicht immer die richtige Ebene für die Abgangsgefäße gezeigt wird. Die Untersuchung des fetalen Herzens ist aufgrund der ständigen Bewegung sowohl des Herzens als auch des Kindes generell anspruchsvoll. Neue Methoden zur Darstellung der bekannten Strukturen müssen daher regelmäßig trainiert und angewandt werden, und es ist denkbar, dass sich dann die Darstellungshäufigkeit noch deutlich steigern lässt. Des Weiteren ist hervorzuheben, dass sich auch unter normalen Bedingungen ohne Nutzung des Biplane Modus die Aorta im Längsschnitt nicht immer darstellen lässt. In dieser Studie war die Darstellung des Aortenbogens bei adipösen Patientinnen signifikant erschwert. Generell nimmt weltweit der Anteil an adipösen Patienten zu. Trotz verbesserter Eindringtiefe des Matrixschallkopfes muss davon ausgegangen werden, dass auch der Ultraschall bei extrem adipöser Bauchdecke Limitationen hat und gerade in diesem Kollektiv die Wahrscheinlichkeit, Fehlbildungen zu übersehen, weiterhin erhöht ist. Dies wurde schon in früheren Studien gezeigt [47]. Speziell auf Untersuchungen des fetalen Herzens bezogen, war bei extrem adipösen Frauen die Darstellung des 4-Kammer-Blickes und des Ausflusstraktes deutlich erschwert im Vergleich zu normalgewichtigen Schwangeren [48]. Zwei Studien aus den USA kommen zu dem Schluss, dass zur Erlangung suffizienter Ultraschallbilder bei adipösen Frauen oftmals mehrere Ultraschalluntersuchungen notwendig sind [49] [50], sodass eine erneute Einbestellung der Patientin hilfreich sein könnte, wenn beim ersten Besuch der Aortenbogen nicht adäquat dargestellt werden kann. In vielen Fällen lieferte die Matrixsonde aber auch im oberen BMI-Bereich gute Bilder, sodass im Vergleich zu älteren Geräten deutliche Fortschritte gemacht werden konnten und viele Feten somit adäquat beurteilt werden können. Somit lässt sich aus den Ergebnissen zwar generell ableiten, dass Limitationen im oberen BMI-Bereich vorliegen, allerdings keine deutlichen. Es kann spekuliert werden, dass durch die gute Eindringtiefe der Matrixsonde die Ultraschalldiagnostik adipöser Schwangerer verbessert wird, vor allem, wenn die Patientin erneut einbestellt wird im Falle einer nicht adäquaten Darstellung. Die Ergebnisse im Bereich der dorsoanterioren Stellung des Feten zeigen signifikant, dass der Aortenbogen bei diesen Feten schlecht darstellbar ist. Im Gegensatz zum BMI sind hier die Ergebnisse deutlich. Auch in diesem Fall kann eine erneute Untersuchung sinnvoll

sein um auf auf eine spontane Drehung des Feten in eine günstigere Position zu warten. Die dorsoanteriore Stellung wird auch in Zukunft zur schlechteren Darstellung der meisten Strukturen führen, da auch verbesserte Ultraschalltechnik nicht verhindern kann, dass Knochen einen Schallschatten im Ultraschall produzieren und Wirbelsäule und Rippen vor den inneren Organen liegen.

6.1.1 Diskussion der Ergebnisse am Beispiel des Hinteren Hirnseitenventrikels (VP)

Hier waren die meisten fehlenden Messungen der künstlichen Intelligenz zu verzeichnen. Auch in der oben erwähnten Studie von Nagasayu et al. wurden die geringsten Erfolge der künstlichen Intelligenz bei der Messung des Ventrikels verzeichnet [41]. Da es sich in dieser Studie um eine relativ geringe Anzahl von 18 Fällen handelt, kann keine Aussage zu eventuellen Prädiktoren für eine fehlende Messung getroffen werden. Hierzu sind weitere Studien mit dem Einschluss von mehr Patientinnen notwendig. Zur Ursachenfindung der fehlenden Messungen wurden die Volumenaufnahmen dieser 18 Patientinnen retrospektiv erneut visuell betrachtet. Hierbei fällt auf, dass nur in 6 Fällen der Ventrikel gut sichtbar ist, in den anderen 12 Fällen ist der Ventrikel nicht optimal dargestellt, ein Beispiel zeigt Abbildung 31. Im rechten oberen Bild ist zwar der Plexus choroideus sichtbar, der Ventrikel erscheint aber schräg angeschnitten, sodass die Vermutung nahe liegt, dass die KI die Ebene nicht korrekt darstellen konnte und somit den Ventrikel nicht als Ventrikel identifizieren konnte. Somit bleiben 6 Fälle, bei denen unklar ist, warum das Programm „SonoCNS“ die Messung nicht durchgeführt hat. Hier stellt sich die Frage, ob das Programm noch dahingehend verbessert werden kann, den Ventrikel in allen Fällen zu erkennen. Des Weiteren könnte es sinnvoll sein, wenn das Programm dem Untersucher beim Nichterkennen des Ventrikels einen Warnhinweis anzeigt, damit die Ventrikelregion nochmals aufmerksam betrachtet wird und etwaige Fehlbildungen ausgeschlossen werden können. In den 353 Fällen, in denen sowohl Untersucher 1 als auch die KI einen Messwert für VP erhoben haben, zeigt sich in der Datenanalyse ein signifikanter Messunterschied, die manuelle Messung ist im Schnitt um 1,06 mm kleiner. Dieser Wert liegt unter der vorher als tolerabel festgelegten Messdifferenz von ≤ 2 mm. Somit ist der Wert zwar statistisch signifikant, hat aber für die Praxis keine Auswirkungen. Dies zeigt auch die Auswertung der Bland-Altman-Plots, bei denen gezeigt werden kann, dass Mensch und

KI äquivalent arbeiten. Diese Äquivalenz gilt auch für Untersucher 2, der im Schnitt -0,79 mm weniger gemessen hat als die KI. Weiterhin ist VP ist der einzige Parameter, bei dem zwei Prädiktoren einen Einfluss auf die Messdifferenz hatten, nämlich die Seitenwandplazenta und das fetale Gewicht. Beide Effekte sind aber sehr gering. Dies relativiert sich wieder im Rahmen der multiplen Regression, sodass davon ausgegangen werden kann, dass keiner der Prädiktoren einen Einfluss auf die Messung mittels künstlicher Intelligenz hat und das Programm auch für den Parameter VP theoretisch bei allen Schwangeren angewandt werden kann.

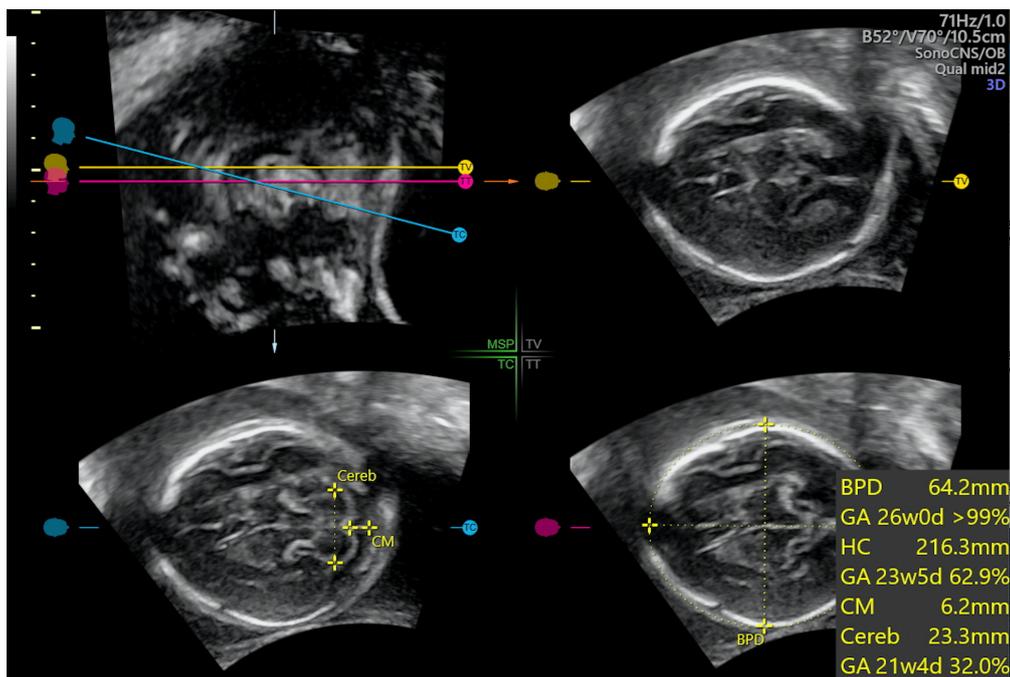


Abbildung 31: Volumenaufnahme ohne Messung des Ventrikels. Im rechten oberen Bild ist die transventrikuläre Ebene mit dem Plexus choroideus dargestellt. Der Ventrikel erscheint schräg angeschnitten, eventuell hat die KI ihn daher nicht als Ventrikel erkannt und nicht ausgemessen.

6.1.2 Diskussion der Ergebnisse am Beispiel der Cisterna magna (CM)

Die Cisterna magna ist eine mit Liquor gefüllte Struktur in der hinteren Schädelgrube. Die Untersuchung der hinteren Schädelgrube spielt in der Pränataldiagnostik eine große Rolle, da Auffälligkeiten in dieser Region ein Hinweis auf eine Vielzahl von Erkrankungen sein können. Ein Durchmesser von >10 mm gilt unabhängig von der Schwangerschaftswoche als auffällig. Die durchschnittliche Weite der Cisterna magna beträgt 5 ± 3 mm [51].

Die Fehlerwertanalyse zeigt in fast 24 % der Fälle (93 Patientinnen) Fehlmessungen der KI für den Parameter CM. Die hohe Anzahl an Fehlern spiegelt sich auch in den Bland-Altman-Plots wider, hier zeigt sich, dass sowohl die Messungen von Untersucher 1 als auch die Messungen von Untersucher 2 nicht äquivalent zur künstlichen Intelligenz sind. Hier stellt sich die Frage, ob die im Vorfeld gewählte tolerable Abweichung von ≤ 2 mm sinnvoll gewählt ist. Interessant wäre auch hier die Interobserver-Variabilität, also ob zwei menschliche Untersucher bei demselben Feten äquivalent gemessen hätten wenn man die gesetzte Grenze von ≤ 2 mm beibehält. Setzt man im Bland-Altman-Plot die kritische Differenz nur um 1 mm höher bei 3 mm an, liegt das Intervall der erlaubten Abweichung bei $-3,87$ - $2,13$ mm, somit würden die LoAs ($-3,56$ - $1,82$) noch unterhalb dieses Intervalls liegen und die Methode könnte als äquivalent betrachtet werden.

Für die KI spricht außerdem, dass trotz fehlender Äquivalenz der Messungen keine Nachteile für die jeweiligen Feten entstanden wären, da alle Werte unter 10 mm waren und somit per definitionem nicht pathologisch waren, genau wie beim menschlichen Untersucher. Es kann allerdings keine Aussage darüber getroffen werden, ob die KI bei Feten mit Fehlbildungen in diesem Bereich ebenfalls funktioniert hätte, da diese aus der Studie ausgeschlossen wurden. Ausgehend von der Messdifferenz scheint die KI auch keine größeren Auffälligkeiten gemessen zu haben, da in diesem Kollektiv nur in 3 Fällen eine Abweichung ≥ 4 mm gemessen wurde, in keinem Fall eine Abweichung $>4,4$ mm. Somit könnte man zu dem Schluss kommen, dass die KI ihre Aufgabe, valide Werte zu erlangen, erfüllt hat. Betrachtet man allerdings die Messergebnisse für den Parameter CM für Untersucher 1 und KI getrennt und lässt die Messdifferenz außer acht, zeigt sich, dass bei der KI 44 Messwerte ≥ 8 mm und $\leq 9,8$ mm sind. Es liegen also viele Werte im oberen Normbereich vor, wohingegen der menschliche Untersucher nur 2 Messungen ≥ 8 mm gemacht hat (8 mm und 8,4 mm). Dies spricht gegen die Verwendung der KI in diesem Bereich, da laut der Studie von Mahony et al. die Werte im Durchschnitt nicht über 8 mm liegen [51] und somit davon auszugehen ist, dass die KI im Bereich der Cisterna magna zu groß misst. Eine Möglichkeit zur Verbesserung der KI-Messung wäre die Nachjustierung des Programms durch veränderte Setzung der Caliper. Bei der retrospektiven Betrachtung der Bilder fällt nämlich auf, dass der menschliche Untersucher den Caliper jeweils an den Innenseiten der Cisterna magna ansetzt, die KI aber an der Außenseite. Somit könnte die Tendenz zur größeren Messung der KI erklärt werden. Weiterhin fällt bei der

Betrachtung einiger Bilder mit einer Cisterna magna im oberen Bereich wie in Abbildung 32 dargestellt auf, dass die hintere Schädelgrube von der KI nicht optimal dargestellt wurde⁶, da auch das Kleinhirn nicht optimal zu erkennen ist. Somit ist die Fehlmessung (die Differenz zum Untersucher 1 betrug in diesem Fall -2,6 mm) am ehesten der inkorrekten Ebene und Calipersetzung geschuldet und könnte durch weitere Programmupdates verbessert werden. Auf diesem Bild ist auch die Ebene des Kopfumfanges nicht korrekt dargestellt, sodass sich auch hier eine große Abweichung zu Untersucher 1 ergab. Somit ist es sicherlich sinnvoll, bei einer auffälligen Abweichung der KI auch die anderen Ebenen derselben Messung genau zu betrachten um weitere Fehler erkennen zu können. KI-Werte ≥ 8 mm sollten ebenfalls genau betrachtet werden.

⁶Die korrekte Darstellung zeigt Abbildung 4

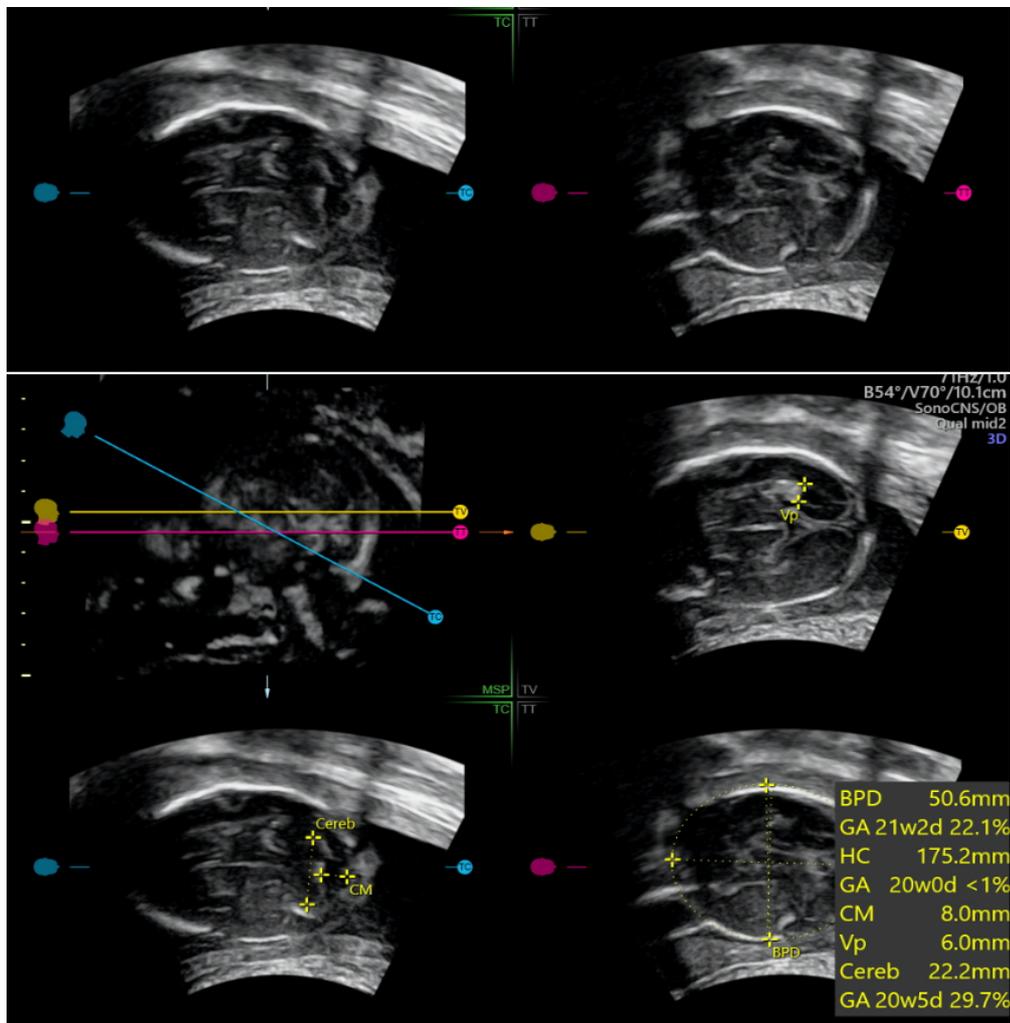


Abbildung 32: Auf den oberen zwei Bildern sieht man die von der KI dargestellten Ebenen, links oben die transzerebelläre, rechts oben die transthalämische Ebene noch ohne Messung. Auf den unteren Bildern sind die automatisch erstellten Messungen zu sehen. Bei genauer Betrachtung fällt auf, dass die KI die hintere Schädelgrube nicht optimal darstellen konnte, da das Cerebellum in der transzerebellären Ebene nicht gut erkennbar ist. In diesem Fall betrug die Messdifferenz der Cisterna magna -2,6 mm, liegt also oberhalb der tolerablen Abweichung und könnte durch die schlechte Bildqualität erklärbar sein. Die Messdifferenz des Cerebellums betrug allerdings nur 1,3 mm und liegt somit im Rahmen.

6.2 Limitation der Methode

Künstliche Intelligenz und fetale Biometrie In der Studie wurde hauptsächlich für jeden Parameter die Messwertdifferenz der beiden Rater überprüft. Es erfolgte keine Einteilung der Bilder nach Qualität, die Bilder der KI wurden zwar visuell beurteilt, die Messebenen des menschlichen Untersuchers allerdings nur in Einzelfällen. Die Überprüfung ist bei Untersuchern mit langjähriger Erfahrung auch nicht das Studienziel gewesen, allerdings kann so nicht in allen Fällen bei Abweichungen über die tolerierten Grenzen hinaus eine Aussage getroffen werden, ob letztlich die künstliche Intelligenz oder der menschliche Untersucher eine Ebene nicht optimal dargestellt hat, da auch kleine Unterschiede im Millimeterbereich einen signifikanten Unterschied bedeuten können. Des Weiteren wurden alle Feten mit Fehlbildungen und Mehrlinge ausgeschlossen, somit kann keine Aussage getroffen werden, ob in diesen Fällen die künstliche Intelligenz auch zuverlässige Ergebnisse liefern könnte. Eine Studie von Welp et al. kommt auch zu dem Schluss, dass die Biometriemessungen der künstlichen Intelligenz insgesamt sehr genau sind, weitere Studien bei Feten mit Fehlbildungen aber noch fehlen [42]. Diese Studien wären auch sinnvoll, um aufzeigen zu können, ob die KI auch im Basisultraschall bei weniger geübten Untersucherinnen und Untersuchern eingesetzt werden könnte, ohne dass sich das Risiko, Fehlbildungen zu übersehen, erhöht. Des Weiteren konnte in diesem Studiendesign nur die Interobserver-Variabilität zwischen der künstlichen Intelligenz und Untersucher 1 sowie getrennt davon zwischen KI und Untersucher 2 überprüft werden. Die Ergebnisse beider menschlicher Untersucher im Vergleich mit der KI ähneln sich zwar, was für die Methode spricht, es handelte sich aber nicht um dieselben Patientinnen. Um die Interobserver-Variabilität besser beurteilen zu können, müssten zwei Untersucher und die künstliche Intelligenz denselben Feten ausmessen. Dies könnte zeigen, dass es bei allen drei Untersuchern zu Abweichungen kommt, die aber generell als akzeptabel angesehen werden könnten. Es könnte aber auch zeigen, dass die menschlichen Untersucher im Vergleich untereinander doch geringere Messunterschiede aufweisen als die KI-Messungen und spräche in dem Fall für die Fortführung der herkömmlichen manuellen Messungen. In diesem Studiendesign schneidet die künstliche Intelligenz im Vergleich zu den beiden erfahrenen Untersuchern ähnlich gut ab, wünschenswert wären aber noch weitere Studien mit weniger erfahrenen Untersucherinnen und Untersuchern, um zu zeigen, dass die künstliche Intelligenz die Diagnostik generell verbessern kann beziehungsweise als sinn-

volle Ergänzung einer Ultraschalluntersuchung auch im Basisultraschall dienen kann. Da die Ultraschalluntersuchung immer auch von der Erfahrung des Untersuchers abhängt, könnte die künstliche Intelligenz in der Ausbildung von jungen Ärztinnen und Ärzten helfen, Fehler zu verhindern und die korrekte Darstellung der fetalen Strukturen zu üben. Die fehlende direkte Überprüfung der KI-Bilder durch einen erfahrenen Untersucher ist ebenfalls wie oben erwähnt ein Schwachpunkt der Studie. In der oben genannten Studie von Salim et al. mussten in einem Drittel der Fälle die automatischen Messungen vom Untersucher korrigiert werden [43]. Durch die direkte Korrektur und Neuaufnahme eines Volumens wäre die Genauigkeit wahrscheinlich noch deutlich verbessert worden. Eine weitere Limitation ist, dass in dieser Untersuchung nur Feten im zweiten Trimenon untersucht wurden. Mehrere Studien zeigen die Messungengenauigkeit im dritten Trimenon auf, allerdings ohne Nutzung der KI [24]. Interessant wären also weitere Untersuchungen zur Anwendung der künstlichen Intelligenz in den Tagen vor der Entbindung um zu prüfen, ob hierdurch die Messungengenauigkeit vermindert werden könnte.

Matrixsonde und fetaler Aortenbogen Auch für diesen Teil der Studie wurden Mehrlinge und Feten mit Fehlbildungen ausgeschlossen, sodass über diese keine Aussage zur Nutzbarkeit der Technik getroffen werden kann. Andere Studien zur Biplanen Bildgebung konnten aber bereits zeigen, dass mit dieser Technik Fehlbildungen im Ausflusstrakt des fetalen Herzens dargestellt werden können [52]. Somit kann die Biplane Technik eine sinnvolle Ergänzung der fetalen Echokardiographie sein. In dieser Studie wurde die Technik von einem Experten für Ultraschall angewandt, sodass unklar bleibt, ob auch weniger erfahrene Anwender den Aortenbogen in dieser Art und Weise sinnvoll darstellen können. Mehrere Autoren beschreiben die Technik aber als relativ simpel und gut durchführbar [53][52], wobei unklar bleibt, wie erfahren die Untersucherinnen und Untersucher in den jeweiligen Studien waren. Weitere Studien, ob die elektronische Matrixsonde unter anderem durch verbesserte Eindringtiefe Fehlbildungen auch im Rahmen der Basisversorgung in der normalen Schwangerenvorsorge ohne Feindiagnostik besser aufzeigen kann sind nötig, um vor allem im low-risk-Kollektiv Feten mit Fehlbildungen zu erkennen und zur adäquaten Weiterbehandlung zu überweisen.

6.3 Ausblick

Durch hochauflösende Ultraschallgeräte und künstliche Intelligenz können stetig mehr fetale Auffälligkeiten detektiert werden. Dies vermittelt auf der einen Seite den Schwangeren ein Gefühl von Sicherheit, kann aber auch zu viel Unsicherheit und Sorgen während der Schwangerschaft führen. Die Schwangere erwartet nach einer unauffälligen Feindiagnostik ein gesundes Kind-dies ist aber nicht zwangsläufig der Fall, es bleiben weiterhin zahlreiche Erkrankungen und Auffälligkeiten, die wir pränatal trotz verbesserter Technik (noch) nicht detektieren können. So hat die Pränataldiagnostik weiterhin ihre Grenzen. Zum anderen entsteht der Eindruck, dass durch künstliche Intelligenz der menschliche Untersucher kaum noch gebraucht wird. Ein komplettes Scannen des Feten von Kopf bis Fuß durch die künstliche Intelligenz erscheint denkbar und in naher Zukunft. Hier stellt sich die Frage, ob Pränatalmedizinerinnen und Pränatalmediziner bald nicht mehr gebraucht werden. In der vorliegenden Studie zeigte sich allerdings, dass die menschliche Kontrolle notwendig ist. Die Kontrolle der künstlichen Intelligenz erfordert menschliche Intelligenz-Untersucherinnen und Untersucher müssen Expertise im Ultraschall und in der fetalen Anatomie mitbringen um automatische Messungen sinnvoll interpretieren zu können. Daher kann und sollte künstliche Intelligenz nicht als Bedrohung sondern als Zugewinn für die ärztliche Tätigkeit betrachtet werden. Auch in anderen Bereichen hat künstliche Intelligenz längst Einzug in unser Leben erhalten ohne dass dies negativ wahrgenommen wird, ein Beispiel ist die Gesichtserkennung moderner Smartphones. Hier wird die Technik als Zugewinn angesehen. Künstliche Intelligenz wird auch in Zukunft nicht über menschliche Eigenschaften wie Empathie und Kreativität verfügen, somit muss sich der Mensch keine Sorgen machen, ganz von der Maschine abgelöst zu werden. Maschinen können nicht adäquat mit Patientinnen kommunizieren. Für die Zukunft ist zu erwarten, dass durch Nutzung von „SonoCNS“ die Darstellung der fetalen Anatomie des Kopfes im zweiten Trimenon verbessert werden kann, da durch standardisierte Anwendung reproduzierbare Bilder entstehen, die auch nachträglich noch abgerufen werden können und möglicherweise die Messungenaugigkeit weiter verringern. Der Algorithmus kann noch verbessert werden, um auch die Cisterna magna besser darzustellen, ähnlich dem Lernprozess des menschlichen Gehirns. Außerdem könnte das Programm sinnvoll in der Ausbildung von jungen Ärztinnen und Ärzten genutzt werden und dabei helfen, Fehlbildungen auch im Basisultraschall zu erkennen. Bei korrekter Anwendung der Technik ist auch eine Zeit-

ersparnis möglich, die für weitere Untersuchungen oder das Patientengespräch genutzt werden kann. Die Nutzung von neuen Technologien hat den Menschen schon weit voran gebracht, die zunehmende Nutzung der künstlichen Intelligenz in der Medizin war somit nur eine Frage der Zeit. Offen bleibt die Frage nach der juristischen Haftung im Falle von Fehldiagnosen oder übersehenen Auffälligkeiten während der Nutzung von künstlicher Intelligenz [54].

Literatur

- [1] Gemeinsamer Bundesausschuss (G-BA). Richtlinien über die ärztliche Betreuung während der Schwangerschaft und nach der Entbindung. https://www.g-ba.de/downloads/62-492-2301/Mu-RL_2020-08-20_iK-2020-11-24.pdf. Accessed: 2021-04-15.
- [2] Mutterschaftsrichtlinien. <https://g-ba.de>. Accessed: 2023-04-07.
- [3] CL Van Velzen, MC Haak, G Reijnders, MEB Rijlaarsdam, CJ Bax, E Pajkrt, J Hrudá, F Galindo-Garre, CM Bilardo, CJM De Groot u. a. „Prenatal detection of transposition of the great arteries reduces mortality and morbidity“. In: *Ultrasound in Obstetrics & Gynecology* 45.3 (2015), S. 320–325.
- [4] Amboss Sonographie. <https://amboss.com>. Accessed: 2023-04-07.
- [5] 4. Auflage Steiner H. Schneider K.-T.M. Dopplersonographie in Gynäkologie und Geburtshilfe, Seite 33. Springer-Verlag, 2017.
- [6] Karl Theo Dussik. „Über die Möglichkeit, hochfrequente mechanische Schwingungen als diagnostisches Hilfsmittel zu verwerten“. In: *Zeitschrift für die gesamte Neurologie und Psychiatrie* 174.1 (1942), S. 153–168.
- [7] Douglass H Howry, Gerald Posakony, C Richard Cushman und Joseph H Holmes. „Three-dimensional and stereoscopic observation of body structures by ultrasound“. In: *Journal of Applied Physiology* 9.2 (1956), S. 304–306.
- [8] Ultraschallmuseum. <https://www.ultraschallmuseum.de/download/Tafeln-7-11Korr-Febr2007-Internet.pdf>. Accessed: 2021-11-26.
- [9] Walter Krause und Richard Soldner. „Ultraschallbildverfahren (B-scan) mit hoher Bildfrequenz für medizinische Diagnostik“. In: *Electromedica* 4 (1967), S. 8–11.
- [10] D Callaghan, TC Rowland und DE Goldman. „Ultrasonic observation of the fetal heart“. In: *Obstet Gynecol* 23 (1964), S. 637.
- [11] G Kossoff, WJ Garrett, DA Carpenter, J Jellins und MJ Dadd. „Principles and classification of soft tissues by grey scale echography“. In: *Ultrasound in medicine & biology* 2.2 (1976), S. 89–105.
- [12] Ch Sohn und G Rudofsky. „Die dreidimensionale Ultraschalldiagnostik“. In: *Ultraschall-diagnostik'89*. Springer, 1990, S. 16–19.

- [13] Ulrich Gembruch, Kurt Hecher und Horst Steiner. *Ultraschalldiagnostik in Geburtshilfe und Gynäkologie*. Springer-Verlag, 2013.
- [14] Ian Donald, John Macvicar und TG Brown. „Investigation of abdominal masses by pulsed ultrasound“. In: *The Lancet* 271.7032 (1958), S. 1188–1195.
- [15] Ian Donald und TG Brown. „I. Demonstration of Tissue Interfaces within the Body by Ultrasonic Echo Sounding“. In: *The British journal of radiology* 34.405 (1961), S. 539–546.
- [16] EC Coles, DG Altman, HB Meire und P Farrant. „Ultrasound estimation of gestational age.“ In: *British medical journal* 1.6175 (1979), S. 1422.
- [17] KP Hanretty, JP Neilson und JEE Fleming. „Re-evaluation of clinical estimation of fetal weight. A comparison with ultrasound“. In: *Journal of Obstetrics and Gynaecology* 10.3 (1990), S. 199–201.
- [18] Horace E Thompson, Joseph H Holmes, Kenneth R Gottesfeld und E Stewart Taylor. „Fetal development as determined by ultrasonic pulse echo techniques“. In: *American Journal of Obstetrics & Gynecology* 92.1 (1965), S. 44–52.
- [19] MHUA Hansmann, H Schumacher und U Voigt. „Mehrparametrische nicht lineare Gewichtsschätzung mittels Ultraschall unter Berücksichtigung des Gestationsalters“. In: *Ultraschalldiagnostik*. Stuttgart: Georg Thieme Verlag (1978), S. 69–71.
- [20] KJ Dornan, M Hansmann, DHA Redford und BK Wittmann. „Fetal weight estimation by real-time ultrasound measurement of biparietal and transverse trunk diameter“. In: *American journal of obstetrics and gynecology* 142.6 (1982), S. 652–657.
- [21] Frank P. Hadlock, R. B. Harrist, Ralph S. Sharman, Russell L. Deter und Seung K. Park. „Estimation of fetal weight with the use of head, body, and femur measurements—A prospective study“. In: *American Journal of Obstetrics and Gynecology* (1985). ISSN: 00029378. DOI: 10.1016/0002-9378(85)90298-4.
- [22] RL Schild, K Fell, R Fimmers, U Gembruch und M Hansmann. „A new formula for calculating weight in the fetus of ≤ 1600 g“. In: *Ultrasound in Obstetrics and Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology* 24.7 (2004), S. 775–780.

- [23] NC Hart, A Hilbert, B Meurer, M Schrauder, M Schmid, J Siemer, M Voigt und RL Schild. „Macrosomia: a new formula for optimized fetal weight estimation“. In: *Ultrasound in Obstetrics and Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology* 35.1 (2010), S. 42–47.
- [24] J Siemer, N Egger, N Hart, B Meurer, A Müller, O Dathe, T Goecke und RL Schild. „Fetal weight estimation by ultrasound: comparison of 11 different formulae and examiners with differing skill levels“. In: *birth* 1 (2008), S. 0–7.
- [25] S Kehl, U Schmidt, S Spaich, RL Schild, M Sütterlin und J Siemer. „What are the limits of accuracy in fetal weight estimation with conventional biometry in two-dimensional ultrasound? A novel postpartum study“. In: *Ultrasound in obstetrics & gynecology* 39.5 (2012), S. 543–548.
- [26] Mark K Friedberg, Norman H Silverman, Anita J Moon-Grady, Elizabeth Tong, Jennifer Nourse, Beatrice Sorenson, Jaimie Lee und Lisa K Hornberger. „Prenatal detection of congenital heart disease“. In: *The Journal of pediatrics* 155.1 (2009), S. 26–31.
- [27] A Lindinger, G Schwedler und H-W Hense. „Prevalence of congenital heart defects in newborns in Germany: Results of the first registration year of the PAN Study (July 2006 to June 2007)“. In: *Klinische Pädiatrie* 222.05 (2010), S. 321–326.
- [28] Rabih Chaoui, Kai-Sven Heling, Nanette Sarioglu, Michael Schwabe, Anja Dankof und Rainer Bollmann. „Aberrant right subclavian artery as a new cardiac sign in second-and third-trimester fetuses with Down syndrome“. In: *American journal of obstetrics and gynecology* 192.1 (2005), S. 257–263.
- [29] GE-Sonostore. <https://ge-sonostore.com>. Accessed: 2023-04-07.
- [30] G Grangé, M Althuser, J Fresson, A Bititi, K Miyamoto, V Tsatsaris und O Morel. „Semi-automated adjusted measurement of nuchal translucency: feasibility and reproducibility“. In: *Ultrasound in obstetrics & gynecology* 37.3 (2011), S. 335–340.
- [31] KO Kagan, H Abele, B Hoffmann-Pröll, D Wallwiener und M Hoopmann. „Intra- und Interobservervariabilität der manuellen und automatischen NT-Messung“. In: *Ultraschall in der Medizin-European Journal of Ultrasound* 31.S 01 (2010), V1_04.
- [32] Kai-Sven Heling Rabih Chaoui. 3D Sonographie in der pränatalen Diagnostik, Kapitel 14. deGruyter, 2016.

- [33] GE Chalouhi, A Gueneuc, G Rameh, H Hamze, LJ Salomon und Y Ville. „Biplane mode for more precise intrauterine procedures“. In: *American Journal of Obstetrics and Gynecology* (2021).
- [34] Georg Borges. „Haftung für KI-Systeme—Konzepte und Adressaten der Haftung“. In: *Computer und Recht* 38.9 (2022), S. 553–560.
- [35] Fujiao He, Yaqin Wang, Yun Xiu, Yixin Zhang und Lizhu Chen. „Artificial intelligence in prenatal ultrasound diagnosis“. In: *Frontiers in medicine* 8 (2021), S. 729978.
- [36] Eurocat. https://eu-rd-platform.jrc.ec.europa.eu/eurocat/eurocat-data/prenatal-screening-and-diagnosis_en. Accessed: 2021-11-26.
- [37] Deutsche Gesellschaft für Ultraschall in der Medizin e.V. www.degum.de. Accessed: 2021-04-15.
- [38] GE-Ultraschall. <https://ge-ultraschall.com>. Accessed: 2023-04-07.
- [39] I Sarris, C Ioannou, P Chamberlain, E Ohuma, F Roseman, L Hoch, DG Altman, AT Papageorghiou, International Fetal und Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st). „Intra-and interobserver variability in fetal ultrasound measurements“. In: *Ultrasound in obstetrics & gynecology* 39.3 (2012), S. 266–273.
- [40] Meng-Jie Lu, Wei-Hua Zhong, Yu-Xiu Liu, Hua-Zhang Miao, Yong-Chang Li und Mu-Huo Ji. In: *The International Journal of Biostatistics* 11.1 (2015), S. 20150038. DOI: doi:9.1515/ijb-2015-0039. URL: <https://doi.org/9.1515/ijb-2015-0039>.
- [41] Y Nagayasu, D Fujita und M Omichi. „VP47. 05: Automating the process of measuring the fetal brain using three-dimensional fetal ultrasonography“. In: *Ultrasound in Obstetrics & Gynecology* 58 (2021), S. 300–300.
- [42] A Welp, MA Gembicki, C Dracopoulos, J Scharf und J Weichert. „VP47. 04: Comparison of two semiautomated volume approaches in clinical use: can coronal and sagittal planes provide a diagnostic benefit?“ In: *Ultrasound in Obstetrics & Gynecology* 58 (2021), S. 300–300.

- [43] I Salim, A Cavallaro, C Ciofalo-Veit, L Rouet, C Raynaud, B Mory, A Collet Billon, G Harrison, D Roundhill und AT Papageorghiou. „Evaluation of automated tool for two-dimensional fetal biometry“. In: *Ultrasound in Obstetrics & Gynecology* 54.5 (2019), S. 650–654.
- [44] JG Martins, T Kawakita, M Gurganus, D Baraki, P Jain, AT Papageorghiou und AZ Abuhamad. „The influence of maternal body mass index on interobserver variability of fetal ultrasound biometry and amniotic fluid assessment in late pregnancy“. In: *Ultrasound in Obstetrics & Gynecology* (2021).
- [45] NJ Dudley und E Chapman. „The importance of quality management in fetal measurement“. In: *Ultrasound in Obstetrics and Gynecology* 19.2 (2002), S. 190–196.
- [46] GE-Healthcare. <https://www.gehealthcare.com/products/ultrasound/voluson/voluson-e10>. Accessed: 2023-09-24.
- [47] Kjersti M Aagaard-Tillery, T Flint Porter, Fergal D Malone, David A Nyberg, Jamie Collins, Christine H Comstock, Gary Hankins, Keith Eddleman, Lorraine Dugoff, Honor M Wolfe u. a. „Influence of maternal BMI on genetic sonography in the FaSTER trial“. In: *Prenatal Diagnosis: Published in Affiliation With the International Society for Prenatal Diagnosis* 30.1 (2010), S. 14–22.
- [48] Henry Adekola, Eleazar Soto, Jing Dai, Jennifer Lam-Rachlin, Navleen Gill, Jocelyn Leon-Peters, Karoline Puder und Jacques S Abramowicz. „Optimal visualization of the fetal four-chamber and outflow tract views with transabdominal ultrasound in the morbidly obese: Are we there yet?“ In: *Journal of Clinical Ultrasound* 43.9 (2015), S. 548–555.
- [49] Beryl Benacerraf. „The use of obstetrical ultrasound in the obese gravida“. In: *Seminars in perinatology*. Bd. 37. 5. Elsevier. 2013, S. 345–347.
- [50] Judith H Chung, Raquel Pelayo, Tamera J Hatfield, Vinita J Speir, Jun Wu und Aaron B Caughey. „Limitations of the fetal anatomic survey via ultrasound in the obese obstetrical population“. In: *The Journal of Maternal-Fetal & Neonatal Medicine* 25.10 (2012), S. 1945–1949.
- [51] BS Mahony, PW Callen, RA Filly und WK Hoddick. „The fetal cisterna magna.“ In: *Radiology* 153.3 (1984), S. 773–776.

- [52] Yi Xiong, Tao Liu, Han Jing Gan, Ying Wu, Jin Feng Xu, Yuen Ha Ting, Tak Yeung Leung und Tze Kin Lau. „Detection of the fetal conotruncal anomalies using real-time three-dimensional echocardiography with live xPlane imaging of the fetal ductal arch view“. In: *Prenatal diagnosis* 33.5 (2013), S. 462–466.
- [53] Y Xiong, M Chen, LW Chan, YH Ting, TY Fung, TY Leung und TK Lau. „A novel way of visualizing the ductal and aortic arches by real-time three-dimensional ultrasound with live xPlane imaging“. In: *Ultrasound in obstetrics & gynecology* 39.3 (2012), S. 316–321.
- [54] L Drukker, JA Noble und AT Papageorghiou. „Introduction to artificial intelligence in ultrasound imaging in obstetrics and gynecology“. In: *Ultrasound in Obstetrics & Gynecology* 56.4 (2020), S. 498–505.

Eidesstattliche Versicherung

„Ich, Julia Fabienne Umlauff, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: Ultraschall in der Pränataldiagnostik-Standardisierung durch künstliche Intelligenz am Beispiel der Neurosonographie und der Darstellung des Aortenbogens (Ultrasound in prenatal diagnosis-Standardisation through artificial intelligence based on the example of neurosonography and demonstration of aortic arch) selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren/innen beruhen, sind als solche in korrekter Zitierung kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) werden von mir verantwortet.

Ich versichere ferner, dass ich die in Zusammenarbeit mit anderen Personen generierten Daten, Datenauswertungen und Schlussfolgerungen korrekt gekennzeichnet und meinen eigenen Beitrag sowie die Beiträge anderer Personen korrekt kenntlich gemacht habe (siehe Anteilserklärung). Texte oder Textteile, die gemeinsam mit anderen erstellt oder verwendet wurden, habe ich korrekt kenntlich gemacht.

Meine Anteile an etwaigen Publikationen zu dieser Dissertation entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem/der Erstbetreuer/in, angegeben sind. Für sämtliche im Rahmen der Dissertation entstandenen Publikationen wurden die Richtlinien des ICMJE (International Committee of Medical Journal Editors; www.icmje.org) zur Autorenschaft eingehalten. Ich erkläre ferner, dass ich mich zur Einhaltung der Satzung der Charité – Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis verpflichte.

Weiterhin versichere ich, dass ich diese Dissertation weder in gleicher noch in ähnlicher Form bereits an einer anderen Fakultät eingereicht habe.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§§156, 161 des Strafgesetzbuches) sind mir bekannt und bewusst.“

01.11.23

Datum

Unterschrift

Lebenslauf

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht

Danksagung

Ich danke meinem Doktorvater Herrn PD Dr. med. Heling für die Vergabe des interessanten Themas und die exzellente Betreuung. Ich danke meinem Freund und liebsten Menschen Joscha von Hein ohne dessen Unterstützung diese Arbeit nicht möglich gewesen wäre. Dank gilt auch meinen Eltern Edith und Dietmar Umlauff für die fortwährende Unterstützung bei der Umsetzung all meiner Pläne.



Name, Vorname: Umlauff, Julia Fabienne

Emailadresse: alex.hildebrandt@charite.de

Matrikelnummer: 227646

PromotionsbetreuerIn: PD Dr. med. Kai-Sven Heling

**Promotionsinstitution/Klinik: Praxis für Pränataldiagnostik in
der Friedrichstrasse 147 in Berlin**

Bescheinigung

Hiermit bescheinige ich, dass Frau Julia Fabienne Umlauff innerhalb der Service Unit Biometrie des Instituts für Biometrie und Klinische Epidemiologie (iBikE) bei mir eine statistische Beratung zu einem Promotionsvorhaben wahrgenommen hat. Folgende Beratungstermine wurden wahrgenommen:

- Termin 1: 22.04.2020
- Termin 2: 08.05.2020
- Termin 3: 10.08.2022

Folgende wesentliche Ratschläge hinsichtlich einer sinnvollen Auswertung und Interpretation der Daten wurden während der Beratung erteilt:

- Verwendung von Bland-Altman-Plots für den Vergleich zweier Messmethoden
- Verwendung multivariabler Regressionsmodelle statt univariabler Variablen-Selektion

Diese Bescheinigung garantiert nicht die richtige Umsetzung der in der Beratung gemachten Vorschläge, die korrekte Durchführung der empfohlenen statistischen Verfahren und die richtige Darstellung und Interpretation der Ergebnisse. Die Verantwortung hierfür obliegt allein dem Promovierenden. Das Institut für Biometrie und Klinische Epidemiologie übernimmt hierfür keine Haftung.

Datum: 26.08.2022



Name des Beraters: [REDACTED]

UNIVERSITÄTSMEDIZIN BERLIN
Institut für Biometrie und
Klinische Epidemiologie
Campus Charité Mitte
Charitéplatz 1 | D-10117 Berlin
Sitz: Reinhardtstr. 58

Unterschrift BeraterIn, Institutsstempel